

# Kapitel 4

## Modelle und Statistiken

In letzter Zeit werden vermehrt Parameter (Gradfolgen, Kernzahlfolgen, etc.) empirischer Graphen (Internet, WWW, Proteine, etc.) berechnet und diskutiert. Insbesondere interessieren dabei die inhärenten Bildungsgesetze, welche die beobachteten Ausprägungen dieser Parameter hervorrufen.

Wir behandeln in diesem Abschnitt einige Modelle und Algorithmen zur Beschreibung und effizienten Erzeugung von Multigraphen, und untersuchen Zusammenhänge zwischen Modellen und Parametern.

### 4.1 Definition (Graphenmodelle)

Ein (Graphen-)Modell  $(\mathcal{K}, P)$  besteht aus einer Klasse  $\mathcal{K}$  von Graphen zusammen mit einem Wahrscheinlichkeitsmaß  $P$  auf dieser Klasse.

Im folgenden sei  $X$  eine Zufallsvariable für Graphen in  $\mathcal{K}$ . Ist  $G = (V, E) \in \mathcal{K}$  ein fester Graph, so wird seine Wahrscheinlichkeit mit  $P(X = G)$  oder kurz  $P(G)$  bezeichnet. Entsprechend bezeichnen wir die Wahrscheinlichkeit, dass ein zufälliger Graph des Modells eine bestimmte Eigenschaft aufweist, z.B. mit  $P(\Delta(G) = 7)$  oder  $P(e \in E)$ . Der Erwartungswert einer Eigenschaft wird mit  $\langle m(G) \rangle$ ,  $\langle d_G(v) \rangle$  usw. bezeichnet.

Wir betrachten zunächst ein Modell für schlichte ungerichtete Graphen. Sei dazu  $\mathcal{G}(n)$  die Klasse der ungerichteten Graphen mit Knotenmenge  $V_n = \{0, \dots, n-1\}$ .

**4.2 Definition (Zufallsgraphen; Gilbert 1959)**

Für jedes  $n \in \mathbb{N}_0$  und  $0 \leq p \leq 1$  heißt das Modell  $\mathcal{G}(n, p) = (\mathcal{G}(n), P)$  mit  $P$  induziert durch

$$P(\{v, w\} \in E) = p \quad \text{unabhängig für alle } v \neq w \in V_n$$

Zufallsgraph (mit  $n$  Knoten und Kantenwahrscheinlichkeit  $p$ ). Wir vereinbaren  $q = 1 - p$ .

**4.3 Lemma**

$$P(G) = p^m q^{\frac{n(n-1)}{2} - m}$$

■ **Beweis:** Da alle Kantenwahrscheinlichkeiten unabhängig sind, gilt

$$P(X = G) = \prod_{v \neq w \in V} P(\{v, w\} \in E) = \prod_{\{v, w\} \in E} p \cdot \prod_{\{v, w\} \notin E} q .$$

□

*Erzeugen* eines Graphen in einem Modell  $(\mathcal{K}, P)$  bedeutet, algorithmisch einen Graphen so zu konstruieren, dass man ein bestimmtes  $G \in \mathcal{K}$  mit Wahrscheinlichkeit  $P(G)$  erhält, d.h. wenn man bei  $k$  Aufrufen mit gleicher Eingabe  $k(G)$  mal den Graphen  $G$  erhalten hat, dann gilt

$$\frac{k(G)}{k} \xrightarrow{k \rightarrow \infty} P(G) .$$

Wir gehen davon aus, dass uns ein Zufallszahlengenerator  $R[0, 1]$  zur Verfügung steht, der unabhängig und gleichverteilt Zahlen im Intervall  $[0, 1]$  erzeugt.

Der naheliegende Algorithmus zur Erzeugung von Zufallsgraphen führt für jedes Knotenpaar ein Zufallsexperiment aus und verbindet die beiden Knoten genau dann, wenn  $R[0, 1] \leq p$ . Die Laufzeit dieses Konstruktionsalgorithmus' ist allerdings in  $\Theta(n^2)$ , unabhängig von  $p$ . Ist  $p$  klein, so ist der Graph höchstwahrscheinlich dünn und die Laufzeit wesentlich länger als der Graph groß.

Um zu einem effizienteren Algorithmus zu gelangen, betrachte die Knotenpaare  $(v, w) \in V \times V$  mit  $v < w$ . Wenn wir diese in irgendeiner Reihenfolge aufzählen, wie lange dauert es dann im naheliegende Algorithmus, bis wieder eine Kante eingefügt wird? Zu einem beliebigen Zeitpunkt kommt die nächste Kante mit Wahrscheinlichkeit

$$q^k p$$

nach genau  $k$  Fehlversuchen.

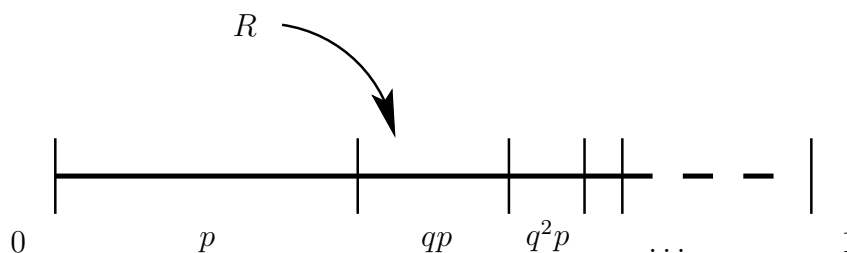
Um mit dem Zufallsgenerator eine Fehlversuchszahl  $k$  mit Wahrscheinlichkeit  $q^k p$  zu erzeugen, ordnen wir jedem  $k$  ein Teilintervall von  $[0, 1]$  der Länge  $q^k p$  zu. Man beachte, dass

$$\sum_{k=0}^{\infty} q^k p = p \cdot \sum_{k=0}^{\infty} q^k = p \cdot \frac{1}{1-q} = 1$$

ist. Sind die Intervalle der Größe nach geordnet, so endet das Intervall von  $k$  bei

$$\sum_{i=0}^k q^i p = p \cdot \sum_{i=0}^k q^i = p \cdot \frac{1 - q^{k+1}}{1 - q} = 1 - q^{k+1}.$$

Zu einer Zufallszahl  $R \leftarrow R[0, 1]$  brauchen wir also nur das kleinste  $k$  zu bestimmen, dessen zugehöriges Intervall hinter  $R$  endet:



$$\begin{aligned} R &< 1 - q^{k+1} \\ q^{k+1} &< 1 - R \\ e^{(k+1) \log q} &< e^{\log(1-R)} \\ k &> \frac{\log(1-R)}{\log q} - 1 \end{aligned}$$

Wir lassen daher  $\left\lfloor \frac{\log(1-R)}{\log q} \right\rfloor$  potentielle Kanten aus. Wegen der Gleichverteilung des Zufallszahlengenerators kann  $\log R$  anstelle von  $\log(1-R)$  verwendet werden.

**Algorithmus 22:** Erzeugen von Zufallsgraphen**Eingabe:** Knotenzahl  $n \in \mathbb{N}_0$ , Kantenwahrscheinlichkeit  $0 < p < 1$ **Ausgabe:** Graph  $G = (\{0, \dots, n-1\}, E)$  $E \leftarrow \emptyset$  $r \leftarrow \frac{1}{\log(1-p)}$  $v \leftarrow 1; w \leftarrow -1$ **while**  $v < n$  **do**     $w \leftarrow w + 1 + \lceil r \cdot \log R[0, 1] \rceil$     **while**  $w \geq v$  **and**  $v < n$  **do**  $w \leftarrow w - v; v \leftarrow v + 1$     **if**  $v < n$  **then**  $E \leftarrow E \cup \{v, w\}$ **4.4 Satz**

Der Algorithmus erzeugt einen Zufallsgraphen  $G = (V, E)$  in Zeit  $\mathcal{O}(n + m)$  und benötigt dafür genau  $m + 1$  Zufallszahlen.

■ **Beweis:** Für die Korrektheit brauchen wir nur noch zu zeigen, dass alle möglichen Paare  $(v, w) \in V \times V$  mit  $v < w$  berücksichtigt werden. Eine lexikographische Aufzählung der Paare entspricht aber einem zeilenweisen Durchlauf durch die untere Dreiecksmatrix der Adjazenzmatrix. Die folgende Skizze zeigt, warum die Indexberechnungen korrekt sind.

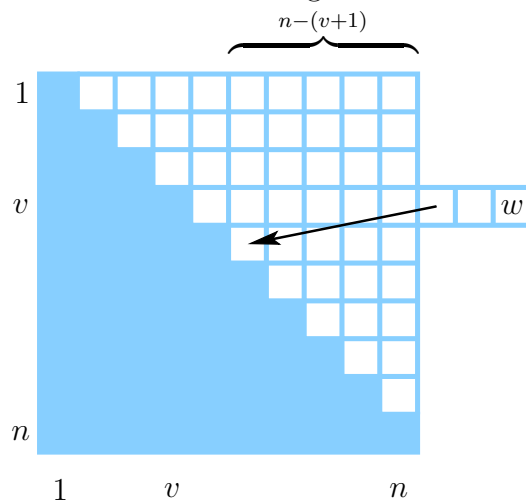


Bild  
ersetzen  
(transponieren)

Zur Laufzeit beachte, dass in jedem (bis auf den letzten) Durchlauf der äußeren Schleife eine Kante erzeugt und in jedem Durchlauf der inneren Schleife die Nachbarschaft eines Knotens abgeschlossen wird.  $\square$

Welche Eigenschaften haben die erzeugten Graphen? Die erwartete Anzahl von Kanten ist

$$\langle m \rangle = \sum_{v \neq w \in V} (P(\{v, w\} \in E) \cdot 1 + P(\{v, w\} \notin E) \cdot 0) = p \cdot \frac{n(n-1)}{2},$$

sodass für festes  $0 < p < 1$  gilt  $\langle m \rangle \in \Theta(n^2)$ .

Da wir vor allem an der Modellierung dünner Graphen interessiert sind, sollte  $p$  also geeignet in Abhängigkeit von  $n$  gewählt werden.

Wählt man die Kantenwahrscheinlichkeit  $p$  abhängig von der Knotenzahl  $n$ , dann treten *Phasenübergänge* auf, unterhalb derer bestimmte Eigenschaften fast immer gegeben sind, aber oberhalb nicht (oder umgekehrt).

#### 4.5 Definition (Eigenschaften fast aller/keines Graphen)

Eine Eigenschaft  $\mathcal{A} \subseteq \mathcal{K}$  haben fast alle Graphen eines Modells  $(\mathcal{K}, P)$ , wenn

$$\lim_{n \rightarrow \infty} P(G \in \mathcal{A}) = 1$$

bzw. hat sie fast kein Graph des Modells, wenn

$$\lim_{n \rightarrow \infty} P(G \in \mathcal{A}) = 0.$$

#### 4.6 Definition (Schwellenfunktion)

Eine Funktion  $t(n)$  heißt Schwelle(nfunktion) einer Eigenschaft  $\mathcal{A}$  von Zufallsgraphen, wenn für alle Wahlen von  $p = p(n)$

$$\lim_{n \rightarrow \infty} P(G \in \mathcal{A}) = \begin{cases} 0 & \text{falls } \frac{p(n)}{t(n)} \xrightarrow{n \rightarrow \infty} 0 \\ 1 & \text{falls } \frac{p(n)}{t(n)} \xrightarrow{n \rightarrow \infty} \infty \end{cases}$$

#### 4.7 Bemerkung

Mit  $t(n)$  ist auch jede andere Funktion in  $\Theta(t(n))$  Schwellenfunktion.

Aus  $\langle m \rangle = p \cdot \frac{n(n-1)}{2}$  können wir bereits folgern, dass unterhalb der Schwelle  $\frac{1}{n^2}$ , d.h. für  $p(n) \in o(\frac{1}{n^2})$ , fast kein Zufallsgraph eine Kante enthält. Analog sind fast alle Zufallsgraphen oberhalb von  $1 - \frac{1}{n^2}$  vollständig.

#### 4.8 Bemerkung

Für zahlreiche interessante Eigenschaften gibt es solche Schwellen. So ist  $\frac{1}{n\sqrt{n}}$  Schwelle für  $\Delta(G) > 1$ ,  $\frac{1}{n}$  Schwelle für das Enthalten eines Kreises (d.h. unterhalb sind fast alle Zufallsgraphen Wälder) und  $\frac{\log n}{n}$  Schwelle dafür, dass fast alle Graphen zusammenhängend sind.

## Exkurs: Die Probabilistische Methode

Das ursprüngliche Anwendungsgebiet von Zufallsgraphen sind *Existenzaussagen* für Graphen mit bestimmten Eigenschaften.

### 4.9 Definition (Ramsey-Zahlen)

Für  $r, s \in \mathbb{N}$  mit  $r, s \geq 2$  hat ein Graph  $G \in \mathcal{G}(n)$  die Eigenschaft  $(r, s)$ , wenn er einen knoteninduzierten Teilgraphen  $G_r \in \mathcal{G}(r)$  enthält, der vollständig verbunden ist (eine Clique  $K_r$  der Größe  $r$ ), oder einen knoteninduzierten Teilgraph  $G_s \in \mathcal{G}(s)$ , der nur aus isolierten Knoten besteht (eine unabhängige Menge der Größe  $s$ ). Das kleinste  $n \in \mathbb{N}$ , für das jeder Graph in  $\mathcal{G}(n)$  die Eigenschaft  $(r, s)$  hat, ist die Ramsey-Zahl  $R(r, s)$ .

### 4.10 Bemerkung

Falls jeder Graph in  $\mathcal{G}(n)$  die Eigenschaft  $(r, s)$  hat, dann auch jeder Graph in  $\mathcal{G}(n')$  mit  $n' > n$ . Ein Graph hat die Eigenschaft  $(r, s)$  genau dann, wenn sein Komplementgraph die Eigenschaft  $(s, r)$  hat.

Es gilt  $R(r, 2) = r$ , denn entweder ist  $G \in \mathcal{G}(r)$  eine Clique, oder es gibt zwei nicht-adjazente (unabhängige) Knoten (für  $n < r$  haben die Cliquen  $K_n$  nicht die Eigenschaft  $(r, 2)$ ). Aus Symmetriegründen folgt  $R(2, s) = s$ . Wir leiten zunächst eine obere Schranke für die Ramsey-Zahlen her.

### 4.11 Lemma

Für  $r, s > 2$  ist  $R(r, s) \leq R(r-1, s) + R(r, s-1)$ .

■ **Beweis:** Seien  $n_r = R(r-1, s)$  und  $n_s = R(r, s-1)$  und  $G = (V, E) \in \mathcal{G}(n)$  für  $n = n_r + n_s$ . Wir zeigen, dass  $G$  die Eigenschaft  $(r, s)$  hat. Für einen beliebigen Knoten  $v \in V$  sei  $N^*(v) = V \setminus (N(v) \cup \{v\})$  die Menge der von  $v$  unabhängigen Knoten. Wegen  $|N(v)| + |N^*(v)| = n - 1$  ist entweder  $|N(v)| \geq n_r$  oder  $|N^*(v)| \geq n_s$ . Ist  $|N(v)| \geq n_r$ , dann gibt es in  $N(v)$  eine Clique der Größe  $r-1$ , die zusammen mit  $v$  eine Clique der Größe  $r$  ergibt, oder eine unabhängige Menge  $N_s$  der Größe  $s$ . Der Fall  $|N^*(v)| \geq n_s$  ist analog, sodass  $G$  in jedem Fall die Eigenschaft  $(r, s)$  hat.  $\square$

### 4.12 Folgerung

$$R(r, s) \leq \binom{r+s-2}{r-1}$$

■ **Beweis:** Durch Induktion über  $r, s \geq 2$ . Zunächst sind  $R(r, 2) = r = \binom{r}{r-1}$  und  $R(2, s) = s = \binom{s}{1}$ . Mit dem Lemma und der Induktionsvorausset-

zung folgt

$$\begin{aligned} R(r, s) &\leq R(r-1, s) + R(r, s-1) \\ &\leq \binom{r+s-3}{r-2} + \binom{r+s-3}{r-1} = \binom{r+s-2}{r-1} \end{aligned}$$

aus der Rekursionsformel für Binomialzahlen.  $\square$

Insbesondere ist also  $R(r, r) \leq \binom{2r-2}{r-1} = \binom{2r-3}{r-1} + \binom{2r-3}{r-2} \leq 2^{2r-3}$  für alle  $r \geq 2$ . Eine (interessantere) untere Schranke wird mit der Probabilistischen Methode bewiesen.

#### 4.13 Satz

$$R(r, r) \geq 2^{\frac{r}{2}}$$

■ **Beweis:** Wir haben bereits  $R(2, 2) = 2$  und (s. oben)  $R(3, 3) \leq 6$ . Der ungerichtete Kreis  $C_5$  hat nicht die Eigenschaft  $(3, 3)$ , sodass  $R(3, 3) = 6$ .

Für  $r \geq 4$  sei  $n < 2^{\frac{r}{2}}$ . Wir zeigen, dass es in  $\mathcal{G}(n)$  einen Graphen geben muss, der nicht die Eigenschaft  $(r, r)$  hat. Im Modell  $\mathcal{G}(n, \frac{1}{2})$  hat jeder Graph  $G \in \mathcal{G}(n)$  Wahrscheinlichkeit

$$P(G) = \frac{1}{2^{\binom{n}{2}}}.$$

Insbesondere induziert eine beliebige Teilmenge  $A \subseteq V_n$  mit  $|A| = r$  eine Clique mit Wahrscheinlichkeit  $2^{-\binom{r}{2}}$ . Wegen  $n < 2^{\frac{r}{2}}$  ist daher die Wahrscheinlichkeit, irgend eine Clique der Größe  $r$  zu finden,

$$\begin{aligned} \sum_{A \subseteq V_n: |A|=r} 2^{-\binom{r}{2}} &= \binom{n}{r} 2^{-\binom{r}{2}} = \frac{n!}{r!(n-r)!} \cdot 2^{-\binom{r}{2}} \\ &= \frac{n(n-1) \cdots (n-r+1)}{r(r-1) \cdots 1} \cdot 2^{-\binom{r}{2}} \leq \frac{n^r}{2^{r-1}} \cdot 2^{-\binom{r}{2}} \\ &< 2^{\frac{r^2}{2} - (r-1) - \binom{r}{2}} = 2^{-\frac{r}{2} + 1} \\ &\leq \frac{1}{2}. \end{aligned}$$

Aus Symmetriegründen ist auch die Wahrscheinlichkeit, dass der Zufallsgraph eine unabhängige Menge der Größe  $r$  enthält, kleiner als  $\frac{1}{2}$ . Da beide Wahrscheinlichkeiten zusammen echt kleiner als 1 sind, muss es also einen Graphen mit  $n$  Knoten geben, der weder eine Clique noch eine unabhängige Menge der Größe  $r$  enthält.  $\square$

Zurück zur Netzwerkanalyse: Eine der am leichtesten zu bestimmenden Kenngrößen empirischer Netzwerke ist die Verteilung der Knotengrade. In Zufallsgraphen ist  $\langle d_G(v) \rangle = p \cdot (n - 1)$ , doch wie sieht die Verteilung aus?

#### 4.14 Lemma

Für Zufallsgraphen gilt

$$P(d_G(v) = k) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$

und unter der Voraussetzung  $(n-1) \cdot p(n) \xrightarrow{n \rightarrow \infty} \lambda > 0$

$$\lim_{n \rightarrow \infty} P(d_G(v) = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} .$$

für alle  $0 \leq k \leq n - 1$ .

■ **Beweis:** Aus der Beobachtung, dass jede der  $\binom{n-1}{k}$  möglichen Nachbarmengen Wahrscheinlichkeit  $p^k q^{n-1-k}$  hat, ergibt sich der erste Teil der Behauptung.

Für den zweiten Teil betrachte

$$\begin{aligned} \binom{n-1}{k} \cdot p^k q^{n-1-k} &= \frac{(n-1)!}{k!(n-1-k)!} \cdot p^k (1-p)^{n-1-k} \\ &= \frac{(n-1)!}{(n-1-k)!(1-p)^k} \cdot \frac{p^k}{k!} \cdot (1-p)^{n-1} \\ &= \frac{(n-1)!}{(n-1-k)!(1-p)^k (n-1)^k} \cdot \frac{(n-1)^k p^k}{k!} \cdot \left(1 - \frac{\lambda}{(n-1)} \cdot \frac{(n-1)p}{\lambda}\right)^{n-1} \end{aligned}$$

Für  $n \rightarrow \infty$  geht der erste Faktor nach Voraussetzung gegen 1, der zweite gegen  $\frac{\lambda^k}{k!}$  und der dritte gegen  $e^{-\lambda}$ .  $\square$



In großen Zufallsgraphen sind die Knotengrade bei festem Erwartungswert (und daher insbesondere sehr kleinem  $p$ ) also in etwa Poisson-verteilt. In vielen interessanten empirischen Netzwerken ist aber beobachtet worden, dass die Knotengrade einem Potenzgesetz folgen, d.h.

$$P(d_G(v) = k) \text{ verhält sich wie } \frac{1}{k^\gamma}$$

für ein  $\gamma > 0$  und zufällig gewähltes  $v \in V$ . Anschaulich bedeutet das, dass die Anzahl der Knoten höheren Grades rapide kleiner wird. Für diese Netzwerke scheint also ein anderes Modell nötig zu sein, in dem einige Knoten als Nachbarn *bevorzugt* werden.

Idee für ein Modell mit Bevorzugung (Barabási und Albert 1999):

“... ausgehend von einer kleinen Anzahl ( $m_0$ ) von Knoten, fügen wir in jedem Schritt einen neuen Knoten mit  $m (\leq m_0)$  Kanten ein, die den neuen Knoten mit  $m$  der bereits im System vorhandenen Knoten verbinden. Um in das Modell Bevorzugung zu integrieren, nehmen wir an, dass die Wahrscheinlichkeit  $\Pi$ , mit der der neue Knoten mit einem Knoten  $i$  verbunden wird, vom Grad  $k_i$  dieses Knotens abhängt, sodass  $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$ . Nach  $t$  Schritten führt das Modell zu einem zufälligen Netzwerk mit  $t + m_0$  Knoten und  $mt$  Kanten.”

Probleme mit dieser Definition:

- Da der Graph nach  $t$  Schritten  $mt$  Kanten hat, müssen die  $m_0$  initialen Knoten isoliert gewesen sein. Dann ist aber  $\Pi$  nicht wohldefiniert.
- Bei gleichzeitiger Wahl der  $m$  Nachbarn eines neuen Knotens reicht für die Auswahl einer Nachbarschaft die Angabe der Randverteilungen für die einzelnen Knoten nicht aus.

**Beispiel:** Betrachte den Fall  $m = 2$  und füge zu  $C_4$  (dem einfachen Kreis mit 4 Knoten) einen weiteren Knoten  $v$  hinzu. Wir legen die Wahrscheinlichkeiten für die sechs mögliche Nachbarschaften von  $v$  willkürlich fest. Die Wahrscheinlichkeit eines Paares, in dem die beiden Nachbarn von  $v$  selbst benachbart sind, sei  $p$  für ein beliebiges  $0 \leq p \leq \frac{1}{4}$ . Entsprechend ist die Wahrscheinlichkeit der beiden nicht benachbarten Paare  $\frac{1-4p}{2}$ . Für jeden Knoten des  $C_4$  ist die Auswahlwahrscheinlichkeit dann  $\frac{1}{4}$  wie gewünscht, unabhängig von  $p$ .

Wegen dieser Schwierigkeiten verwenden wir das folgende präzisere Modell. Da es über gerichteten Multigraphen definiert ist, sei im folgenden  $\mathcal{G}(n)$  die Klasse der gerichteten Multigraphen mit Knotenmenge  $V_n = \{1, \dots, n\}$ .

#### 4.15 Definition (Bevorzugung; Bollobás und Riordan 2001)

Für  $n \in \mathbb{N}_0$  ist das Multigraphenmodell mit Bevorzugung,  $\mathcal{G}_1^n = (\mathcal{G}(n), P)$ , wie folgt induktiv definiert. Für  $G = (V, E) \in \mathcal{G}(n+1)$  sei  $G' = G[V_n]$  der von den Knoten  $\{1, \dots, n\}$  induzierte Teilgraph von  $G$ . Dann ist

$$P(G) = \begin{cases} 0 & \text{falls } d_G^+(n+1) \neq 1 \text{ oder } m(G) - m(G') \neq 1 \\ P(G') \cdot \frac{d_{G'}(v)}{2n+1} & \text{falls } (n+1, v) \in E, v \leq n \\ P(G') \cdot \frac{1}{2n+1} & \text{falls } (n+1, v) \in E, v = n+1. \end{cases}$$

Zu  $b \in \mathbb{N}$  erhalten wir das Modell  $\mathcal{G}_b^n$  aus  $\mathcal{G}_1^{nb}$  durch Kontraktion von jeweils  $b$  aufeinanderfolgenden Knoten, d.h. ein Knoten  $1 \leq i \leq n$  eines Graphen im Modell  $\mathcal{G}_b^n$  entspricht den Knoten  $ib, \dots, ib+b-1$  eines Graphen im Modell  $\mathcal{G}_1^{nb}$ .

#### 4.16 Bemerkung

Wir können einen Multigraphen im Modell  $\mathcal{G}_1^n$  also als Ergebnis eines Prozesses ansehen, in dem zu jedem Zeitpunkt ein Knoten hinzukommt. Dieser erhält genau eine ausgehende Kante, deren Zielknoten mit Wahrscheinlichkeit proportional zu seinem Knotengrad (der Summe aus Ein- und Ausgangsgrad) gewählt wird, wobei für den neuen Knoten nur der Ausgangsgrad (das bereits feststehende Ende der Kante) gezählt wird. Bereits hochgradige Knoten werden also weiter bevorzugt.

Für den Erzeugungsalgorithmus gehen wir diesmal von Zufallszahlengeneratoren  $R\{0, \dots, r\}$  aus, die gleichverteilt und unabhängig ganze Zahlen aus  $\{0, \dots, r\}$  erzeugen. Die Idee des Algorithmus besteht darin, die gewünschte Auswahlwahrscheinlichkeit eines Knotens durch eine entsprechende Anzahl von Kopien zu bewirken.

**Algorithmus 23:** Erzeugen von Multigraphen mit Bevorzugung**Eingabe** : Knotenzahl  $n \in \mathbb{N}_0$ , Ausgangsgrad  $b \in \mathbb{N}_0$ **Daten** : Array  $N[0 \dots 2nb - 1]$  von Knoten  
(entspricht Liste der Kanten)**Ausgabe**: Multigraph  $G = (\{0, \dots, n - 1\}, E)$  $E \leftarrow \emptyset$  $m \leftarrow 0$ **foreach**  $v = 0, \dots, n - 1$  **do**    **foreach**  $j = 0, \dots, b - 1$  **do**         $N[2m] \leftarrow v$          $w \leftarrow N[R\{0, \dots, 2m\}]$          $N[2m + 1] \leftarrow w$          $E \leftarrow E \cup \{\{v, w\}\}$     inkrementiere  $m$ **4.17 Satz**

Der Algorithmus erzeugt einen Multigraphen mit Bevorzugung  $G = (V, E)$  in  $\mathcal{O}(n + m)$  Zeit und Platz und benötigt dafür genau  $m$  Zufallszahlen.

**4.18 Satz (Bollobás, Riordan, Spencer und Tusnády 2001)**Für  $a, b \in \mathbb{N}$  sei

$$\delta(a, b) = \frac{2b(b+1)}{(a+b)(a+b+1)(a+b+2)}$$

dann gilt für Multigraphen mit Bevorzugung  $\mathcal{G}_b^n$ 

$$P \left( (1 - \varepsilon) \cdot \delta(a, b) \leq \frac{|\{v \in V_n : d_G^-(v) = a\}|}{n} \leq (1 + \varepsilon) \cdot \delta(a, b) \right) \xrightarrow{n \rightarrow \infty} 1$$

für alle  $\varepsilon > 0$  und  $0 \leq a \leq n^{\frac{1}{15}}$ .

Insbesondere verhält sich die Eingangsgradverteilung also ungefähr wie  $\frac{1}{a^3}$ . Da der Ausgangsgrad  $b$  konstant ist, folgt der Knotengrad im zum symmetrisierten Multigraphen gehörigen ungerichteten Multigraphen ebenfalls einem Potenzgesetz. Dies war für die oben zitierte Modellidee aus Experimenten gefolgert worden.

**4.19 Bemerkung**

Mit einer ähnlichen Idee lassen sich Multigraphen mit fest vorgegebenen Gradfolgen erzeugen (zufälliges Matching).

Zum Abschluss betrachten wir noch ein Modell für ungerichtete Multigraphen, das durch die folgenden Standardvermutungen über Personenbekanntschaften motiviert ist:

- Jede Person hat eine beschränkte Anzahl von Bekannten  
(*kleiner durchschnittlicher Knotengrad ohne Ausreißer*)
- Viele Bekannte kennen sich untereinander  
(*lokal dichte Teilgraphen*)
- Man ist mit den meisten anderen Personen über eine kurze Kette von Bekannten verbunden  
(*kleiner durchschnittlicher Abstand, „Die Welt ist ein Dorf!“*)

Im folgenden sei  $\mathcal{G}(n)$  die Menge der ungerichteten Multigraphen mit Knotenmenge  $V_n = \{0, \dots, n-1\}$ .

Während die erste Standardvermutung mit dem mittleren Knotengrad bereits formalisiert ist, können für die anderen beiden aus den bereits definierten Größen Dichte und Distanz passende neue abgeleitet werden.

#### 4.20 Definition (Cluster-Koeffizient)

Sei  $G = (V, E)$  ein Multigraph und  $G[V'] \subseteq G$  der von einer Teilmenge  $V' \subseteq V$  induzierte Teilgraph. Die Dichte

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

der (echten) Nachbarschaft eines Knotens  $v \in V$  heißt Cluster-Koeffizient  $C(v)$  von  $v$ . Die Cluster-Koeffizient  $C(G)$  von  $G$  ist der mittlere Cluster-Koeffizient seiner Knoten,

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

#### 4.21 Definition (Charakteristische Weglänge)

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

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

Graphen mit zwei der drei Eigenschaften lassen sich leicht konstruieren.

**4.22 Definition (Torus)**

Seien  $d, k \in \mathbb{N}$ ,  $V_k = \{0, \dots, k - 1\}$  und  $V_k^d = \underbrace{V_k \times \dots \times V_k}_{d \text{ mal}}$ . Der Graph

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

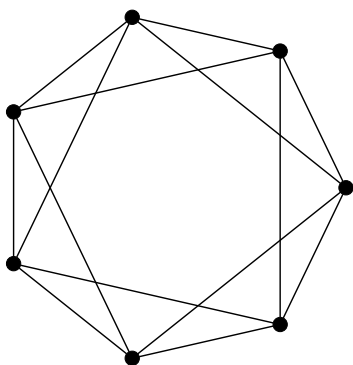
$$\{(v_1, \dots, v_d), (w_1, \dots, w_d)\} \in E \iff 1 \leq \sum_{i=1}^d \min\{v_i - w_i, w_i - v_i \pmod k\} \leq r$$

heißt  $(k, d, r)$ -Torus. Wir identifizieren ihn mit dem durch die Bijektion

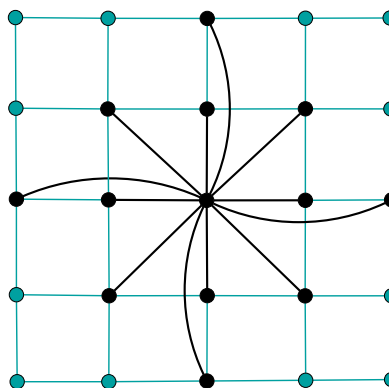
$$(v_1, \dots, v_d) \leftrightarrow \sum_{i=1}^d v_i \cdot k^{i-1}$$

bestimmten isomorphen Graphen in  $\mathcal{G}(k^d)$ .

**4.23 Beispiel**



$(7, 1, 2)$ -Torus



Nachbarn eines Knotens in  $(k, 2, 2)$ -Torus mit  $k \geq 5$

Wir analysieren nur eindimensionale Tori. Für höhere Dimensionen ergeben sich analoge Resultate.

#### 4.24 Lemma

In einem  $(n, 1, r)$ -Torus  $G = (V, E)$  mit  $r \leq \frac{n-1}{2}$  gilt

$$(i) \quad d(G) = 2r$$

$$(ii) \quad C(G) = \frac{3(r-1)}{4(r-\frac{1}{2})}$$

$$(iii) \quad L(G) \geq \lfloor \frac{n-1}{4r} \rfloor$$

■ **Beweis:** Zwei Knoten  $v \neq w \in V_n$  sind genau dann adjazent, wenn  $w \equiv i \pmod{n}$  für ein  $i \in \{v-r, \dots, v-1, v, \dots, v+r\}$ . Jeder Knoten hat damit genau  $2r$  Nachbarn.

Die  $2r$  Nachbarn eines Knotens  $v$  bilden  $\frac{2r(2r-1)}{2}$  Paare. Betrachte den Knoten  $n$  mit Nachbarschaft  $N(n) = \{n-r, \dots, n-1, 1, \dots, r\}$ . Die Knoten in  $\{n-r, \dots, n-1\}$  und  $\{1, \dots, r\}$  sind jeweils paarweise adjazent, und jeder Knoten  $i \in \{1, \dots, r\}$  hat  $r-i$  weitere Nachbarn in  $\{n-r, \dots, n-1\}$ . Insgesamt ergibt sich daher

$$\begin{aligned} C(n) &= \frac{2 \cdot \frac{r(r-1)}{2} + \sum_{i=1}^r (r-i)}{\frac{2r(2r-1)}{2}} = \frac{r(r-1) + \sum_{i=1}^{r-1} i}{r(r-1) + r^2} \\ &= \frac{r(r-1) + \frac{(r-1)r}{2}}{r(r-1) + r(r-1) \cdot \frac{r}{r-1}} = \frac{1 + \frac{1}{2}}{1 + \frac{r}{r-1}} = \frac{\frac{3}{2}}{\frac{2r-1}{r-1}} \\ &= \frac{3(r-1)}{4(r-\frac{1}{2})} \end{aligned}$$

und aus Symmetriegründen  $C(G) = C(n)$ .

Schließlich beobachten wir leicht, dass der Knoten  $n$  jeweils  $2r$  Knoten mit Abstand  $1, 2, \dots, \lfloor \frac{n-1}{2r} \rfloor$  (und höchstens  $2r$  Knoten mit um 1 größerem Abstand). Die Summe der Abstände zu den  $n-1$  anderen Knoten ist daher mindestens

$$2r \cdot \sum_{i=1}^{\lfloor \frac{n-1}{2r} \rfloor} i = 2r \cdot \frac{\lfloor \frac{n-1}{2r} \rfloor (\lfloor \frac{n-1}{2r} \rfloor + 1)}{2}$$

Dies gilt analog für alle Knoten, sodass

$$L(G) \geq \frac{r}{n-1} \left\lfloor \frac{n-1}{2r} \right\rfloor^2 + \frac{r}{n-1} \left\lfloor \frac{n-1}{2r} \right\rfloor \geq \left\lfloor \frac{n-1}{4r} \right\rfloor.$$

□

Für festes  $r \ll n$  hat ein  $(n, 1, r)$ -Torus also konstanten Knotengrad und einen konstant hohen Cluster-Koeffizienten. Allerdings ist die charakteristische Weglänge linear in  $n$ . Die Idee des nun folgenden Graphenmodells ist, durch Umleitung einiger Kanten die erwarteten Abstände zu verkürzen, um so auch der dritten Eigenschaft zu entsprechen.

#### 4.25 Definition (Kleine Welt; Strogatz und Watts 1998)

Zu  $k, d, r \in \mathbb{N}$  und  $0 \leq p \leq 1$  heißt das Graphenmodell  $\mathcal{S}(k, d, r, p) = (\mathcal{G}(k^d), P)$  kleine Welt der Dimension  $d$ , falls  $P$  die Wahrscheinlichkeit angibt, dass der Graph aus einem  $(k, d, r)$ -Torus entstanden ist, indem jede Kante unabhängig mit Wahrscheinlichkeit  $p$  genau einen neuen Endknoten erhalten hat.

Wir beschränken uns weiterhin auf Dimension 1. Der Erzeugungsalgorithmus durchläuft die Kanten eines  $(n, 1, r)$ -Torus, fügt sie aber nur mit Wahrscheinlichkeit  $1 - p$  in den Graphen ein. Mit Wahrscheinlichkeit  $p$  wird eine neue Kante von einem der beiden Endknoten zu einem zufälligen anderen eingefügt. Dabei können Mehrfachkanten entstehen.

---

#### Algorithmus 24: Erzeugen kleiner Welten der Dimension 1

---

**Eingabe:** Knotenzahl  $n \in \mathbb{N}$

Nachbarschaftsparameter  $r \in \mathbb{N}$ ,  $r \leq \frac{n-1}{2}$

Änderungswahrscheinlichkeit  $0 < p < 1$

**Ausgabe:** Graph  $G = (\{0, \dots, n-1\}, E)$

$E \leftarrow \emptyset$

**foreach**  $v = 0, \dots, n-1$  **do**

**foreach**  $w = v+1, \dots, v+r$  **do**

$w \leftarrow w \bmod n$

$R \leftarrow R[0, 1]$

**if**  $\frac{p}{2} \leq R \leq 1 - \frac{p}{2}$  **then**

$E \leftarrow E \cup \{\{v, w\}\}$

**else**

**if**  $R < \frac{p}{2}$  **then**

$E \leftarrow E \cup \{\{v, v + \lceil \frac{2R(n-1)}{p} \rceil \bmod n\}\}$

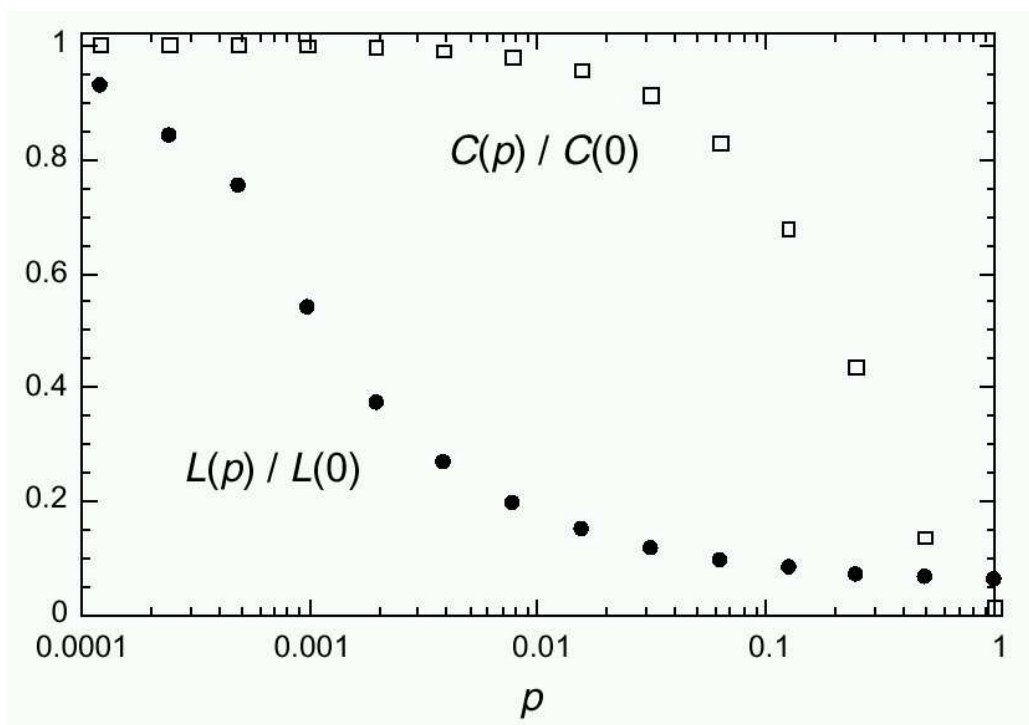
**else**

$E \leftarrow E \cup \{\{w, w + \lceil \frac{2(1-R)(n-1)}{p} \rceil \bmod n\}\}$

**4.26 Satz**

Der Algorithmus erzeugt eine kleine Welt  $G = (V, E)$  in  $\mathcal{O}(n + m)$  Zeit und Platz und benötigt dafür genau  $m$  Zufallszahlen.

Die folgende Grafik aus dem Artikel von Strogatz und Watts (1998) zeigt den Cluster-Koeffizienten und die charakteristische Weglänge in Abhängigkeit von der Änderungswahrscheinlichkeit ( $n = 1000$  und  $r = 5$ ).



Insbesondere ist für  $p \approx \frac{1}{50}$  der Cluster-Koeffizient noch sehr hoch, die charakteristische Weglänge aber stark gefallen.

Diese Ergebnisse sind experimentell. Lassen sie sich analytisch bestätigen? Wie ist das asymptotische Verhalten? Gibt es für ein  $p = p(n)$  einen optimalen Kompromiss?