

**Ein Prinzip moderater Abweichungen für die Größe  
der größten Komponente in einem  
Erdős-Rényi-Zufallsgraphen im superkritischen Fall.**

Inauguraldissertation zur  
Erlangung des akademischen Grades  
eines Doktors der Naturwissenschaften  
durch den Fachbereich Mathematik und Informatik  
der Westfälischen Wilhelms-Universität Münster

vorgelegt von  
**Jens Ameskamp**  
aus Münster

– 2010 –

Dekan: Prof. Dr. Dr. h.c. Joachim Cuntz  
Erster Gutachter: Prof. Dr. Matthias Löwe  
Zweite Gutachterin: Prof. Dr. Nina Gantert

Tag der mündlichen Prüfung: 18. Juni 2010  
Tag der Promotion: 18. Juni 2010

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Vorbereitungen</b>	<b>7</b>
2.1	Graphen . . . . .	7
2.1.1	Ein Modell für Graphen . . . . .	7
2.1.2	Zufällige Graphen . . . . .	9
2.1.3	Das Dualitätsprinzip . . . . .	10
2.1.4	Der Explorationsprozess . . . . .	11
2.2	Galton-Watson-Prozesse . . . . .	15
2.2.1	Galton-Watson-Prozesse mit poissonverteilten Zuwächsen	22
2.3	Tailabschätzungen der Größe der ersten Komponente . . . . .	29
2.3.1	Der superkritische Fall: $\lambda > 1$ . . . . .	30
2.3.2	Der subkritische Fall: $\lambda < 1$ . . . . .	31
2.3.3	Der kritische Fall: $\lambda = 1$ . . . . .	35
<b>3</b>	<b>Das Erscheinen der größten Komponente</b>	<b>47</b>
3.1	Der subkritische Fall: $\lambda < 1$ . . . . .	48
3.2	Der superkritische Fall: $\lambda > 1$ . . . . .	53
3.2.1	Die Größe der zweitgrößten Komponente . . . . .	61
3.3	Der kritische Fall: $\lambda = 1$ . . . . .	63
<b>4</b>	<b>Ein Prinzip moderater Abweichungen</b>	<b>67</b>
4.1	Ein Prinzip moderater Abweichungen für Binomialverteilungen .	68

4.2	Weitere Untersuchungen des Explorationsprozesses $S_t$ . . . . .	71
4.2.1	Die Konvergenzgeschwindigkeit der Folge $(1 + \frac{x}{n})^n$ . . . . .	72
4.2.2	Asymptotische Momente von $S_{\lfloor \beta n \rfloor}$ . . . . .	73
4.3	Das Prinzip moderater Abweichungen für $ C_{\leq k_n} $ . . . . .	75
4.4	Das Prinzip moderater Abweichungen für $ C_{\max} $ . . . . .	83
4.5	Übertragung auf $P_{n,M}$ . . . . .	87
<b>5</b>	<b>Ein Zentraler Grenzwertsatz</b>	<b>91</b>
	<b>Literaturverzeichnis</b>	<b>97</b>

# 1 Einleitung

Die Grundlagen zur Theorie zufälliger Graphen legte Paul Erdős mit seinen Arbeiten [Erd47] und [Erd57]. In diesen Arbeiten wurde diese Theorie zunächst nur zum Beweis der Existenz von Graphen mit bestimmten Ramsey-Eigenschaften verwendet. Angestoßen von weiterer Forschung von Erdős zusammen mit Alfréd Rényi (z. B. [ER59], [ER60]) hat sie sich seitdem zu einem eigenständigen und aktiven Forschungsfeld entwickelt. Einen detaillierten Überblick über das Gebiet bieten [Bol85] und, darauf aufbauend, [JLR00].

Zwei Modelle für eine zufallsabhängige Wahl eines Graphen auf einer festen Anzahl  $n$  von Knoten haben in der Literatur besondere Aufmerksamkeit erfahren. Unter der Bezeichnung  $G(n, M)$  führten Erdős und Rényi in [ER59] ein Modell zufälliger Graphen ein, in dem ein Graph gleichverteilt aus der Menge der Graphen auf  $n$  Knoten mit genau  $M$  Kanten gewählt wird. Die vorliegende Arbeit beschäftigt sich hauptsächlich mit dem in [Gil59] zuerst untersuchten Modell  $G(n, p)$ , bei dem jede Kante unabhängig von den anderen Kanten mit Wahrscheinlichkeit  $p$  realisiert wird.

In [ER60] beobachteten Erdős und Rényi, dass die Eintrittswahrscheinlichkeit monotoner Ereignisse mit wachsender Größe des Graphen entweder verschwindet oder gegen 1 strebt und zwar abhängig davon, ob der dem Modell zu Grunde liegende Parameter unterhalb oder oberhalb eines bestimmten Schwellenwertes liegt. So geht zum Beispiel die Wahrscheinlichkeit dafür, dass ein gemäß  $G(n, M)$

gewählter Graph verbunden ist, gegen 1, wenn  $M$  schneller als linear mit  $n$  wächst, ansonsten gegen 0 (vgl. [Bol85]). Ein solcher Phasenübergang genanntes Verhalten stellten die Autoren in [ER60] und [ER61] bei der Untersuchung der Größe der größten Komponente in einem gemäß  $G(n, M)$  gewählten Graphen ebenfalls fest: Wächst  $M$  mit  $n$  wesentlich langsamer als  $\frac{n}{2}$ , so ist die größte Komponente für große  $n$  von der Ordnung  $\ln n$ , falls aber  $M$  wesentlich schneller mit  $n$  wächst als  $\frac{n}{2}$ , so nimmt sie einen signifikanten Teil des Graphen ein. Diese Arbeit präsentiert in Kapitel 3 ähnliche Ergebnisse über  $G(n, p)$ , um die mathematischen Grundlagen und eine thematische Einordnung des Prinzips moderater Abweichungen für die Größe der größten Komponente zu schaffen. Als Grundlage für viele der vorgestellten Aussagen diente [Hof09].

Anschließend an diese Einleitung widmet sich **Kapitel 2** dieser Arbeit der Darstellung der benötigten mathematischen Grundlagen. Zunächst werden in Abschnitt 2.1 beide oben vorgestellten Modelle zufälliger Graphen mathematisch präzisiert und ein einfacher Mechanismus zur Erforschung von Teilen eines Graphen vorgestellt. Dieser *Explorationsprozess* genannte Mechanismus arbeitet wie folgt: Startend in einer beliebigen Anzahl Knoten werden schrittweise alle mit diesen Knoten verbundenen Knoten erforscht, was dann iterativ mit den neuen Knoten wiederholt wird. Als Resultat dieses Prozesses erhält man den kleinsten Teilgraphen, der die gewählten Startknoten enthält. Kenntnisse über einen solchen Teilgraphen bieten eine Möglichkeit, Aussagen über die größte Komponente des Graphen zu gewinnen. Insbesondere im sogenannten superkritischen Parameterbereich, in dem man viele Knoten in der größten Komponente erwartet, ist es wahrscheinlich, dass die größte Komponente Teil des auf diese Weise erforschten Teilgraphen ist, wenn nur die Anzahl der Startknoten hinreichend groß ist. Umgekehrt wird sich in Abschnitt 3.2.1 zeigen, dass die kleineren Komponenten im Gegensatz zur größten Komponente nur von der Größenordnung  $\ln n$  sind, so dass sich der Fehler, wenn man statt der größten Komponente

---

diesen Teilgraphen betrachtet, leicht kontrollieren lässt.

Diese Zusammenhänge rechtfertigen eine genauere Untersuchung des Explorationsprozesses. Legt man das Modell  $G(n, p)$  zu Grunde, so lässt sich die Verteilung des Explorationsprozesses zu jedem Zeitpunkt exakt bestimmen (s. Abschnitt 2.1.4), was im Verlauf dieser Arbeit weitreichende Schlüsse ermöglicht.

Eine weitere Beobachtung, die zu einem besseren Verständnis des Explorationsprozesses führt, ist die Ähnlichkeit des Prozesses mit einem klassischen Galton-Watson-Prozess. Damit lassen sich einige aus der Literatur bekannte Resultate über Galton-Watson-Prozesse zu Aussagen über den Explorationsprozess übertragen. Deswegen wird in Abschnitt 2.2 eine Einführung in die Theorie der Galton-Watson-Prozesse auf Basis von [Als02] gegeben und die Verbindung zum Explorationsprozess hergestellt.

Das Kapitel 2 schließt mit einer Zusammenfassung von Ergebnissen über die Tails der Verteilung der Größe der ersten Komponente eines gemäß  $G(n, p)$  verteilten Graphen, die im folgenden Kapitel Verwendung finden.

In **Kapitel 3** wird an Hand schwacher Gesetze der großen Zahlen für die Größe der größten Komponente der oben angesprochene Phasenübergang im Modell  $G(n, p)$  dargestellt und zugleich die Grundlage für wichtige Beweisschritte in den Kapiteln 4 und 5 gelegt.

Um einen Zugang zur Verteilung der größten Komponente eines Graphen zu bekommen, untersucht man hier die Anzahl  $Z_{>k}$  der Knoten in Komponenten mit mehr als  $k$  Knoten. Diese Zufallsgröße ist eng mit der Größe der größten Komponente verwandt, da für jedes  $k$  das Ereignis  $|C_{\max}| > k$  bereits impliziert, dass  $Z_{>k} > k$ . Umgekehrt folgt natürlich aus  $Z_{>k} > k$  auch  $|C_{\max}| > k$ . So ist es

natürlich, für den Beweis der schwachen Gesetze in diesem Kapitel,

$$E_\lambda Z_{>k} = nP_\lambda(|C_{(1)}| > k)$$

zu betrachten, was die Untersuchung dieser Tails in Abschnitt 2.3 rechtfertigt.

Das Resultat unterscheidet sich je nach Verhalten des Modellparameters  $p$  für wachsendes  $n$  erheblich. Der in diesem Modell kritische Wert von  $p$  ist  $\frac{1}{n}$ . Fällt  $p$  mit  $n$  deutlich schneller als  $\frac{1}{n}$ , also  $pn \rightarrow \lambda < 1$ , so ist die größte Komponente von der Ordnung  $\ln n$ . Ist hingegen  $pn \rightarrow \lambda > 1$ , so ist sie von der Ordnung  $n$ . Abschnitt 3.3 beschäftigt sich schließlich mit dem Fall  $pn \rightarrow 1$ . In diesem Fall ist kein triviales schwaches Gesetz der großen Zahlen mehr wahr (vgl. [Ald97]), aber man kann mit den selben Mitteln, die zu den schwachen Gesetzen in den anderen Fällen geführt haben, zeigen, dass sich die Größe der größten Komponente mit positiver Wahrscheinlichkeit im Bereich  $n^{\frac{2}{3}}$  bewegt.

**Kapitel 4** beschäftigt sich mit dem Hauptthema dieser Arbeit, einem Beweis für ein Prinzip moderater Abweichungen für die Größe der größten Komponente eines zufälligen Graphen. Um den Begriff eines Prinzips moderater Abweichungen erläutern zu können, ist der Begriff eines Prinzips großer Abweichungen von Nöten.

Eine Folge  $(\mu_n)_{n \in \mathbb{N}}$  von Maßen auf  $(\mathbb{R}, \mathfrak{B})$  genügt einem Prinzip großer Abweichungen, wenn es eine Skala  $(a_n)_{n \in \mathbb{N}}$  und eine Ratenfunktion  $I$  gibt, so dass

$$\liminf_{n \in \mathbb{N}} \frac{1}{a_n} \ln \mu_n(U) \geq - \inf_{x \in U} I(x)$$

und

$$\limsup_{n \in \mathbb{N}} \frac{1}{a_n} \ln \mu_n(A) \leq - \inf_{x \in A} I(x)$$

für alle offenen Mengen  $U$  und abgeschlossenen Mengen  $A$  (vgl. [Ell85]). Oft findet man ein Prinzip großer Abweichungen auf der Skala der für das Modell geltenden



---

Gesetze der großen Zahlen. Im Fall von unabhängigen, identisch verteilten reellwertigen Zufallsgrößen  $X_n$  zeigt [Var85] ein Prinzip großer Abweichungen auf der Skala  $n$  für die normierten Partialsummen

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Auch die in dieser Arbeit untersuchten Größe der größten Komponente eines zufälligen Graphen genügt im kritischen Fall, normiert mit  $n^{-1}$ , einem Prinzip großer Abweichungen mit Skala  $n$  (vgl. [O'C98]).

Man spricht dann von einem Prinzip moderater Abweichungen, wenn die gewählte Skala zwischen der eines Gesetzes der großen Zahlen und der eines zentralen Grenzwertsatzes für das untersuchte Modell liegt. Die Arbeit [EL03] behandelt ein Prinzip moderater Abweichungen für unabhängige, zentrierte, identisch verteilte Zufallsgrößen. Dort findet sich, dass für  $\alpha \in (\frac{1}{2}, 1)$  die mit  $n^{-\alpha}$  skalierten Partialsummen

$$Y_n = \frac{1}{n^\alpha} \sum_{i=1}^n X_i$$

einem Prinzip moderater Abweichungen auf der Skala  $n^{2\alpha-1}$  genügen. Ist  $\sigma^2 \in \mathbb{R}^+$  so gewählt, dass er Zentrale Grenzwertsatz für dieses Modell (z. B. [Als00]) den Schluss

$$\frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^n X_i \rightarrow \mathfrak{N}(0, \sigma^2) \quad \text{in Verteilung}$$

erlaubt, so ist die Ratenfunktion im Prinzip moderater Abweichungen schon durch

$$I(x) = \frac{x^2}{2\sigma^2}$$

gegeben. Diese Ratenfunktion hängt hier also nicht wie im Satz von Cramér (vgl. [Ell85]) von der genauen Verteilung der betrachteten Zufallsgrößen ab, was für ein Prinzip moderater Abweichungen typisch ist. Die Größe der größten Komponente eines zufälligen Graphen zeigt ein analoges Verhalten, wie in Kapitel 4 gezeigt wird.

Kernidee des Beweises in Kapitel 4 ist es, an Stelle der schwerer zu fassenden größten Komponente des Graphen den von den ersten  $k_n$  Knoten aufgespannten Teilgraphen zu untersuchen. Dieser ist mit den in Kapitel 2 vorgestellten Methoden besser zugänglich. Wenn  $k_n$  dann passend gewählt wird, kann sichergestellt werden, dass dieser Teilgraph sich nicht zu sehr von der größten Komponente unterscheidet, so dass das zunächst für die Größe dieses Teilgraphen gezeigte Prinzip moderater Abweichungen auf die Größe der größten Komponente übertragbar ist.

Anschließend wird in Abschnitt 4.5 dieses Ergebnis, welches für das Modell  $G(n, p)$  hergeleitet wurde, noch mit Hilfe der in Abschnitt 2.1.2 vorgestellten Zusammenhänge zwischen den Modellen auf  $G(n, M)$  übertragen.

Ein Zentraler Grenzwertsatz für die Größe der größten Komponente wurde von Pittel [Pit90] gefunden. Alternative Beweise entstanden unter anderem in [ML98] und [BBV00]. Da der Beweis den Prinzip moderater Abweichungen in dieser Arbeit von dem elementaren Beweis des Zentralen Grenzwertsatzes aus [Hof09] inspiriert wurde, wird dieser schließlich in **Kapitel 5** vorgestellt.

## 2 Vorbereitungen

In diesem Kapitel wird zunächst das Konzept eines Graphen mathematisch modelliert. Anschließend werden zwei Modelle für die zufällige Wahl eines Graphen mit einer festen Anzahl von Knoten eingeführt. Das wichtigste Hilfsmittel zur Erforschung der Größe einer Komponente eines Graphen, der Explorationsprozess, sowie seine Vergleichbarkeit mit klassischen Galton-Watson-Verzweigungsprozessen sind Thema der folgenden Abschnitte.

### 2.1 Graphen

#### 2.1.1 Ein Modell für Graphen

Wie üblich (vgl. [Die06]) wird ein Graph  $G$  durch seine Knoten  $V(G)$  und seine Kanten  $E(G)$  repräsentiert. In dieser Arbeit werden ausschließlich endliche, ungerichtete Graphen betrachtet. Außerdem ist hier eine spezielle Bezeichnung der Knoten nicht von Bedeutung, so dass sich das vorgestellte Modell auf  $V(G) \subset N$  beschränkt.

**Definition 2.1.** Sei  $n \in N$ . Ein Tupel  $G = (V(G), E(G))$  mit Knoten  $V(G) = \{1, \dots, n\}$  und Kanten  $E(G) \subset \{\{i, j\} : i, j \in V(G), i \neq j\}$  heißt *Graph  $G$  auf  $n$  Knoten*. Zwei Knoten  $i, j \in V(G)$  heißen *verbunden* – man schreibt  $i \leftrightarrow j$  – wenn  $\{i, j\} \in E(G)$ .

Ausgehend von dieser Struktur lassen sich Konzepte, die in den Gebieten, die Graphen zur Modellierung benutzen, von Bedeutung sind, wiederfinden. Im Rahmen von Netzwerken zum Beispiel ist es weniger wichtig, ob zwei Knoten direkt miteinander verbunden sind, als ob sie, notfalls über Umwege, kommunizieren können.

**Definition 2.2.** Zwei Knoten  $i, j$  eines Graphen  $G$  *kommunizieren*, geschrieben  $i \leftrightarrow j$ , genau dann, wenn es ein  $k \in \mathbb{N}$  und Knoten  $i_1, \dots, i_k \in V(G)$  gibt, mit  $i_1 = i, i_k = j$  und  $i_m \leftrightarrow i_{m+1}$  für alle  $1 \leq m < k$ . Der Graph  $G$  heißt *verbunden*, wenn je zwei Knoten von  $G$  kommunizieren.

Die in dieser Arbeit behandelten Graphen werden im Allgemeinen nicht verbunden sein, zerfallen also in voneinander getrennte Bereiche, innerhalb derer alle Knoten miteinander kommunizieren.

**Definition 2.3.** Ein Teilgraph  $G'$  von  $G$  ist ein Graph, für den  $V(G') \subset V(G)$  und  $E(G') \subset \{\{i, j\} \in E(G) : i, j \in V(G')\}$ .

Der Schwerpunkt in dieser Arbeit wird auf die Größe solcher Teilgraphen gelegt und nicht auf ihre Struktur, deshalb wird im Folgenden immer  $E(G') = \{\{i, j\} \in E(G) \mid i, j \in V(G')\}$  angenommen.

**Definition 2.4.** Eine *Komponente*  $C$  eines Graphen  $G$  ist ein verbundener Teilgraph von  $G$ , so dass für alle Knoten  $i \in V(C)$  und  $j \in V(G) \setminus V(C)$  gilt:  $i \not\leftrightarrow j$ .

**Bemerkung 2.5.** Da Kommunikation wie in Definition 2.2 definiert eine Äquivalenzrelation auf der Menge  $V(G)$  darstellt, und die Komponenten des Graphen gerade die Äquivalenzklassen bezüglich dieser Relation darstellen, zerfällt jeder Graph eindeutig in untereinander nicht verbundene Komponenten. Die größte Komponente des Graphen bezeichnet man als  $C_{\max}(G)$ .

**Definition 2.6.** Für einen Knoten  $i \in V(G)$  bezeichnet  $C_{(i)}(G)$  die Komponente mit  $i \in C_{(i)}(G)$  und für  $k \in \{1, \dots, n\}$  ist  $C_{\leq k}(G) = \cup_{i=1}^k C_{(i)}(G)$  der von den ersten  $k$  Knoten aufgespannte Teilgraph.

Der Übersichtlichkeit halber wird im Folgenden bei allen hier eingeführten Größen der zu Grunde liegende Graph, falls keine Verwechslungsgefahr besteht, nicht genannt.

### 2.1.2 Zufällige Graphen

Es gibt in der Literatur mehrere Ansätze zur Einführung einer Wahrscheinlichkeitsverteilung auf der Menge aller Graphen auf  $n$  Knoten. Die beiden am häufigsten verwendeten werden hier vorgestellt.

Ein Ansatz, eingeführt von Erdős und Rényi in [ER59] als  $G(n, M)$ , ist es, für eine gegebene Anzahl  $M \in N$  von Kanten einen Graphen nach der Gleichverteilung auf der Menge aller Graphen mit genau  $M$  Kanten zu wählen. Nennt man dieses Wahrscheinlichkeitsmaß  $P_{n,M}$ , so ist

$$P_{n,M}(\{G\}) = \begin{cases} 0 & \text{für } |E(G)| \neq M \\ \left(\frac{n(n-1)}{2M}\right)^{-1} & \text{für } |E(G)| = M. \end{cases}$$

Das in dieser Arbeit hauptsächlich betrachtete stochastische Modell auf Graphen mit  $n$  Knoten wurde kurze Zeit später in [Gil59] als  $G(n, p)$  vorgestellt. Die Größe  $p \in (0, 1)$  bezeichnet die Wahrscheinlichkeit, dass eine Kante – unabhängig von den anderen – realisiert wird. In dieser Arbeit wird nur das Verhalten des Graphen für  $n \rightarrow \infty$  und  $p \in \mathcal{O}(n^{-1})$  untersucht. Sei also  $p = \frac{\lambda}{n}$  für  $\lambda > 0$ , so

bezeichne

$$P_{n,\lambda}(\{G\}) = \left(\frac{\lambda}{n}\right)^{|E(G)|} \left(1 - \frac{\lambda}{n}\right)^{\binom{n}{2} - |E(G)|}.$$

Diese beiden Modelle stehen in enger Beziehung zueinander. So ist die Anzahl der Kanten in einem gemäß  $P_{n,\lambda}$  verteilten Graphen verteilt wie  $\mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}$ , weswegen sich  $P_{n,\lambda}$  durch  $P_{n,M}$  wie folgt darstellen lässt:

$$\begin{aligned} P_{n,\lambda}(A) &= \sum_{M=0}^{\frac{n(n-1)}{2}} P_{n,\lambda}(A \mid |E(G)| = M) P_{n,\lambda}(|E(G)| = M) \\ &= \sum_{M=0}^{\frac{n(n-1)}{2}} P_{n,M}(A) \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(M). \end{aligned}$$

Diese Eigenschaft wird in Abschnitt 3.2.1 verwendet, um das Prinzip moderater Abweichungen von  $P_{n,\lambda}$  auf  $P_{n,M}$  zu übertragen.

Der Index  $n$  wird im Folgenden nur genannt, wenn ansonsten Mehrdeutigkeiten entstehen würden.

### 2.1.3 Das Dualitätsprinzip

Eine Frage, die sich dann bei der Untersuchung des Graphen stellt, ist, wie sich der Rest des Graphen verhält, wenn man eine Komponente nicht mehr berücksichtigt. Eine Antwort darauf ist für die Beweise in Abschnitt 3.2 und den Kapiteln 4 und 5 unerlässlich.

**Lemma 2.7.** *Sei  $G$  ein gemäß  $P_{n,\lambda}$  verteilter Graph und  $G \setminus C_{(1)}$  dieser Graph, aus dem die erste Komponente entfernt wurde. Dann ist für jedes  $k \in \{0, \dots, n\}$*

$$P_{n,\lambda}^{G \setminus C_{(1)} \mid |C_{(1)}|=k} = P_{n-k, \lambda \frac{n-k}{n}}.$$

**Beweis:** Ist  $|C_{(1)}| = k$ , so sind Kanten in  $G \setminus C_{(1)}$  unabhängig von der Struktur von  $C_{(1)}$  weiterhin unabhängig voneinander mit Wahrscheinlichkeit  $\frac{\lambda}{n} = \lambda \frac{n-k}{n(n-k)}$  realisiert. Da weiterhin  $n - k$  Knoten in  $G \setminus C_{(1)}$  sind, folgt die Behauptung.  $\square$

In Abschnitt 3.2 wird gezeigt, dass für  $\lambda > 1$  und wachsendes  $n$  eine Komponente eines gemäß  $P_{n,\lambda}$  verteilten Graphen eine Anzahl von Knoten der Ordnung  $n$  beinhaltet und damit wesentlich größer wird als die anderen. Eine dem obigen Lemma ähnliche Aussage erlaubt es, auch das Grenzverhalten der Größe der zweitgrößten Komponente anzugeben, was in Abschnitt 3.2.1 geschieht. Die Überlegungen, die zu dieser Aussage führen, sind in [AS00, Abschnitt 10.5] dargelegt.

**Bemerkung 2.8.** Seien  $\mu_\lambda < 1 < \lambda$  konjugiert, d. h.

$$\lambda e^{-\lambda} = \mu_\lambda e^{-\mu_\lambda}$$

und sei  $G$  ein gemäß  $P_{n,\lambda}$  verteilter Graph. Dann ist

$$P_{n,\lambda}^{G \setminus C_{\max}} = P_{(1-\zeta_\lambda)n, \mu_\lambda} + \mathcal{O}(1).$$

### 2.1.4 Der Explorationsprozess

Die Verteilung der Größe eines Teilgraphen eines zufälligen Graphen ist nicht leicht zugänglich. Deswegen benötigt man einen leichter zu fassenden Mechanismus, der Aussagen über die zu untersuchenden Zufallsgrößen gestattet. Dafür wird in diesem Abschnitt ein Prozess vorgestellt, der ausgehend von  $k$  Startknoten den von diesen Knoten aufgespannten Teilgraphen erforscht.

Zu Beginn des Prozesses, der im folgenden *Explorationsprozess* genannt wird, ist jeder der  $k$  Startknoten *aktiv*; alle anderen Knoten starten *neutral*. In jedem

Schritt des Prozesses wird ein zufälliger aktiver Knoten ausgewählt und *inaktiv* gesetzt. Zusätzlich werden alle mit diesem Knoten verbundenen neutralen Knoten aktiviert. Der Prozess stoppt, wenn kein aktiver Knoten mehr vorhanden ist. Am Ende des Prozesses sind offenbar alle Knoten von  $C_{\leq k}$  inaktiv und alle anderen Knoten neutral. Es bezeichne  $S_t$  die Anzahl der aktiven Knoten nach dem  $t$ -ten Schritt des Prozesses.

**Bemerkung 2.9.** Da in jedem Schritt ein inaktiver Knoten hinzukommt, gilt

$$|C_{\leq k}| = \min\{t \in N : S_t = 0\}.$$

**Bemerkung 2.10.** Nach Konstruktion ist klar, dass  $S_0 = k$  ist und für jedes  $t$  mit  $S_{t-1} > 0$ , wenn also noch mindestens ein aktiver Knoten vorhanden ist,

$$S_t = S_{t-1} + X_t - 1$$

gilt. Dabei ist  $X_t$  die Anzahl der neutralen Knoten ist, die im  $t$ -ten Schritt des Prozesses aktiviert werden und  $S_t = 0$ , falls  $S_{t-1} = 0$ .

**Bemerkung 2.11.** Im Rahmen der Untersuchung der Verteilung von  $S_t$  unter  $P_\lambda$  stellt man zunächst fest, dass für jedes  $t \in N$  im  $t$ -ten Schritt mit  $s$  aktiven,  $t$  inaktiven und  $n - t - s$  neutralen Knoten jeder neutrale Knoten unabhängig von den anderen mit Wahrscheinlichkeit  $\frac{\lambda}{n}$  mit dem aktuell zu erforschenden Knoten verbunden ist und somit aktiviert wird. Daraus folgt

$$P_\lambda^{X_t | S_{t-1}=s} = \mathfrak{B}_{n-(t-1)-s, \frac{\lambda}{n}}.$$

In Bemerkung 2.9 wird klar, dass vor allem der erste Zeitpunkt, an dem  $S_t = 0$  gilt, eine Rolle spielt. Um die Verteilung von  $S_t$  besser handhabbar zu machen, wird daher ein leicht abgeänderter Prozess betrachtet, für den

$$S_t = S_{t-1} + X_t - 1$$



für alle  $t \in N$  gilt. Offenbar ist dieser Prozess bis zum ersten Nulldurchlauf mit dem Originalprozess identisch, weswegen sich alle für diese Arbeit wesentlichen Aussagen übertragen lassen. Im Folgenden wird mit dem Explorationsprozess  $S_t$  dieser abgeänderte Prozess bezeichnet.

Für die Untersuchung des Explorationsprozesses werden in den folgenden Abschnitten zwei verschiedene Herangehensweisen vorgestellt. Zum einen ist es möglich, genaue Aussagen über die Verteilung des Prozesses zu machen, zum anderen lässt sich der Prozess gut mit den bekannten Galton-Watson-Prozessen vergleichen.

## Die Verteilung des Explorationsprozesses

In diesem Abschnitt wird [Hof09, Proposition 4.26] folgend die Verteilung von  $S_t$  hergeleitet.

**Lemma 2.12.** *Für alle  $t \in \{1, \dots, n\}$  ist*

$$S_t + t - k \sim \mathfrak{B}_{n-k, 1 - (1 - \frac{\lambda}{n})^t}.$$

*Ist zusätzlich  $s < t$ , so gilt*

$$P_\lambda^{S_t + (t-s) - S_s | S_s = j} = \mathfrak{B}_{n-s-j, 1 - (1 - \frac{\lambda}{n})^{t-s}}.$$

Für den Beweis wird eine elementare Eigenschaft der Binomialverteilung benötigt, um von den in Bemerkung 2.11 angesprochenen bedingten Verteilungen der Zuwächse des Prozesses auf die Verteilung des Gesamtprozesses schließen zu können.

**Lemma 2.13.** *Sei  $X \sim \mathfrak{B}_{n,p}$  und  $P(Y \in \cdot | X = x) \sim \mathfrak{B}_{x,q}$  für  $p, q \in [0, 1]$ , so gilt*

$$Y \sim \mathfrak{B}_{n,pq}.$$

**Beweis:** Es wird gezeigt, dass die Zahldichte von  $Y$  die geforderte Form hat.

Sei dazu  $y \in \{0, \dots, n\}$ . Dann ist

$$\begin{aligned}
 P(Y = y) &= \sum_{x=y}^n P(Y = y|X = x)P(X = x) \\
 &= \sum_{x=y}^n \binom{x}{y} q^y (1 - q)^{x-y} \binom{n}{x} p^x (1 - p)^{n-x} \\
 &= (pq)^y (1 - p)^{n-y} \sum_{x=0}^{n-y} \binom{n}{x+y} \binom{x+y}{y} \left(\frac{p(1-q)}{1-p}\right)^x \\
 &= \binom{n}{y} (pq)^y (1 - p)^{n-y} \sum_{x=0}^{n-y} \binom{n-y}{x} \left(\frac{p(1-q)}{1-p}\right)^x \\
 &= \binom{n}{y} (pq)^y (1 - p)^{n-y} \left(1 + \frac{p(1-q)}{1-p}\right)^{n-y} \\
 &= \binom{n}{y} (pq)^y (1 - pq)^{n-y},
 \end{aligned}$$

was zu beweisen war. □

**Beweis von Lemma 2.12:** Es ist bersichtlicher, fur den Beweis die Anzahl  $N_t$  der neutralen Knoten zu betrachten. Offenbar gilt

$$\begin{aligned}
 N_t &= n - t - S_t \\
 &= n - t - S_{t-1} - X_t + 1 \\
 &= N_{t-1} - X_t
 \end{aligned}$$

und

$$\begin{aligned}
 P_\lambda^{X_t|N_{t-1}=m} &= P_\lambda^{X_t|S_{t-1}=n-(t-1)-m} \\
 &= \mathfrak{B}_{n-(t-1)-(n-(t-1)-m), \frac{\lambda}{n}} \\
 &= \mathfrak{B}_{m, \frac{\lambda}{n}}.
 \end{aligned}$$

Damit ist

$$\begin{aligned}
 P_\lambda(N_t = x | N_{t-1} = y) &= P_\lambda(y - X_t = x | N_{t-1} = y) \\
 &= P_\lambda(X_t = y - x | N_{t-1} = y) \\
 &= \mathfrak{B}_{y, \frac{\lambda}{n}}(\{y - x\}) \\
 &= \mathfrak{B}_{y, 1 - \frac{\lambda}{n}}(\{x\}).
 \end{aligned}$$

Mit Hilfe von Lemma 2.13 lässt sich nun rekursiv die Verteilung von  $N_t$  bestimmen. Für  $s < t$  ist demnach

$$P_\lambda(N_t = i | N_s = m) = \mathfrak{B}_{m, (1 - \frac{\lambda}{n})^{t-s}}(\{i\}),$$

woraus die zweite zu beweisende Aussage folgt:

$$\begin{aligned}
 &P_\lambda(S_t + (t - s) - S_s = i | S_s = j) \\
 &= P_\lambda(n - t - N_t + (t - s) - (n - s - N_s) = i | n - s - N_s = j) \\
 &= P_\lambda(N_t = N_s - i | N_s = n - s - j) \\
 &= \mathfrak{B}_{n-s-j, (1 - \frac{\lambda}{n})^{t-s}}(\{n - s - i - j\}) \\
 &= \mathfrak{B}_{n-s-j, 1 - (1 - \frac{\lambda}{n})^{t-s}}(\{i\}).
 \end{aligned}$$

Die erste Aussage folgt aus  $S_0 = k$  mit  $s = 0$ . □

## 2.2 Galton-Watson-Prozesse

Betrachtet man die Verteilung der Zuwächse des Explorationsprozesses

$$P_\lambda^{X_t | S_{t-1}=s} = \mathfrak{B}_{n-(t-1)-s, \frac{\lambda}{n}},$$

so fällt auf, dass sich diese insbesondere für kleine  $s$  und  $t$  durch  $\mathfrak{B}_{n, \frac{\lambda}{n}}$  approximieren ließe. Auch wenn  $s$  und  $t$  nicht klein gegenüber  $n$  sind, so gilt doch immer

$$\mathfrak{B}_{n-(t-1)-s, \frac{\lambda}{n}} \preceq \mathfrak{B}_{n, \frac{\lambda}{n}},$$

so dass sich in jedem Fall ein sinnvoller Vergleich ergibt:

$$\begin{aligned} P_\lambda^{X_t} &= \sum_{s \in N_0} P_\lambda^{X_t | S_{t-1}=s} P(S_{t-1} = s) \\ &\preceq \sum_{s \in N_0} \mathfrak{B}_{n, \frac{\lambda}{n}} P(S_{t-1} = s) \\ &= \mathfrak{B}_{n, \frac{\lambda}{n}}. \end{aligned}$$

Insbesondere ist  $\mathfrak{B}_{n, \frac{\lambda}{n}}$  unabhängig von  $s$  und  $t$ , so dass diese Überlegungen einen Vergleich des Explorationsprozesses mit Galton-Watson-Prozessen mit binomiellen Zuwächsen nahelegen. In diesem Abschnitt werden die verwendeten Schreibweisen und Resultate vorgestellt.

**Bemerkung 2.14.** Die übliche Definition eines Galton-Watson-Prozesses (z. B. [Als00]) modelliert die Situation, dass im jedem Schritt des Prozesses alle Individuen unabhängig voneinander gemäß einer gemeinsamen Verteilung Nachkommen bekommen und dann sterben. Der Wert des Prozesses zur Zeit  $t$  stellt damit die Größe der  $t$ -ten Generation dar. Sei  $k \in N$  und  $\nu$  ein Maß auf  $N_0$ , so ist der Galton-Watson-Prozess  $Z'$  mit Start in  $k$  und Reproduktionsverteilung  $\nu$  gegeben durch:

$$\begin{aligned} Z'_0 &= k \\ Z'_{t+1} &= \sum_{i=1}^{Z'_t} X_{t,i} \end{aligned}$$

wobei  $(X_{t,i})_{(t,i) \in N^2}$  eine Familie unabhängiger, gemäß  $\nu$  verteilter Zufallsgrößen ist. In dieser Schreibweise ist es leicht zu sehen, dass die Nachkommenschaften jedes Individuums der ersten Generation unabhängig sind und wieder einen Galton-Watson-Prozess bilden, was im Beweis von Lemma 2.25 ausgenutzt wird.

Der so definierte Galton-Watson-Prozess ist dem Explorationsprozess aus Abschnitt 2.1.4 wegen der Generationenstruktur nicht sehr ähnlich. Deswegen wird im Folgenden unter dem Begriff Galton-Watson-Prozess meist ein anderer Prozess verstanden, in dem in jedem Schritt nur ein einzelnes Individuum Nachkommen bekommt und stirbt.

**Definition 2.15.** Ein Prozess  $(Z_t)_{t \in \mathbb{N}}$  wird *Galton-Watson-Prozess mit Start in  $k$  und Reproduktionsverteilung  $\nu$*  genannt, wenn

$$Z_0 = k$$

$$Z_{t+1} = \begin{cases} Z_t + X_t - 1 & \text{für } Z_t > 0 \\ 0 & \text{für } Z_t = 0 \end{cases}$$

gilt für  $k \in \mathbb{N}$ ,  $\nu(N_0) = 1$  und eine Familie von unabhängigen, gemäß  $\nu$  verteilten Zufallsgrößen  $(X_t)_{t \in \mathbb{N}}$ . Die Größe  $X_t$  wird als die Anzahl der Nachkommen des  $t$ -ten Individuums interpretiert und auch als *Zuwachs* des Prozesses zum Zeitpunkt  $t$  bezeichnet.

In dieser Definition wird durch einen Vergleich mit Bemerkung 2.10 die Ähnlichkeit des Explorationsprozesses mit einem geeigneten Galton-Watson-Prozess deutlich.

**Bemerkung 2.16.** Sei  $(S_t)_{t \in N_0}$  der Explorationsprozess mit Start in den ersten  $k$  Knoten auf einem gemäß  $P_{n,\lambda}$  verteilten Graphen und  $(Z_t)_{t \in N_0}$  ein Galton-Watson-Prozess mit Start in  $k$  und Reproduktionsverteilung  $\mathfrak{B}_{n,\frac{\lambda}{n}}$ . Da, wie in der Einleitung des Abschnittes hergeleitet, die Zuwächse von  $(S_t)_{t \in N_0}$  von den Zuwächsen von  $(Z_t)_{t \in N_0}$  stochastisch dominiert werden, gilt auch für jedes  $t \in N_0$

$$S_t \preceq Z_t.$$

Diese Ähnlichkeit überträgt sich auch auf die Größe des von  $k$  Knoten aufgespannten Teilgraphen, die mit der Gesamtpopulation des Verzweigungsprozesses

korrespondiert, was sich aus Bemerkung 2.9 zusammen mit folgender Definition ergibt:

**Definition 2.17.** Ist  $(Z_t)_{t \in N}$  eine Galton-Watson-Prozesses, so bezeichnet man

$$T = \inf\{t \in N : Z_t = 0\},$$

mit der Konvention  $\inf \emptyset = \infty$ , als *Gesamtpopulation* von  $(Z_t)_{t \in N}$ .

**Bemerkung 2.18.** Ist  $T$  die Gesamtpopulation eines Galton-Watson-Prozesses mit Reproduktionsverteilung  $\mathfrak{B}_{n, \frac{\lambda}{n}}$  und Start in  $k$  und  $C_{\leq k}$  der von den ersten  $k$  Knoten aufgespannte Teilgraph eines nach  $P_{n, \lambda}$  verteilten Graphen, so erkennt man an Definition 2.17 zusammen mit Bemerkung 2.9, dass  $|C_{\leq k}| \preceq T$ , also für alle  $m \in N$

$$P(|C_{\leq k}| > m) \leq P(T > m).$$

Einen so einfachen Vergleich kann man für eine untere Schranke nicht erwarten. Dennoch lässt sich eine nützliche Abschätzung finden, die besser wird, je kleiner die betrachteten Teilgraphen werden.

**Lemma 2.19.** Ist  $T_n$  die Gesamtpopulation eines Galton-Watson-Prozesses mit Reproduktionsverteilung  $\mathfrak{B}_{n, \frac{\lambda}{n}}$  und Start in  $k$ , so gilt

$$P(|C_{\leq k}| > m) \geq P(T_{n-m} > m).$$

**Beweis:** Diese Ungleichung ist leicht zu sehen, wenn man eine leicht abgewandelte Version des in Abschnitt 2.1.4 vorgestellten Explorationsprozesses betrachtet. Wie in Abschnitt 2.1.4 ist zu Beginn der Exploration jeder der  $k$  Startknoten aktiv. Zusätzlich bezeichnet man  $m - k$  nicht aktive Knoten mit *verboten*, die übrigen Knoten starten neutral. Wenn jetzt in einem Schritt des Prozesses ein zufälliger aktiver Knoten ausgewählt wird und den Status inaktiv bekommt,

aktiviert man nur alle mit diesem Knoten verbundenen neutralen, nicht verbotenen Knoten. Für jeden auf diese Weise aktivierten Knoten setzt man einen verbotenen Knoten auf neutral, so dass zu jedem Zeitpunkt des Prozesses die Summe aus aktiven, inaktiven und verbotenen Knoten genau  $m$  ergibt. Der Prozess stoppt, wenn keine aktiven oder keine verbotenen Knoten mehr verbleiben. Da zu jedem Zeitpunkt  $n - m$  neutrale, nicht verbotene Knoten existieren, ist die Anzahl der neu aktivierten Knoten in jedem Schritt binomialverteilt mit Parametern  $n - m$  und  $p$ , so dass die Anzahl der aktiven Knoten ein Galton-Watson-Prozess mit Start in  $k$  und Reproduktionsverteilung  $\mathfrak{B}_{n-m, \frac{\lambda}{n}}$  ist.  $\square$

Mit Hilfe dieser Schranken lassen sich jetzt Aussagen über die Gesamtpopulation eines Galton-Watson-Prozesses auf die in dieser Arbeit behandelten Fragestellungen übertragen. Dwa hat in [Dwa69] die Verteilung der Gesamtpopulation untersucht und eine alternative Darstellung gefunden.

**Satz 2.20** ([Dwa69, Theorem 1]). *Bezeichne  $P_k$  die Verteilung eines Galton-Watson-Prozesses  $(Z_n)_{n \in \mathbb{N}}$  mit Reproduktionsverteilung  $\nu$  und Start in  $k$  und sei  $(X_n)_{n \in \mathbb{N}}$  die Familie der Zuwächse von  $(Z_n)_{n \in \mathbb{N}}$  wie in Definition 2.15. Dann gilt für die Gesamtpopulation  $T$  von  $(Z_n)_{n \in \mathbb{N}}$ :*

$$P_k(T = n) = \frac{k}{n} P(X_1 + \cdots + X_n = n - k).$$

**Beweis:** Der Beweis, der hier gegeben wird, ist angelehnt an den Beweis von [Hof09, Theorem 3.19] und erfolgt über vollständige Induktion über  $n$ . Für  $n = 1$  sind beide Seiten der Gleichung nur für  $k = 1$  nicht 0 und

$$P_1(T = 1) = P_1(Z_1 = 0) = \nu(\{0\}) = P(X_1 = 0).$$

Gelte die Aussage nun für  $n - 1$ , so betrachtet man

$$P_k(T = n) = \sum_{s=0}^{\infty} P_k(T = n | X_1 = s) P(X_1 = s).$$

Der Prozess  $(Z_n)_{n \in \mathbb{N}}$  ist ein Markow-Prozess, so dass

$$P_k(T = n | X_1 = s) = P_{k+s-1}(T = n - 1),$$

was zu

$$P_k(T = n) = \sum_{s=0}^{\infty} P_{k+s-1}(T = n - 1)P(X_1 = s)$$

führt. Bezeichnet man  $S_n := \sum_{i=1}^n X_i$  und wendet die die Induktionsvoraussetzung an, so ergibt sich

$$P_{k+s-1}(T = n - 1) = \frac{k - 1 + s}{n - 1} P(S_{n-1} = n - k - s)$$

und unter der Verwendung der Markow-Eigenschaft des Prozesses  $(S_n)_{n \in \mathbb{N}}$

$$\begin{aligned} (n - 1)P_k(T = n) &= \sum_{s=0}^{\infty} (k - 1 + s)P(S_{n-1} = n - k - s)P(X_1 = s) \\ &= \sum_{s=0}^{\infty} (k - 1 + s)P(S_n = n - k | X_1 = s)P(X_1 = s) \\ &= \sum_{s=0}^{\infty} (k - 1 + s)P(X_1 = s | S_n = n - k)P(S_n = n - k) \\ &= P(S_n = n - k) \left( (k - 1) + \sum_{s=0}^{\infty} sP(X_1 = s | S_n = n - k) \right) \\ &= P(S_n = n - k) ((k - 1) + \mathbb{E}(X_1 | S_n = n - k)). \end{aligned}$$

Da der Wert von  $S_n$  unabhängig ist von Permutationen der Summanden, ist

$$\begin{aligned} \mathbb{E}(X_1 | S_n = n - k) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i | S_n = n - k) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i | S_n = n - k) \\ &= \frac{1}{n} \mathbb{E}(S_n | S_n = n - k) \\ &= \frac{n - k}{n}, \end{aligned}$$



so dass

$$\begin{aligned} P_k(T = n) &= \frac{1}{n-1} \left( k - 1 + \frac{n-k}{n} \right) P(S_n = n-k) \\ &= \frac{k}{n} P(S_n = n-k), \end{aligned}$$

was zu zeigen war. □

In Abschnitt 3 wird eine Abschätzung der erwarteten Gesamtpopulation von bestimmten Galton-Watson-Prozessen benötigt, die hier abschließend hergeleitet werden soll. Mit Hilfe der Bemerkung 2.14 lässt sich die Gesamtpopulation  $T$  des Prozesses einfacher darstellen als die Summe der Größen der einzelnen Generationen:

$$T = \sum_{t=0}^{\infty} Z'_t.$$

Diese Darstellung erlaubt eine Abschätzung des Erwartungswertes von  $T$ .

**Lemma 2.21.** *Sei  $(Z'_t)_{t \in \mathbb{N}}$  ein Galton-Watson-Prozess im Sinne von Bemerkung 2.14,  $\mu$  der Erwartungswert der Reproduktionsverteilung von  $(Z'_t)_{t \in \mathbb{N}}$  und  $Z'_0 = k$ . Für  $\mu < 1$  gilt*

$$\mathbb{E}T = k \frac{1}{1 - \mu}.$$

**Beweis:** Der Erwartungswert von  $Z'_t$  lässt sich rekursiv angeben, da die Unabhängigkeit von  $X_{t+1,i}$  von  $Z'_t$  folgende Umformungen erlaubt:

$$\begin{aligned} \mathbb{E}Z'_{t+1} &= \mathbb{E}(\mathbb{E}(Z'_{t+1}|Z'_t)) \\ &= \mathbb{E}\left(\sum_{s=0}^{\infty} \mathbb{E}(Z'_{t+1}|Z'_t = s) \mathbb{1}_{\{Z'_t=s\}}\right) \\ &= \mathbb{E}\left(\sum_{s=0}^{\infty} \mathbb{E}\left(\sum_{i=1}^s X_{t+1,i}|Z'_t = s\right) \mathbb{1}_{\{Z'_t=s\}}\right) \\ &= \mathbb{E}\left(\sum_{s=0}^{\infty} s\mu \mathbb{1}_{\{Z'_t=s\}}\right) \\ &= \mu \mathbb{E}Z'_t. \end{aligned}$$

Es folgt also mit  $Z'_0 = k$

$$\mathbb{E}Z'_t = k\mu^t.$$

Damit ist

$$\begin{aligned}\mathbb{E}T &= \mathbb{E} \sum_{t=0}^{\infty} Z'_t \\ &= \sum_{t=0}^{\infty} k\mu^t \\ &= k \frac{1}{1-\mu}.\end{aligned}$$

□

### 2.2.1 Galton-Watson-Prozesse mit poissonverteilten Zuwächsen

Zur Vereinfachung der Untersuchung von Galton-Watson-Prozessen mit binomiellen Zuwächsen bietet es sich auf Grund des Poissonschen Grenzwertsatzes (z. B. [Als00])

$$\mathfrak{B}_{n, \frac{\lambda}{n}} \rightarrow Poi_{\lambda} \quad \text{in totaler Variation}$$

an, poissonverteilte Zuwächse an Stelle der binomiellen zu betrachten.

In diesem Abschnitt werden daher Fehlerabschätzungen für diese Approximation sowie für diese Arbeit wichtige Aussagen über Galton-Watson-Prozesse mit poissonverteilten Zuwächsen vorgestellt.

**Lemma 2.22.** *Sei  $X \sim \mathfrak{B}_{n,p}$  und  $Y \sim Poi_{\lambda}$  mit  $\lambda = np$ , so gibt es eine Kopplung  $(X', Y')$  von  $X$  und  $Y$ , so dass*

$$P(X' \neq Y') \leq \frac{\lambda^2}{n}.$$

**Beweis:** Es genügt, für  $X \sim \mathfrak{B}_{1,p}$  und  $Y \sim Poi_p$  eine Kopplung  $(X', Y')$  anzugeben, in der  $P(X' \neq Y') \leq p^2$  gilt, denn seien  $(X_i)_{i \in N}$  und  $(Y_i)_{i \in N}$  Familien von unabhängigen, identisch verteilten Zufallsgrößen mit  $X_1 \sim \mathfrak{B}_{1,p}$  und  $Y_1 \sim Poi_p$ , so dass  $\sum_{i=1}^n X_i \sim \mathfrak{B}_{n,p}$  und  $\sum_{i=1}^n Y_i \sim Poi_{\lambda}$ , und seien die individuellen Kopplungen  $(X'_i, Y'_i)$  unabhängig voneinander gewählt, so ist  $(\sum_{i=1}^n X'_i, \sum_{i=1}^n Y'_i)$  eine Kopplung von  $\sum_{i=1}^n X_i$  und  $\sum_{i=1}^n Y_i$ , und es gilt

$$\begin{aligned} P\left(\sum_{i=1}^n X'_i \neq \sum_{i=1}^n Y'_i\right) &\leq P\left(\bigcup_{1 \leq i \leq n} \{X'_i \neq Y'_i\}\right) \\ &\leq \sum_{i=1}^n P(X'_i \neq Y'_i) \\ &\leq np^2 = \frac{\lambda^2}{n}. \end{aligned}$$

Eine passende Kopplung von  $X$  und  $Y$  ist die maximale Kopplung (s. [Tho00, Kapitel 1.4]); sie hat folgende Form:

$$P(X' = x, Y' = y) = \begin{cases} 1 - p & \text{für } x = 0, y = 0 \\ 0 & \text{für } x = 0, y > 0 \\ e^{-p} - (1 - p) & \text{für } x = 1, y = 0 \\ \frac{p^y}{y!} e^{-p} & \text{für } x = 1, y > 0 \end{cases}$$

Es ist leicht nachzurechnen, dass diese Definition von  $(X', Y')$  tatsächlich eine Kopplung von  $X$  und  $Y$  ist. Außerdem gilt

$$\begin{aligned} P(X' \neq Y') &= 1 - P(X' = Y') \\ &= 1 - (1 - p) - (pe^{-p}) \\ &= p(1 - e^{-p}) \\ &\leq p^2, \end{aligned}$$

was den Beweis abschließt. □

Dieses Lemma hilft, folgende wichtige Aussage zu beweisen:

**Lemma 2.23.** *Seien  $T$  und  $T^*$  die Gesamtpopulationen von Galton-Watson-Prozessen  $(Z_n)_{n \in \mathbb{N}}$  und  $(Z_n^*)_{n \in \mathbb{N}}$  mit Reproduktionsverteilung  $\mathfrak{B}_{n,p}$  (für  $T$ ) beziehungsweise  $Poi_\lambda$  (für  $T^*$ ) mit  $\lambda = np$  und sei  $Z_0 = Z_0^*$ . Dann gilt für jedes  $m \in \mathbb{N}$*

$$P(T \geq m) = P(T^* \geq m) + r_{m,\lambda},$$

mit

$$|r_{m,\lambda}| \leq \frac{2\lambda^2}{n} \sum_{s=1}^{m-1} P(T^* \geq s).$$

**Beweis:** Kern des Beweises ist es, eine geeignete Abhängigkeit der beiden auftretenden Prozesse zu wählen. Dafür betrachtet man in gewissem Sinne die Versionen der Zuwächse mit der höchsten Abhängigkeit, die auch im Beweis von Lemma 2.22 gewählt wurde und überträgt das dort gewonnene Resultat auf den gesamten Prozess. Seien  $(X_i)_{i \in \mathbb{N}}$  und  $(X_i^*)_{i \in \mathbb{N}}$  die Zuwächse von  $(Z_n)_{n \in \mathbb{N}}$  und  $(Z_n^*)_{n \in \mathbb{N}}$ , also  $X_i \sim \mathfrak{B}_{n,p}$  und  $X_i^* \sim Poi_\lambda$  für alle  $i \in \mathbb{N}$  und zusätzlich so gewählt, dass

$$P(X_i \neq X_i^*) \leq \frac{\lambda^2}{n},$$

was Lemma 2.22 erlaubt. Betrachtet man

$$P(T \geq m) = P(T \geq m, T^* \geq m) + P(T \geq m, T^* < m)$$

$$P(T^* \geq m) = P(T \geq m, T^* \geq m) + P(T^* \geq m, T < m)$$

und bildet die Differenz dieser Gleichungen, so ergibt sich

$$|P(T \geq m) - P(T^* \geq m)| \leq P(T^* \geq m, T < m) + P(T \geq m, T^* < m).$$

Tritt eines der Ereignisse auf der rechten Seite dieser Ungleichung ein, so kann das nur dadurch zu Stande kommen, dass in einem Schritt vor  $m$  die Zuwächse

der beiden Prozesse nicht identisch waren, also

$$\begin{aligned}
 P(T \geq m, T^* < m) &\leq P(\exists 1 \leq s \leq m-1 : X_i = X_i^* \forall i < s, X_s \neq X_s^*, T \geq m) \\
 &\leq \sum_{s=1}^{m-1} P(X_i = X_i^* \forall i < s, X_s \neq X_s^*, T \geq s) \\
 &\leq \sum_{s=1}^{m-1} P(X_i = X_i^* \forall i < s, X_s \neq X_s^*, T^* \geq s) \\
 &\leq \sum_{s=1}^{m-1} P(X_s \neq X_s^*, T^* \geq s) \\
 &= \sum_{s=1}^{m-1} P(X_s \neq X_s^*) P(T^* \geq s) \\
 &\leq \frac{\lambda^2}{n} \sum_{s=1}^{m-1} P(T^* \geq s).
 \end{aligned}$$

Das andere Ereignis lässt sich analog behandeln; insgesamt folgt die Behauptung.

□

Um die recht umfangreiche Berechnung des Restterms in einigen einfachen Fällen später abzukürzen, stellt man fest:

**Korollar 2.24.** *In der Situation von Lemma 2.23 ist*

$$P(T \geq m) = P(T^* \geq m) + \mathcal{O}\left(\frac{m}{n}\right).$$

**Beweis:** Zum Beweis genügt hier die einfache Abschätzung

$$\begin{aligned}
 |r_{m,\lambda}| &\leq \frac{2\lambda^2}{n} \sum_{s=1}^{m-1} P(T^* \geq s) \\
 &\leq \frac{2\lambda^2(m-1)}{n} \in \mathcal{O}\left(\frac{m}{n}\right).
 \end{aligned}$$

□

Im Verlauf der Untersuchung der Größe der größten Komponente eines zufälligen Graphen korrespondiert im Fall  $\lambda > 1$  das Auftreten der größten Komponente mit dem Überleben der Population in dem zum Vergleich herangezogenen Verzweigungsprozess. So wird der Explorationsprozess, der in zufällig gewählten Knoten des Graphen startet, etwa mit der Wahrscheinlichkeit des Überlebens der Population des Verzweigungsprozesses die größte Komponente des Graphen erforschen. Dieses Ergebnis kommt zum Beispiel in Satz 3.6 zum Ausdruck.

**Lemma 2.25.** *Sei  $\zeta_\lambda$  die Überlebenswahrscheinlichkeit eines Galton-Watson-Prozesses  $(Z_t)_{t \in \mathbb{N}}$  mit Reproduktionsverteilung  $\text{Poi}_\lambda$ ,  $\lambda > 1$  und  $Z_0 = 1$ . Dann gilt*

$$1 - \zeta_\lambda = e^{-\lambda \zeta_\lambda}.$$

**Beweis:** Sei  $X_1$  die Anzahl der Nachkommen des ersten Individuums. Jeder dieser Nachkommen starte dann einen eigenen, von den anderen unabhängigen Galton-Watson-Prozess mit Gesamtpopulation  $T_i$ ,  $1 \leq i \leq X_1$ , wie es die Darstellung aus Bemerkung 2.14 erlaubt. Dann lässt sich die Gesamtpopulation des Gesamtprozesses darstellen als

$$T = 1 + \sum_{i=1}^{X_1} T_i,$$

woraus sich folgern lässt

$$\begin{aligned}
 1 - \zeta_\lambda &= P(T < \infty) \\
 &= P\left(\sum_{i=1}^{X_1} T_i < \infty\right) \\
 &= \sum_{j=0}^{\infty} P\left(\sum_{i=1}^j T_i < \infty \mid X_1 = j\right)P(X_1 = j) \\
 &= \sum_{j=0}^{\infty} P(\forall 1 \leq i \leq j : T_i < \infty)P(X_1 = j) \\
 &= \sum_{j=0}^{\infty} (1 - \zeta_\lambda)^j \frac{\lambda^j}{j!} e^{-\lambda} \\
 &= e^{-\lambda} e^{\lambda(1-\zeta_\lambda)} \\
 &= e^{-\lambda\zeta_\lambda},
 \end{aligned}$$

was zu zeigen war. □

Um Aussagen über einen typischen Verlauf des Prozesses machen zu können, steht das folgende Lemma zur Verfügung, das die Wahrscheinlichkeit eines Aussterbens einer großen Population durch eine exponentiell in der Größe fallende Schranke begrenzt.

**Lemma 2.26.** *Bezeichne  $T$  die Gesamtpopulation eines Galton-Watson-Prozesses  $(Z_t)_{t \in \mathbb{N}}$  mit Reproduktionsverteilung  $\text{Poi}_\lambda$  für  $\lambda > 1$  und Start in  $k$ . Dann gibt es  $C > 0$  und  $I > 0$ , so dass für alle  $m \in \mathbb{N}$*

$$P(m < T < \infty) \leq Ce^{-mI}.$$

**Beweis:** Nach Definition der Gesamtpopulation  $T$  ist  $T = n$  äquivalent zu  $Z_n = 0 \wedge Z_i > 0$  für alle  $i < n$ . In diesem Fall ist  $0 = Z_n = Z_0 - n + \sum_{i=1}^n X_i$ . Der Satz von Cramér (z. B. [OV05, Theorem 1.1]) liefert dann für ein geeignetes

$I > 0$  wegen  $\mathbb{E}X_1 = \lambda > 1$ :

$$\begin{aligned}
 P(m < T < \infty) &= \sum_{n=m}^{\infty} P(T = n) \\
 &\leq \sum_{n=m}^{\infty} P\left(\sum_{i=1}^n X_i = n - k\right) \\
 &\leq \sum_{n=m}^{\infty} P\left(\sum_{i=1}^n X_i < n\right) \\
 &\leq \sum_{n=m}^{\infty} e^{-nI} \\
 &= e^{-mI} \frac{1}{1 - e^{-I}},
 \end{aligned}$$

was das Lemma beweist. □

Zum Schluss der allgemeinen Untersuchungen von Galton-Watson-Prozessen mit poissonverteilten Zuwächsen erhält man durch die Anwendung von Satz 2.20 eine Abschätzung der Verteilung der Gesamtpopulation im kritischen und subkritischen Fall.

**Lemma 2.27.** *Sei  $0 < \lambda \leq 1$  und  $T^*$  die Gesamtpopulation eines Galton-Watson-Prozesses mit Reproduktionsverteilung  $Poi_\lambda$ . Dann gilt*

$$\frac{1}{\lambda\sqrt{2\pi t^3}} e^{-\frac{1}{12n}} e^{-I_\lambda t} \leq P_\lambda(T^* = t) \leq \frac{1}{\lambda\sqrt{2\pi t^3}} e^{-I_\lambda t}$$

mit

$$I_\lambda = \lambda - 1 - \ln \lambda.$$

**Beweis:** Satz 2.20 erlaubt die Umformung

$$\begin{aligned}
 P_\lambda(T^* = t) &= \frac{1}{t} Poi_{\lambda t}(t - 1) \\
 &= \frac{1}{t} \frac{(\lambda t)^{t-1}}{(t-1)!} e^{-\lambda t} \\
 &= \frac{1}{t} \frac{t^t}{t!} e^{-t} e^{-I_\lambda t}.
 \end{aligned}$$



Die Stirling-Formel liefert

$$\sqrt{2\pi t} \left(\frac{t}{e}\right)^t \leq t! \leq \sqrt{2\pi t} \left(\frac{t}{e}\right)^t e^{\frac{1}{12t}},$$

was die Behauptung zeigt. □

## 2.3 Tailabschätzungen der Größe der ersten Komponente

Eine bedeutende Hilfsgröße, die in den Beweisen in Kapitel 3 verwendet wird, ist die Anzahl der Knoten in großen Komponenten

$$\begin{aligned} Z_{>k} &:= \#\{v \in V : |C_{(v)}| > k\} \\ &= \sum_{v=1}^n \mathbb{1}_{\{|C_{(v)}| > k\}}, \end{aligned}$$

da sie einen leichteren Zugang zur Größe der größten Komponente ermöglicht. Für alle  $k \leq |C_{\max}|$  ist  $Z_{>k} \geq |C_{\max}|$  und für  $k > |C_{\max}|$  gilt  $Z_{>k} = 0$ , so dass

$$|C_{\max}| > k \iff Z_{>k} > k \iff Z_{>k} > 0.$$

Die erwartete Anzahl der Knoten in großen Komponenten  $\mathbb{E}_\lambda Z_{>k}$  liefert einen guten Ansatz, um das Verhalten von  $Z_{>k}$  zu verstehen, und wegen

$$\begin{aligned} \mathbb{E}_\lambda Z_{>k} &= \mathbb{E}_\lambda \sum_{v=1}^n \mathbb{1}_{\{|C_{(v)}| > k\}} \\ &= nP_\lambda(|C_{(1)}| > k), \end{aligned}$$

ist eine Untersuchung der Tails der Größe der ersten Komponente, also des Verhaltens von  $P_\lambda(|C_{(1)}| > k)$ , für große  $k$  hilfreich. Diese wird im folgenden Abschnitt durchgeführt.

### 2.3.1 Der superkritische Fall: $\lambda > 1$

**Lemma 2.28.** *Sei  $\lambda > 1$  und  $k_n = K \ln n$  für ein hinreichend großes  $K$ , dann ist*

$$P(|C_{(1)}| > k_n) = \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right).$$

**Beweis:** Der Beweis ergibt sich aus dem im vorigen Abschnitt behandelten Vergleich der Größe der Zusammenhangskomponenten in einem zufälligen Graphen mit der Gesamtpopulation eines Galton-Watson-Prozesses mit geeigneter Reproduktionsverteilung. Sei also  $T_\lambda$  die Gesamtpopulation eines Galton-Watson-Prozesses mit Reproduktion gemäß einer Poissonverteilung mit Parameter  $\lambda$ .

Für die obere Schranke liefert Lemma 2.23 zusammen mit Lemma 2.26 für

$$I_\lambda = \lambda - 1 - \ln \lambda$$

und  $K > I_\lambda^{-1}$ :

$$\begin{aligned} P(|C_{(1)}| > k_n) &\leq P(T_\lambda > k_n) + \mathcal{O}\left(\frac{k_n}{n}\right) \\ &= P(T_\lambda = \infty) + P(k_n < T_\lambda < \infty) + \mathcal{O}\left(\frac{k_n}{n}\right) \\ &\leq \zeta_\lambda + \mathcal{O}(e^{-k_n I_\lambda}) + \mathcal{O}\left(\frac{k_n}{n}\right) \\ &= \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right), \end{aligned}$$

wie gewünscht.

Für die untere Schranke betrachtet man analog für  $\lambda_n = \lambda(1 - \frac{k_n}{n})$

$$\begin{aligned} P(|C_{(1)}| > k_n) &\geq P(T_{\lambda_n} > k_n) + \mathcal{O}\left(\frac{k_n}{n}\right) \\ &= \zeta_{\lambda_n} + \mathcal{O}\left(\frac{k_n}{n}\right). \end{aligned}$$

Lemma 2.25 hat gezeigt, dass  $\zeta_\lambda$  die Gleichung  $F(\lambda, \zeta_\lambda) := 1 - \zeta_\lambda - e^{-\lambda\zeta_\lambda} = 0$  löst. Die Ableitung dieses Ausdruckes nach der zweiten Komponente verschwindet an keiner Stelle  $(\lambda, \zeta_\lambda)$  für  $\lambda > 1$ , denn

$$\begin{aligned} \frac{d}{dy} F(\lambda, y) \Big|_{y=\zeta_\lambda} &= \lambda e^{-\lambda y} - 1 \Big|_{y=\zeta_\lambda} \\ &= \lambda(1 - \zeta_\lambda) - 1 \end{aligned}$$

und mit Hilfe von Lemma 2.25

$$\begin{aligned} \lambda(1 - \zeta_\lambda) - 1 = 0 &\Leftrightarrow \zeta_\lambda = 1 - \frac{1}{\lambda} \\ &\Leftrightarrow \frac{1}{\lambda} = e^{1-\lambda} \\ &\Leftrightarrow 1 = \lambda e^{1-\lambda}. \end{aligned}$$

Weil  $\lambda e^{1-\lambda} \xrightarrow{\lambda \rightarrow 1} 1$  und

$$\frac{d}{d\lambda} \lambda e^{1-\lambda} = (1 - \lambda^2) e^{1-\lambda} < 0$$

ist diese Gleichung aber für kein  $\lambda > 1$  erfüllt. Der Satz über implizite Funktionen [Heu00, Sätze 169.1ff] liefert unter dieser Voraussetzung die Differenzierbarkeit der Funktion  $\lambda \rightarrow \zeta_\lambda$  und damit  $\zeta_{\lambda_n} = \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right)$ .  $\square$

### 2.3.2 Der subkritische Fall: $\lambda < 1$

Über den Vergleich mit dem im Abschnitt 2.1.4 hergeleiteten Explorationsprozess lassen sich gute untere und obere Schranken für die Tails der Verteilung der Größe der ersten Komponente im subkritischen Fall angeben.

**Korollar 2.29.** *Für  $\lambda < 1$  ist für jedes  $k \in \mathbb{N}$  die Ungleichung*

$$P_\lambda(|C_{(1)}| > k) \leq \frac{1}{\lambda} e^{-kI_\lambda}$$

*erfüllt, mit  $I_\lambda = \lambda - \ln \lambda - 1$ .*

Im Kapitel 4 wird eine leichte Verallgemeinerung dieses Korollars für eine Abschätzung benötigt, die an Stelle der vorhergehenden Aussage hier bewiesen wird.

**Lemma 2.30.** *Für  $\lambda < 1$  ist für alle  $k, m \in \mathbb{N}$  mit  $m < k$*

$$P_\lambda(|C_{\leq m}| > k) \leq e^{-kI_\lambda - m \ln \lambda}$$

mit  $I_\lambda = \lambda - \ln \lambda - 1$ .

**Beweis:** Ruft man sich in Erinnerung, dass die Größe des von den ersten  $m$  Knoten aufgespannten Teilgraphen gerade die Zeit des ersten Nulldurchgangs des Explorationsprozesses mit Start in  $m$  ist (Bemerkung 2.9) und verwendet die abkürzende Schreibweise  $p = \frac{\lambda}{n}$ , so kann man mit Hilfe von Lemma 2.12 folgende Abschätzung vornehmen:

$$\begin{aligned} P_\lambda(|C_{\leq m}| > k) &\leq P_\lambda(S_k > 0) \\ &\leq P(B(n - m, (1 - (1 - p)^k)) \geq k - m), \end{aligned}$$

wenn  $B(n, p)$  eine unter  $P$  wie  $\mathfrak{B}_{n,p}$  verteilte Zufallsgröße ist. Wegen

$$1 - (1 - p)^k \Big|_{p=0} = 0 \text{ und } pk \Big|_{p=0} = 0,$$

sowie

$$\begin{aligned} \frac{d}{dp}(1 - (1 - p)^k) &= k(1 - p) \\ &\leq k = \frac{d}{dp}(kp) \end{aligned}$$

ist  $1 - (1 - p)^k \leq kp$  für alle  $p \in (0, 1)$  und  $k \in \mathbb{N}$ , so dass

$$P(B(n - m, (1 - (1 - p)^k)) \geq k - m) \leq P(B(n, kp) \geq k - m).$$

An dieser Stelle benutzt man die exponentielle Markow-Ungleichung und die einfache Ungleichung

$$1 + x \leq e^x$$

um zu zeigen:

$$\begin{aligned} P(B(n, kp) \geq k - m) &\leq \min_{t \in \mathbb{R}^+} e^{-(k-m)t} (E e^{tB(n, kp)})^n \\ &= \min_{t \in \mathbb{R}^+} e^{-(k-m)t} (1 - kp + kpe^t)^n \\ &\leq \min_{t \in \mathbb{R}^+} e^{-(k-m)t} (1 - kp + kpe^t)^n \\ &\leq \min_{t \in \mathbb{R}^+} e^{-(k-m)t} e^{nkp(e^t-1)} \\ &= \min_{t \in \mathbb{R}^+} e^{-k(t-\lambda(e^t-1))+mt}. \end{aligned}$$

Insbesondere ist damit für  $t = -\ln \lambda > 0$

$$\begin{aligned} P_\lambda(|C_{\leq m}| > k) &\leq e^{-k(-\ln \lambda - \lambda(e^{-\ln \lambda} - 1)) - m \ln \lambda} \\ &= e^{-kI_\lambda - m \ln \lambda}, \end{aligned}$$

was das Lemma beweist. □

Für die untere Schranke steht mit Lemma 2.19 ein nicht ganz so starkes Hilfsmittel für die Abschätzung bereit. Dementsprechend erhält man für diese nur eine Aussage für große  $k$ .

**Lemma 2.31.** *Für jedes  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{N}$ , so dass für alle  $k > K$*

$$P_\lambda(|C_{(1)}| > k) \geq \frac{1}{\lambda_n} \frac{e^{-(I_{\lambda_n} + \varepsilon)k}}{1 - e^{-(I_{\lambda_n} + \varepsilon)}} + \mathcal{O}\left(\frac{k}{n}\right),$$

mit  $\lambda_n = \lambda \frac{n-k}{n}$ .

**Beweis:** Sei hier  $T_\lambda^*$  unter  $P$  verteilt wie die Gesamtpopulation eines Galton-Watson-Prozesses mit Reproduktionverteilung  $Poi_\lambda$ , so liefert Lemma 2.19 zusammen mit Korollar 2.24

$$P_\lambda(|C_{(1)}| > k) \geq P(T_{\lambda_n}^* > k) + \mathcal{O}\left(\frac{k}{n}\right).$$

Über das Lemma 2.27 lässt sich eine untere Schranke für diese Wahrscheinlichkeit angeben:

$$\begin{aligned} P(T_{\lambda_n}^* > k) &= \sum_{i=k+1}^{\infty} P(T_{\lambda_n}^* = i) \\ &\geq \frac{1}{\lambda_n} \sum_{i=k+1}^{\infty} \frac{1}{\sqrt{2\pi i^3}} e^{-I_{\lambda_n} i}. \end{aligned}$$

Da  $i^{-\frac{3}{2}}$  langsamer als exponentiell fällt, gibt es für jedes  $\varepsilon > 0$  ein  $K \in \mathbb{N}$ , so dass für alle  $i > K$

$$\frac{1}{\sqrt{2\pi i^3}} \geq e^{-\varepsilon i}.$$

Für  $k > K$  ist dann

$$\begin{aligned} \frac{1}{\lambda_n} \sum_{i=k+1}^{\infty} \frac{1}{\sqrt{2\pi i^3}} e^{-I_{\lambda_n} i} &\geq \frac{1}{\lambda_n} \sum_{i=k+1}^{\infty} e^{-(I_{\lambda_n} + \varepsilon) i} \\ &= \frac{1}{\lambda_n} \frac{e^{-(I_{\lambda_n} + \varepsilon) k}}{1 - e^{-(I_{\lambda_n} + \varepsilon)}}, \end{aligned}$$

was zu zeigen war. □

Schließlich wird hier noch eine obere Schranke für die erwartete Größe der ersten Komponente angegeben.

**Korollar 2.32.** *Für  $\lambda < 1$  ist*

$$\mathbb{E}_\lambda |C_{(1)}| \leq \frac{1}{1 - \lambda}.$$

**Beweis:** Sei  $T$  die Gesamtpopulation eines Galton-Watson-Prozesses mit Start in 1 und Reproduktionsverteilung  $\mathfrak{B}_{n, \frac{\lambda}{n}}$ . Das Lemma 2.18 ermöglicht zusammen mit Lemma 2.21 die Abschätzung

$$\mathbb{E}_\lambda |C_{(1)}| \leq \mathbb{E}T = \frac{1}{1 - \lambda}.$$

□

### 2.3.3 Der kritische Fall: $\lambda = 1$

**Lemma 2.33.** Für  $\lambda = 1$  gibt es Konstanten  $c_1, c_2 > 0$ , so dass

$$\frac{c_1}{\sqrt{k}} \leq P_1(|C_{(1)}| \geq k) \leq \frac{c_2}{\sqrt{k}}.$$

**Beweis:** Für den Beweis der oberen Schranke vergleicht man wie oben erläutert die Größe der ersten Komponente des zufälligen Graphen mit der Gesamtpopulation eines Galton-Watson-Prozesses mit Reproduktionsverteilung  $Poi_1$  und Start in 1. Die Abschätzungen in Bemerkung 2.18 und Lemma 2.23 liefern

$$P_1(|C_{(1)}| \geq k) \leq P_1(T^* \geq k) + r_{n,k},$$

mit

$$|r_{n,k}| \leq \frac{2}{n} \sum_{s=1}^k P_1(T^* \geq s).$$

Um die gewünschte Schranke zu erreichen, betrachtet man zunächst mit Hilfe

von Lemma 2.27

$$\begin{aligned}
 P_1(T^* \geq s) &= \sum_{t=s}^{\infty} P_1(T^* = t) \\
 &\leq \sum_{t=s}^{\infty} \frac{1}{\sqrt{2\pi t^3}} \\
 &\leq \frac{1}{2\pi} \int_s^{\infty} \frac{1}{(t-1)^3} dt \\
 &= \frac{2}{3\sqrt{2\pi}} \frac{1}{\sqrt{s-1}} \\
 &\leq \frac{C}{\sqrt{s}},
 \end{aligned}$$

für ein geeignetes  $C > 0$ . Für die Abschätzung des Restterms führt dies zu

$$\begin{aligned}
 |r_{n,k}| &\leq \frac{2}{n} \sum_{s=1}^k P_1(T^* \geq s) \\
 &\leq \frac{2}{n} \sum_{s=1}^k \frac{C}{\sqrt{s}} \\
 &\leq \frac{2C}{n} \int_1^k \frac{1}{\sqrt{s-1}} ds \\
 &= \frac{4C}{n} \sqrt{k-1} \\
 &\leq \frac{4C}{\sqrt{k}},
 \end{aligned}$$

da  $k < n$ . Insgesamt ist mit  $c_2 = 5C$  die Ungleichung

$$P_1(|C_{(1)}| \geq k) \leq \frac{c_2}{\sqrt{k}}$$

erfüllt.

Für den Beweis der unteren Schranke wird Lemma 2.19 zusammen mit Lemma 2.23 und der gerade hergeleiteten Abschätzung von  $|r_{n,k}|$  herangezogen, um für



$k \leq rn^{\frac{2}{3}}$  und  $\lambda_n = 1 - rn^{-\frac{1}{3}}$  die Abschätzung

$$\begin{aligned}
 P_1(|C_{(1)}| \geq k) &\geq P_{n-k, \frac{\lambda}{n}}(T \geq k) \\
 &\geq P_{\lambda_n}(T^* \geq k) - r_{n,k} \\
 &\geq P_{\lambda_n}(T^* \geq k) - \frac{4C\sqrt{k}}{n} \\
 &\geq P_{\lambda_n}(T^* \geq k) - \frac{4C\sqrt{r}}{n^{\frac{2}{3}}}
 \end{aligned}$$

zu erhalten. Verwendet man wieder Satz 2.20 für eine Umformung des ersten Summanden, so ergibt sich

$$\begin{aligned}
 P_{\lambda_n}(T^* \geq k) &= \sum_{t=k}^{\infty} P_{\lambda_n}(T^* = t) \\
 &= \sum_{t=k}^{\infty} \frac{1}{t} \text{Poi}_{t\lambda_n}(t-1) \\
 &= \sum_{t=k}^{\infty} \frac{1}{t} \frac{(\lambda_n t)^{t-1}}{(t-1)!} e^{-\lambda_n t} \\
 &= \sum_{t=k}^{\infty} \frac{1}{t} \frac{t^{t-1}}{(t-1)!} e^{-t} \lambda_n^{t-1} e^{-(\lambda_n-1)t} \\
 &= \sum_{t=k}^{\infty} \frac{1}{t} \text{Poi}_1(t-1) \frac{1}{\lambda_n} e^{-(\lambda_n - \ln \lambda_n - 1)t} \\
 &\geq \sum_{t=k}^{\infty} P_1(T^* = t) e^{-I_{\lambda_n} t},
 \end{aligned}$$

mit  $I_\lambda = \lambda - \ln \lambda - 1$ . Anwendung des Lemmas 2.27 führt dann zu

$$\begin{aligned}
 P_{\lambda_n}(T^* \geq k) &\geq \sum_{t=k}^{\infty} \frac{1}{t} P_1(T^* = t) e^{-I_{\lambda_n} t} \\
 &\geq \sum_{t=k}^{\infty} \frac{1}{\sqrt{2\pi t^3}} e^{-\frac{1}{12n}} e^{-I_{\lambda_n} t} \\
 &\geq \sum_{t=k}^{2k} \frac{1}{\sqrt{2\pi t^3}} e^{-\frac{1}{12n}} e^{-I_{\lambda_n} t} \\
 &\geq k \frac{1}{\sqrt{16\pi k^3}} e^{-\frac{1}{12n}} e^{-2I_{\lambda_n} k} \\
 &\geq \frac{1}{8\sqrt{\pi}} \frac{1}{\sqrt{k}} e^{-2I_{\lambda_n} k}.
 \end{aligned}$$

Wegen

$$\ln(1-x) \geq -x - \frac{x^2}{2}$$

für  $0 \leq x < 1$  ist

$$\begin{aligned}
 I_{\lambda_n} &= \lambda_n - 1 - \ln(1 - (1 - \lambda_n)) \\
 &\leq \lambda_n - 1 + 1 - \lambda_n + \frac{(1 - \lambda_n)^2}{2} \\
 &= \frac{1}{2} r^2 n^{-\frac{2}{3}},
 \end{aligned}$$

so dass für  $k \leq rn^{\frac{2}{3}}$  die Ungleichung

$$e^{-2I_{\lambda_n} k} \geq e^{-r^3}$$

erfüllt ist. Schließlich ist deswegen

$$\begin{aligned}
 P_{\lambda_n}(T^* \geq k) &\geq \frac{1}{8\sqrt{\pi}} \frac{1}{\sqrt{k}} e^{-2I_{\lambda_n} k} \\
 &\geq \frac{1}{8e^{r^3}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{k}}
 \end{aligned}$$

und

$$\begin{aligned}
 P_1(|C_{(1)}| \geq k) &\geq \frac{1}{8e^{r^3}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{k}} - \frac{4C\sqrt{r}}{n^{\frac{2}{3}}} \\
 &\geq \frac{c_1}{\sqrt{k}},
 \end{aligned}$$

mit  $c_1 = \frac{1}{8e^{r^3}\sqrt{\pi}} - 4C\sqrt{r}$ . □

**Lemma 2.34.** *Es gibt ein  $K > 0$ , so dass*

$$E_1|C_{(1)}| \leq Kn^{\frac{1}{3}}.$$

**Beweis:** Die im Fall  $\lambda \neq 1$  erfolgreiche Abschätzung über die stochastische Dominiertheit durch einen Galton-Watson-Prozess mit binomieller Reproduktionsverteilung liefert für  $\lambda = 1$  die Schranke  $E_1|C_{(1)}| \in \mathcal{O}(\sqrt{n})$ , ist hier also nicht stark genug. Deswegen betrachtet man Umformung

$$\begin{aligned} \chi_n(\lambda) &:= E_\lambda|C_{(1)}| \\ &= E_\lambda \sum_{v=1}^n \mathbb{1}_{\{v \leftrightarrow 1\}} \\ &= \sum_{v=2}^n P_\lambda(v \leftrightarrow 1) + 1 \\ &= (n-1)\tau_n(\lambda) + 1 \end{aligned}$$

mit

$$\tau_n(\lambda) := P_\lambda(1 \leftrightarrow 2).$$

Ziel der folgenden Ausführungen ist die Herleitung einer hinreichend scharfen oberen Schranke für  $\tau_n(1)$ . Für ihre Herleitung stellt man zunächst fest, dass

$$\tau_n(\lambda) = \sum_{E \subset \{(i,j) \in \{1,\dots,n\}^2 : i < j\}} \left(\frac{\lambda}{n}\right)^{|E|} \left(1 - \frac{\lambda}{n}\right)^{\frac{n(n-1)}{2} - |E|}$$

ein Polynom von Grad kleiner als  $n^2$  in  $\lambda$  ist und deswegen differenzierbar. Der Hauptsatz der Differential- und Integralrechnung erlaubt deswegen eine Darstellung von  $\tau_n(1)$  der Form

$$\tau_n(1) = \tau_n(\lambda) + \int_\lambda^1 \tau_n'(\alpha) d\alpha$$

für jedes  $\lambda \in (0, 1)$ . Das folgende Lemma 2.35 zeigt für die Ableitung von  $\tau_n$  für jedes  $a \in (0, 1)$  die Ungleichung

$$\tau_n'(\lambda) \leq an^{-\frac{2}{3}}\chi_n(1) + \frac{C}{a}n^{-\frac{1}{3}} + \mathcal{O}\left(\frac{1}{n}\right).$$

Mit dieser Abschätzung ergibt sich mit Hilfe der Umformung

$$\begin{aligned}\chi_n(\lambda) &= (n-1)\tau_n(\lambda) + 1 \\ \Leftrightarrow \tau_n(\lambda) &= \frac{\chi_n(\lambda) - 1}{n-1}\end{aligned}$$

und Korollar 2.32 die Abschätzung

$$\begin{aligned}\tau_n(1) &= \tau_n(\lambda) + \int_{\lambda}^1 \tau_n'(\alpha) d\alpha \\ &\leq \tau_n(\lambda) + \int_{\lambda}^1 an^{-\frac{2}{3}}\chi_n(1) + \frac{C}{a}n^{-\frac{1}{3}} + \mathcal{O}\left(\frac{1}{n}\right) d\alpha \\ &\leq \frac{\lambda}{(n-1)(1-\lambda)} + (1-\lambda) \left( an^{-\frac{2}{3}}\chi_n(1) + \frac{C}{a}n^{-\frac{1}{3}} + \mathcal{O}\left(\frac{1}{n}\right) \right).\end{aligned}$$

Wählt man  $\lambda = 1 - n^{-\frac{1}{3}}$  und  $a = \frac{1}{2}$ , so wird dies für ein geeignetes  $C' > 0$  zu

$$\begin{aligned}\tau_n(1) &\leq \frac{1 - n^{-\frac{1}{3}}}{(n-1)n^{\frac{1}{3}}} + \frac{1}{2}n^{-1}\chi_n(1) + 2Cn^{-\frac{2}{3}} + \mathcal{O}\left(\frac{1}{n}\right) \\ &\leq C'n^{-\frac{2}{3}} + \frac{1}{2}n^{-1}((n-1)\tau_n(1) + 1) + \mathcal{O}\left(\frac{1}{n}\right) \\ &\leq C'n^{-\frac{2}{3}} + \frac{1}{2}\tau_n(1) + \frac{1}{2n} + \mathcal{O}\left(\frac{1}{n}\right).\end{aligned}$$

Diese Ungleichung ist schließlich äquivalent zu

$$\tau_n(1) \leq 2C'n^{-\frac{2}{3}} + \frac{1}{4n} + \mathcal{O}\left(\frac{1}{n}\right).$$

Dies ist für den Abschluss des Beweises scharf genug, denn

$$\begin{aligned}\chi_n(\lambda) &= (n-1)\tau_n(\lambda) + 1 \\ &\leq 2C'n^{\frac{1}{3}} + \frac{1}{4} + \mathcal{O}(1).\end{aligned}$$

Wählt man  $K$  groß genug, so folgt die Behauptung.  $\square$

**Lemma 2.35.** Für  $\lambda < 1$  gibt es ein  $C > 0$ , so dass für jedes  $a \in (0, 1)$

$$\tau'_n(\lambda) \leq \left( an^{-\frac{2}{3}} \chi_n(1) + \frac{C}{a} n^{-\frac{1}{3}} \right) \left( 1 + \frac{\lambda}{n - \lambda} \right)$$

gilt.

**Beweis:** Nach Definition ist

$$\tau'_n(\lambda) = \lim_{\varepsilon \rightarrow 0} \frac{\tau_n(\lambda + \varepsilon) - \tau_n(\lambda)}{\varepsilon}.$$

Sei  $G_{n,\lambda} \sim P_{n,\lambda}$ . Wählt man  $P$  als die Kopplung der Graphen  $(G_{n,\lambda})_{\lambda \in (0,1]}$ , in der für jede Kante  $(s, t)$  eine auf  $(0, 1)$  gleichverteilte Zufallsgröße  $U_{st}$  existiert, unter der

$$\begin{aligned} s \leftrightarrow_{\lambda} t &:\Leftrightarrow (s, t) \in E(G_{n,\lambda}) \\ &\Leftrightarrow U_{st} < \frac{\lambda}{n}, \end{aligned}$$

so erkennt man, dass  $\tau_n$  in  $\lambda$  monoton wachsend ist. Im Folgenden wird  $i \leftrightarrow_{\lambda} 2$  geschrieben, wenn die Knoten  $i$  und  $2$  im Graphen  $G_{n,\lambda}$  verbunden sind und analog  $i \leftrightarrow_{\lambda} j$ , wenn sie in  $G_{n,\lambda}$  kommunizieren. Damit ist

$$\begin{aligned} \tau_n(\lambda + \varepsilon) - \tau_n(\lambda) &= P(1 \leftrightarrow_{\lambda+\varepsilon} 2) - P(1 \leftrightarrow_{\lambda} 2) \\ &= P(1 \leftrightarrow_{\lambda+\varepsilon} 2, 1 \not\leftrightarrow_{\lambda} 2). \end{aligned}$$

Da dieses Ereignis nur eintritt, wenn eine Kante  $(s, t)$ , die die Verbindung zwischen 1 und 2 komplettiert, zwischen  $\lambda$  und  $\lambda + \varepsilon$  entsteht, gilt

$$\begin{aligned} &P(1 \leftrightarrow_{\lambda+\varepsilon} 2, 1 \not\leftrightarrow_{\lambda} 2) \\ &\leq P(\exists s, t \in \{1, \dots, n\} : 1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2) \\ &\leq \sum_{s < t=1}^n P(1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2). \end{aligned}$$

Für eine solche Kante  $(s, t)$  gilt

$$\begin{aligned} & P(1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2) \\ &= P(1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2 \mid s \not\leftrightarrow_{\lambda} t) P(s \not\leftrightarrow_{\lambda} t). \end{aligned}$$

Unter der Bedingung  $s \not\leftrightarrow_{\lambda} t$  hängt dann  $s \leftrightarrow_{\lambda+\varepsilon} t$  nicht mehr von den übrigen Ereignissen ab, so dass

$$\begin{aligned} & P(1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2 \mid s \not\leftrightarrow_{\lambda} t) P(s \not\leftrightarrow_{\lambda} t) \\ &= P(s \leftrightarrow_{\lambda+\varepsilon} t \mid s \not\leftrightarrow_{\lambda} t) P(s \not\leftrightarrow_{\lambda} t) \\ &\quad \times P(1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2 \mid s \not\leftrightarrow_{\lambda} t) \\ &= \frac{\varepsilon}{n} \left(1 - \frac{\lambda}{n}\right)^{-1} P(1 \leftrightarrow_{\lambda} s, t \leftrightarrow_{\lambda} 2, s \leftrightarrow_{\lambda+\varepsilon} t, 1 \not\leftrightarrow_{\lambda} 2). \end{aligned}$$

Daraus lässt sich schließen:

$$\begin{aligned} \tau'_n(\lambda) &= \frac{1}{n} \left(1 - \frac{\lambda}{n}\right)^{-1} \sum_{s < t=1}^n P_{\lambda}(1 \leftrightarrow s, t \leftrightarrow 2, 1 \not\leftrightarrow 2) \\ &\leq \frac{1}{n} \left(1 - \frac{\lambda}{n}\right)^{-1} \sum_{s < t=1}^n E_{\lambda} \left( \mathbb{1}_{\{1 \leftrightarrow s\}} \mathbb{1}_{\{t \leftrightarrow 2\}} \mathbb{1}_{\{1 \not\leftrightarrow 2\}} \right) \\ &\leq \frac{1}{n} \left(1 - \frac{\lambda}{n}\right)^{-1} E_{\lambda} \left( |C_{(1)}| |C_{(2)}| \mathbb{1}_{\{1 \not\leftrightarrow 2\}} \right). \end{aligned}$$

Kennt man nun die Größe der ersten Komponente, so liefert Bemerkung 2.8 für zufällige Graphen eine Verteilung des Graphen ohne diese Komponente, so

dass

$$\begin{aligned}
 & E_\lambda (|C_{(1)}||C_{(2)}|\mathbb{1}_{\{1 \leftrightarrow 2\}}) \\
 &= E_\lambda \left( \sum_{l=1}^n l|C_{(2)}|\mathbb{1}_{\{1 \leftrightarrow 2\}}\mathbb{1}_{\{|C_{(1)}|=l\}} \right) \\
 &= \sum_{l=1}^n l E_\lambda (E_\lambda (|C_{(2)}|\mathbb{1}_{\{1 \leftrightarrow 2\}}\mathbb{1}_{\{|C_{(1)}|=l\}}|\mathbb{1}_{\{1 \leftrightarrow 2\}}\mathbb{1}_{\{|C_{(1)}|=l\}})) \\
 &= \sum_{l=1}^n l E_\lambda (\mathbb{1}_{\{1 \leftrightarrow 2\}}\mathbb{1}_{\{|C_{(1)}|=l\}} E_\lambda (|C_{(2)}|\mathbb{1}_{\{1 \leftrightarrow 2\}}\mathbb{1}_{\{|C_{(1)}|=l\}})) \\
 &= \sum_{l=1}^n l E_\lambda (\mathbb{1}_{\{1 \leftrightarrow 2\}}\mathbb{1}_{\{|C_{(1)}|=l\}}) \chi_{n-l}(\lambda \frac{n-l}{n}) \\
 &\leq \sum_{l=1}^n l P_\lambda(|C_{(1)}|=l) \chi_{n-l}(\lambda \frac{n-l}{n}).
 \end{aligned}$$

Zur Abschätzung dieser Summe unterscheidet man zwischen den kleinen und den großen Werten von  $l$ . Dazu wählt man ein  $a \in (0, 1)$  und betrachtet zunächst die Summe mit Indizes kleiner als  $a^2 n^{\frac{2}{3}}$ . Für solche  $l$  ist der Fehler der Abschätzung

$$\begin{aligned}
 \chi_{n-1}(\lambda \frac{n-l}{n}) &\leq \chi_n(\lambda \frac{n-l}{n}) \\
 &\leq \chi_n(\lambda) \\
 &\leq \chi_n(1)
 \end{aligned}$$

vergleichsweise klein, so dass man hoffen kann, in diesem Teil der Summe nicht

zu große Fehler zu begehen:

$$\begin{aligned}
 \sum_{l=1}^{a^2 n^{\frac{2}{3}}} l P_{\lambda}(|C_{(1)}| = l) \chi_{n-l} \left( \lambda \frac{n-l}{n} \right) &\leq \chi_n(1) \sum_{l=1}^{a^2 n^{\frac{2}{3}}} l P_{\lambda}(|C_{(1)}| = l) \\
 &= \chi_n(1) \sum_{l=1}^{a^2 n^{\frac{2}{3}}} \sum_{i=1}^l P_{\lambda}(|C_{(1)}| = l) \\
 &= \chi_n(1) \sum_{l=1}^{a^2 n^{\frac{2}{3}}} \sum_{i=1}^l P_{\lambda}(|C_{(1)}| = l) \\
 &\leq \chi_n(1) \sum_{i=1}^{a^2 n^{\frac{2}{3}}} \sum_{l=i}^n P_{\lambda}(|C_{(1)}| = l) \\
 &= \chi_n(1) P_{\lambda}(|C_{(1)}| \geq a^2 n^{\frac{2}{3}}).
 \end{aligned}$$

An dieser Stelle liefert Lemma 2.33 zusammen mit dem oben bereits verwendeten Integralvergleich die Existenz einer Konstante  $C > 0$ , so dass

$$\begin{aligned}
 \chi_n(1) \sum_{i=1}^{a^2 n^{\frac{2}{3}}} P_{\lambda}(|C_{(1)}| \geq l) &\leq \chi_n(1) \sum_{i=1}^{a^2 n^{\frac{2}{3}}} \frac{C}{2\sqrt{i}} \\
 &\leq \chi_n(1) \frac{C}{2} \int_0^{a^2 n^{\frac{2}{3}}} \frac{1}{\sqrt{i}} di \\
 &= \chi_n(1) C a n^{\frac{1}{3}},
 \end{aligned}$$

was einen der in der Behauptung geforderten Terme darstellt.

Für Indizes die nicht kleiner als  $a^2 n^{\frac{2}{3}}$  sind, erlaubt Korollar 2.32 zusammen mit

$$\frac{1}{1 - \lambda \frac{n-l}{n}} = \frac{n}{(1 - \lambda)n + \lambda l} \leq \frac{n}{l}$$



folgende Umformungen:

$$\begin{aligned}
 \sum_{l=a^2 n^{\frac{2}{3}}}^n l P_\lambda(|C_{(1)}| = l) \chi_{n-l}(\lambda \frac{n-l}{n}) &\leq \sum_{l=a^2 n^{\frac{2}{3}}}^n l P_\lambda(|C_{(1)}| = l) \chi_n(\lambda \frac{n-l}{n}) \\
 &\leq \sum_{l=a^2 n^{\frac{2}{3}}}^n l P_\lambda(|C_{(1)}| = l) \frac{1}{1 - \lambda \frac{n-l}{n}} \\
 &\leq n \sum_{l=a^2 n^{\frac{2}{3}}}^n P_\lambda(|C_{(1)}| = l) \\
 &\leq n P_\lambda(|C_{(1)}| \geq a^2 n^{\frac{2}{3}}).
 \end{aligned}$$

Hier folgert man analog zur obigen Argumentation mit Hilfe von Lemma 2.33

$$\begin{aligned}
 n P_\lambda(|C_{(1)}| \geq a^2 n^{\frac{2}{3}}) &\leq n P_1(|C_{(1)}| \geq a^2 n^{\frac{2}{3}}) \\
 &\leq n \frac{C'}{a} n^{-\frac{1}{3}}.
 \end{aligned}$$

Damit ergibt sich die Behauptung. □



### 3 Das Erscheinen der größten Komponente

Ein bemerkenswertes Phänomen beim Betrachten des  $G(n, p)$ -Modells zufälliger Graphen ist die abrupte Veränderung der asymptotischen Größe der größten Komponente des Graphen, wenn  $\lambda = np$  um 1 herum variiert.

Ist  $\lambda < 1$ , so sind die großen Komponenten des Graphen klein gegen seine Gesamtgröße. Satz 3.1 zeigt hier eine Größenordnung von  $\ln n$ . Für  $\lambda = 1$  wächst die größte Komponente mit positiver Wahrscheinlichkeit deutlich schneller, nämlich so wie  $n^{\frac{2}{3}}$ , wie in Satz 3.13 gezeigt wird. Erst im Fall  $\lambda > 1$  nimmt die größte Komponente einen signifikanten Anteil der Gesamtgröße des Graphen ein. Setzt man das Auftreten einer Komponente mit Größenordnung  $n$  mit dem Überleben eines ähnlichen Galton-Watson-Prozesses mit poissonverteilten Zuwächsen gleich, so erbt dieses Modell also in gewissem Sinne damit über die in Abschnitt 2.2.1 beschriebenen Mechanismen das bekannte (z. B. [Als02]) Verhalten dieses Verzweigungsprozesses nur im superkritischen Fall eine positive Überlebenswahrscheinlichkeit zu haben.

Dieses Verhalten von zufälligen Graphen wurde erstmals in [ER60] und [ER61] untersucht. In diesem Kapitel soll dieses Phänomen an Hand schwacher Gesetze der großen Zahlen im Fall  $\lambda \neq 1$  dargestellt werden. Für  $\lambda = 1$  gilt kein solches Gesetz; stattdessen wird eine Abschätzung der Wahrscheinlichkeit einer Größe der größten Komponente im Bereich  $n^{\frac{2}{3}}$  gezeigt. Die Beweise sind an das Vorgehen in [Hof09] angelehnt.

### 3.1 Der subkritische Fall: $\lambda < 1$

Wie man durch den Vergleich mit den oben genannten Verzweigungsprozessen erwartet, gilt für die größte Komponente eines zufälligen Graphen im Fall  $\lambda < 1$

$$\frac{|C_{\max}|}{n} \rightarrow 0.$$

Tatsächlich liefert erst eine Skalierung mit  $\ln n$  einen nicht trivialen Grenzwert.

**Satz 3.1.** *Für  $\lambda < 1$  und  $I_\lambda = \lambda - 1 - \ln \lambda$  gilt*

$$\frac{|C_{\max}|}{\ln n} \longrightarrow I_\lambda^{-1} \quad \text{stochastisch.}$$

Wesentlicher Bestandteil des Beweises hier und im Rest des Kapitels in die Untersuchung der Anzahl der Knoten in großen Komponenten  $Z_{>k}$ . Wie bereits in der Einleitung zu Abschnitt 2.3 erwähnt, erlaubt die Äquivalenz

$$|C_{\max}| > k \iff Z_{>k} > 0$$

einen leichteren Zugang zu der schwerer zu untersuchenden Größe  $|C_{\max}|$ . Um Aussagen über  $Z_{>k}$  herzuleiten, verwendet man die leichte Darstellbarkeit von  $\mathbb{E}_\lambda Z_{>k}$  durch die Tails der Größe der ersten Komponente via

$$\mathbb{E}_\lambda Z_{>k} = nP_\lambda(|C_{(1)}| > k),$$

die in Abschnitt 2.3 mit Hilfe der Methoden aus Kapitel 2 untersucht wurden. Technisch zerfällt der Beweis des schwachen Gesetzes der großen Zahlen in diesem Fall in die beiden Lemmata.

**Lemma 3.2.** *Für jedes  $a > I_\lambda^{-1}$  ist*

$$\lim_{n \rightarrow \infty} P_\lambda(|C_{\max}| > a \ln n) = 0.$$

**Beweis:**

Die Ungleichung von Markow zusammen mit der Abschätzung aus Korollar 2.29 liefert für  $k = a \ln n$

$$\begin{aligned}
 P_\lambda(|C_{\max}| > a \ln n) &= P_\lambda(Z_{>a \ln n} > 0) \\
 &\leq \mathbb{E}_\lambda Z_{>a \ln n} \\
 &= n P_\lambda(|C_{(1)}| > a \ln n) \\
 &\leq n \frac{1}{\lambda} e^{-a I_\lambda \ln n} \\
 &= \frac{1}{\lambda} e^{-(a I_\lambda - 1) \ln n},
 \end{aligned}$$

was wegen  $a > I_\lambda^{-1}$  und somit  $a I_\lambda - 1 > 0$  gegen 0 strebt. □

**Lemma 3.3.** *Für jedes  $a < I_\lambda^{-1}$  ist*

$$\lim_{n \rightarrow \infty} P_\lambda(|C_{\max}| < a \ln n) = 0.$$

Die Abschätzung der Wahrscheinlichkeit  $P_\lambda(|C_{\max}| < a \ln n)$  in diesem Lemma erfolgt mit Hilfe der Ungleichung von Tschebyschow über eine Varianzabschätzung für die Zufallsgröße  $Z_{>k}$ . Diese Abschätzung wird hier vorgezogen.

**Lemma 3.4.** *Für jedes  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{N}$ , so dass für jedes  $k > K$*

$$\text{Var}_\lambda Z_{>k} \leq n \frac{1}{\lambda} \frac{e^{-(I_\lambda - \varepsilon)k}}{1 - e^{-(I_\lambda - \varepsilon)}}.$$

Da ein Zwischenschritt aus dem Beweis dieses Lemmas im späteren Abschnitt 3.3 noch einmal Verwendung findet, zeigt man zunächst:

**Lemma 3.5.** *Für jedes  $\lambda > 0$  gilt*

$$\text{Var}_\lambda Z_{>k} \leq n \mathbb{E}_\lambda |C_{(1)}| \mathbb{1}_{\{|C_{(1)}| > k\}}.$$

**Beweis:** Nach Definition ist

$$\begin{aligned}
 \text{Var } Z_{>k} &= \text{Var} \sum_{v=1}^n \mathbb{1}_{\{|C_{(v)}|>k\}} \\
 &= \sum_{i,j=1}^n P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k) - P_\lambda(|C_{(i)}| > k)P_\lambda(|C_{(j)}| > k) \\
 &= \sum_{i,j=1}^n P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k) - P_\lambda(|C_{(i)}| > k)P_\lambda(|C_{(j)}| > k) \\
 &= \sum_{i,j=1}^n P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k, i \not\leftrightarrow j) - P_\lambda(|C_{(i)}| > k)P_\lambda(|C_{(j)}| > k) \\
 &\quad + \sum_{i,j=1}^n P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k, i \leftrightarrow j).
 \end{aligned}$$

Um die Abhängigkeit der Ereignisse  $\{|C_{(i)}| > k\}$  und  $\{|C_{(j)}| > k\}$  in den Griff zu bekommen, verwendet man im Fall  $i \not\leftrightarrow j$  das Dualitätsprinzip aus Bemerkung 2.8, so dass

$$\begin{aligned}
 &P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k, i \not\leftrightarrow j) \\
 &= \sum_{l=k+1}^n P_\lambda(|C_{(j)}| > k \mid |C_{(i)}| = l, i \not\leftrightarrow j)P_\lambda(|C_{(i)}| = l, i \not\leftrightarrow j) \\
 &\leq \sum_{l=k+1}^n P_\lambda(|C_{(j)}| > k)P_\lambda(|C_{(i)}| = l) \\
 &= P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k).
 \end{aligned}$$

Damit ist

$$\begin{aligned}
 \text{Var } Z_{>k} &\leq \sum_{i,j=1}^n P_\lambda(|C_{(i)}| > k, |C_{(j)}| > k) \\
 &= \sum_{i,j=1}^n P_\lambda(|C_{(i)}| > k, j \in C_{(i)}) \\
 &= \sum_{i=1}^n \mathbb{E}_\lambda \mathbb{1}_{\{|C_{(i)}|>k\}} \sum_{j=1}^n \mathbb{1}_{\{j \in C_{(i)}\}} \\
 &= \sum_{i=1}^n \mathbb{E}_\lambda |C_{(i)}| \mathbb{1}_{\{|C_{(i)}|>k\}} \\
 &= n \mathbb{E}_\lambda |C_{(1)}| \mathbb{1}_{\{|C_{(1)}|>k\}}.
 \end{aligned}$$

□

**Beweis von Lemma 3.4:** Ausgehend von der vorhergehenden Aussage ergibt sich

$$\begin{aligned} \text{Var } Z_{>k} &\leq n \mathbb{E}_\lambda |C_{(1)}| \mathbb{1}_{\{|C_{(1)}|>k\}} \\ &= n \sum_{l=k+1}^n l P_\lambda(|C_{(1)}| = l) \\ &\leq n \sum_{l=k}^n (l+1) P_\lambda(|C_{(1)}| > l) \\ &\leq n \sum_{l=k}^n (l+1) \frac{1}{\lambda} e^{-I_\lambda l}, \end{aligned}$$

wegen Korollar 2.29. Für jedes  $\varepsilon > 0$  gibt es dann ein  $K \in \mathbb{N}$ , so dass für alle  $l > K$

$$\frac{1}{\lambda}(l+1) < e^{\varepsilon l}.$$

Damit gilt

$$\begin{aligned} \frac{1}{\lambda} \sum_{l=k}^n (l+1) e^{-I_\lambda l} &\leq \sum_{l=k}^{\infty} e^{-(I_\lambda - \varepsilon)l} \\ &= \frac{e^{-(I_\lambda - \varepsilon)k}}{1 - e^{-(I_\lambda - \varepsilon)}}. \end{aligned}$$

□

**Beweis von Lemma 3.3:** Der Beweis dieses Lemmas verwendet die Ungleichung von Tschebyschow für die Abschätzung von

$$\begin{aligned} P_\lambda(|C_{\max}| < a \ln n) &= P_\lambda(Z_{>a \ln n} = 0) \\ &\leq P_\lambda(|Z_{>a \ln n} - \mathbb{E}_\lambda Z_{>a \ln n}| < \mathbb{E}_\lambda Z_{>a \ln n}) \\ &\leq \frac{\text{Var}_\lambda Z_{>a \ln n}}{\mathbb{E}_\lambda^2 Z_{>a \ln n}}. \end{aligned}$$

### 3 Das Erscheinen der größten Komponente

---

Sei nun  $\varepsilon > 0$ . Korollar 2.29 und Lemma 3.4 liefert die Existenz eines  $N \in \mathbb{N}$  so, dass für alle  $n > N$

$$\begin{aligned} \text{Var } Z_{>a \ln n} &\leq n \frac{e^{-(I_\lambda - \varepsilon)a \ln n}}{1 - e^{-(I_\lambda - \varepsilon)}} \\ \text{und } \mathbb{E}_\lambda Z_{>a \ln n} &= n P_\lambda(|C_{(1)}| > a \ln n) \\ &\geq n \frac{1}{\lambda_n} \frac{e^{-(I_\lambda + \varepsilon)a \ln n}}{1 - e^{-(I_\lambda + \varepsilon)}} + \mathcal{O}(1). \end{aligned}$$

Mit Hilfe dieser Ungleichungen folgt

$$\begin{aligned} \frac{\text{Var } {}_\lambda Z_{>a \ln n}}{\mathbb{E}_\lambda^2 Z_{>a \ln n}} &\leq \frac{\lambda_n^2 e^{-(I_\lambda - \varepsilon)a \ln n}}{n e^{-2(I_\lambda + \varepsilon)a \ln n}} \frac{1 - e^{-(I_\lambda + \varepsilon)}}{1 - e^{-(I_\lambda - \varepsilon)}} + \mathcal{O}(1) \\ &= \lambda_n^2 \frac{1 - e^{-(I_\lambda + \varepsilon)}}{1 - e^{-(I_\lambda - \varepsilon)}} e^{(I_\lambda + 3\varepsilon - 1)a \ln n}. \end{aligned}$$

Wählt man  $\varepsilon$  klein genug, um

$$(I_\lambda + 3\varepsilon)a < 1$$

sicherzustellen, was wegen  $a < I_\lambda^{-1}$  möglich ist, so strebt dieser Ausdruck mit wachsendem  $n$  gegen 0.  $\square$

**Beweis von Satz 3.1:** Der Beweis ist mit den Lemmata 3.2 und 3.3 erbracht, denn für jedes  $\varepsilon > 0$  ist

$$\begin{aligned} &\lim_{n \rightarrow \infty} P_\lambda \left( \left| \frac{1}{\ln n} |C_{\max}| - I_\lambda^{-1} \right| > \varepsilon \right) \\ &\leq \lim_{n \rightarrow \infty} P_\lambda(|C_{\max}| < (I_\lambda^{-1} - \varepsilon) \ln n) + \lim_{n \rightarrow \infty} P_\lambda(|C_{\max}| > (I_\lambda^{-1} + \varepsilon) \ln n) \\ &= 0, \end{aligned}$$

was zu zeigen war.  $\square$



### 3.2 Der superkritische Fall: $\lambda > 1$

In diesem Abschnitt wird der superkritische Fall untersucht. Der Galton-Watson-Prozess mit  $Poi_\lambda$  hat eine positive Wahrscheinlichkeit  $\zeta_\lambda$  nicht auszusterben, was ebenfalls – nicht überraschend – die relative Anzahl Knoten in der größten Komponente ist.

**Satz 3.6.** *Sei  $\lambda > 1$  und  $\zeta_\lambda$  die Überlebenswahrscheinlichkeit eines Galton-Watson-Prozesses mit  $Poi_\lambda$ -verteilten Zuwächsen. Dann gilt für jedes  $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} P_\lambda \left( \left| \frac{1}{n} |C_{\max}| - \zeta_\lambda \right| > \varepsilon \right) = 0.$$

Statt diesen Satz direkt für  $|C_{\max}|$  zu zeigen, wird wie im subkritischen Fall zunächst die Anzahl der Knoten in großen Komponenten

$$Z_{>k_n} := \sum_{v=1}^n \mathbb{1}_{\{|C_{(v)}| > k_n\}}$$

untersucht und das Gesetz der großen Zahlen für  $Z_{>k_n}$  klassisch über eine Varianzabschätzung und die Ungleichung von Tschebyschow bewiesen.

Um aus einem schwachen Gesetz der großen Zahlen für  $Z_{>k}$  auf ein solches für  $|C_{\max}|$  schließen zu können, ist es nötig zu zeigen, dass sich  $Z_{>k}$  und  $|C_{\max}|$  nur mit kleiner Wahrscheinlichkeit voneinander unterscheiden. Falls es mehrere große Komponenten im Graphen gäbe, so wäre  $Z_{>k}$ , wenn  $k$  klein genug ist um die großen Komponenten zu erfassen, wesentlich größer als  $|C_{\max}|$ . Um diesen Fall auszuschließen, zeigen die folgenden zwei Aussagen, dass es nur mit exponentiell kleiner Wahrscheinlichkeit zwei Komponenten der Größenordnung  $n$  gibt.

**Lemma 3.7.** *Sei  $\lambda > 1$  und  $\lim_{n \rightarrow \infty} k_n = \infty$ . Dann existieren für jedes  $\alpha < \zeta_\lambda$  ein  $C > 0$  und ein  $J > 0$ , so dass*

$$P_\lambda(k_n < |C_{(1)}| \leq \alpha n) \leq C e^{-k_n J}.$$

Im Beweis verwendet man die Darstellung von  $|C_{(1)}|$  durch den Explorationsprozess  $S_t$  aus Abschnitt 2.1.4. Weiter benötigt man eine Aussage über  $S_t$ , die in leicht allgemeinerer Form im Kapitel 4 noch einmal verwendet wird. Sie wird deswegen hier vorgezogen.

**Lemma 3.8.** *Für  $\lambda > 1$  und  $\alpha < \zeta_\lambda$  gibt es ein  $\delta > 0$ , so dass für den Explorationsprozess  $(S_t)_{t \in \mathbb{N}}$  mit Start in  $m$  Knoten für alle  $t < \alpha n$*

$$P_\lambda(S_t = 0) \leq e^{-(t+m)\frac{\delta^2}{2\lambda}}.$$

**Beweis:** Für  $t < \alpha n$  stellt man zunächst  $P_\lambda(S_t = 0)$  mit Hilfe von Lemma 2.12 dar:

$$\begin{aligned} P_\lambda(S_t = 0) &= \mathfrak{B}_{n-m, 1 - (1 - \frac{\lambda}{n})^{t+m}}(t+m) \\ &\leq P_\lambda(X \leq t+m), \end{aligned}$$

für eine Zufallsgröße  $X \sim \mathfrak{B}_{n, 1 - \exp(-\frac{\lambda}{n}(t+m))}$ . Um hier eine obere Schranke mit Hilfe der Chernoff-Schranke (z. B. [Hof09, Theorem 2.18]) angeben zu können, benötigt man eine Abschätzung von  $\mathbb{E}X$ . Für jedes  $\alpha < \zeta_\lambda$  gibt es ein  $\delta > 0$ , so dass

$$1 - \lambda\alpha < e^{-\lambda\alpha} < 1 - (1 + \delta)\alpha.$$

Es folgt

$$\begin{aligned} \mathbb{E}X &= n \left( 1 - e^{-\lambda \frac{t+m}{n}} \right) \\ &\leq n \left( 1 - \left( 1 - \lambda \frac{t+m}{n} \right) \right) \\ &= \lambda(t+m) \end{aligned}$$

und

$$\begin{aligned}\mathbb{E}X &= n \left(1 - e^{-\lambda \frac{t+m}{n}}\right) \\ &\geq n(1 + \delta) \frac{t+m}{n} \\ &= (1 + \delta)(t+m).\end{aligned}$$

Damit liefert die Chernoff-Schranke die Abschätzung

$$\begin{aligned}P_\lambda(S_t = 0) &\leq P_\lambda(X \leq t+m) \\ &\leq P_\lambda(X \leq \mathbb{E}X - \delta(t+m)) \\ &\leq \exp\left(-\delta^2(t+m)^2 \frac{1}{2\mathbb{E}X}\right) \\ &\leq \exp\left(-\frac{\delta^2(t+m)}{2\lambda}\right),\end{aligned}$$

was zu beweisen war. □

**Beweis von Lemma 3.7:** Es gilt  $|C_{(1)}| = \min\{t \in N : S_t = 0\}$  und damit

$$\begin{aligned}P_\lambda(k_n < |C_{(1)}| \leq \alpha n) &= \sum_{t=k_n+1}^{\alpha n} P_\lambda(|C_{(1)}| = t) \\ &\leq \sum_{t=k_n}^{\alpha n} P_\lambda(S_t = 0).\end{aligned}$$

Lemma 3.8 liefert mit  $m = 1$  dann

$$P_\lambda(S_t \leq 0) \leq e^{-\frac{t\delta^2}{2\lambda}},$$

für ein kleines  $\delta > 0$ . Abschließend ist, wenn  $J = \frac{\delta^2}{2\lambda} > 0$  gewählt wird,

$$\begin{aligned} P_\lambda(k_n < |C_{(1)}| \leq \alpha n) &\leq \sum_{t=k_n}^{\alpha n} P_\lambda(S_t = 0) \\ &\leq \sum_{t=k_n}^{\alpha n} e^{-tJ} \\ &\leq e^{-k_n J} \sum_{t=0}^{\infty} e^{-tJ} \\ &\leq e^{-k_n J} \int_0^{\infty} e^{-(t-1)J} dt, \end{aligned}$$

was zu zeigen war. □

**Korollar 3.9.** Für  $\alpha < \zeta_\lambda$  und  $k_n = K \ln n$  mit hinreichend großem  $K$  gibt es ein  $\delta > 0$ , so dass

$$P_\lambda(\exists v : k_n < |C_{(v)}| \leq \alpha n) \in \mathcal{O}(n^{-\delta}).$$

**Beweis:** Die Markow-Ungleichung liefert

$$\begin{aligned} P_\lambda(\exists v : k_n < |C_{(v)}| \leq \alpha n) &= P_\lambda(Z_{>k_n} - Z_{>\alpha n} \geq 1) \\ &\leq \mathbb{E}(Z_{>k_n} - Z_{>\alpha n}) \\ &= nP_\lambda(k_n < |C_{(1)}| \leq \alpha n) \\ &\leq nCe^{-k_n J} \in \mathcal{O}(n^{-\delta}), \end{aligned}$$

für  $K$  groß genug. □

**Lemma 3.10.** Für jedes  $n \in \mathbb{N}$  und  $1 \leq k \leq n$  ist

$$\text{Var}(Z_{>k}) \leq (\lambda k + 1)nk.$$

**Beweis:** Zunächst stellt man fest, dass für

$$Z_{\leq k} = \sum_{v=1}^n \mathbb{1}_{\{|C_{(v)}| \leq k\}}$$

gilt  $Z_{\leq k} = n - Z_{>k}$  und deswegen

$$\text{Var}(Z_{\leq k}) = \text{Var}(Z_{>k}).$$

Die Betrachtung von  $Z_{\leq k}$  ergibt in dieser Situation eine leichtere Abschätzung.

Schreibt man die Varianz aus und teilt die Summe dann auf nach verbundenen und nicht verbundenen Knoten, so erhält man

$$\begin{aligned} \text{Var}(Z_{\leq k}) &= \text{Var}\left(\sum_{v=1}^n \mathbb{1}_{\{|C(v)| \leq k\}}\right) \\ &= \sum_{i,j=1}^n P_\lambda(|C(i)| \leq k, |C(j)| \leq k) - P_\lambda(|C(i)| \leq k)P_\lambda(|C(j)| \leq k) \\ &= \sum_{i,j=1}^n P_\lambda(|C(i)| \leq k, |C(j)| \leq k, i \not\leftrightarrow j) - P_\lambda(|C(i)| \leq k)P_\lambda(|C(j)| \leq k) \\ &\quad + \sum_{i,j=1}^n P_\lambda(|C(i)| \leq k, |C(j)| \leq k, i \leftrightarrow j). \end{aligned}$$

Für  $i \leftrightarrow j$  ist natürlich  $C(i) = C(j)$ , so dass der letzte Term leicht zugänglich ist:

$$\begin{aligned} \sum_{i,j=1}^n P_\lambda(|C(i)| \leq k, |C(j)| \leq k, i \leftrightarrow j) &= \sum_{i,j=1}^n P_\lambda(|C(i)| \leq k, i \leftrightarrow j) \\ &= \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{|C(i)| \leq k\}} \sum_{j=1}^n \mathbb{1}_{i \leftrightarrow j}) \\ &= \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{|C(i)| \leq k\}} |C(i)|) \\ &= nk. \end{aligned}$$

Da alle auftretenden Wahrscheinlichkeiten positiv sind, kann man zur Betrachtung des ersten Terms die Summanden mit Hilfe der Definition bedingter Wahr-

scheinlichkeiten umschreiben zu

$$\begin{aligned}
 & P_\lambda(|C_{(i)}| \leq k, |C_{(j)}| \leq k, i \not\leftrightarrow j) \\
 &= \sum_{l=1}^{k-1} P_\lambda(|C_{(i)}| = l, |C_{(j)}| \leq k, i \not\leftrightarrow j) \\
 &= \sum_{l=1}^{k-1} \left( P_\lambda(|C_{(i)}| = l) P_\lambda(i \not\leftrightarrow j \mid |C_{(i)}| = l) P_\lambda(|C_{(i)}| = l) \right. \\
 &\quad \left. \cdot P_\lambda(i \leftrightarrow j \mid |C_{(i)}| = l) P_\lambda(|C_{(j)}| \leq k \mid |C_{(i)}| = l, i \leftrightarrow j) \right) \\
 &\leq \sum_{l=1}^{k-1} P_\lambda(|C_{(i)}| = l) P_\lambda(|C_{(j)}| \leq k \mid |C_{(i)}| = l, i \leftrightarrow j).
 \end{aligned}$$

Die Verteilung der letzten Faktoren ist mit der Bemerkung 2.8 zugänglich. So ist

$$\begin{aligned}
 & P_{n,\lambda}(|C_{(j)}| \leq k \mid |C_{(i)}| = l, i \not\leftrightarrow j) \\
 &= P_{n-l,\lambda}(|C_{(1)}| \leq k) \\
 &= P_{n-l,\lambda}(|C_{(1)}| \leq k) - P_{n,\lambda}(|C_{(1)}| \leq k) + P_{n,\lambda}(|C_{(1)}| \leq k).
 \end{aligned}$$

$P_{n,\lambda}$  lässt sich nun auf folgende Weise darstellen: Startend mit  $P_{n-l,\lambda}$  fügt man die fehlenden  $l$  Knoten hinzu und jede hinzugefügte Kante wird unabhängig von den anderen realisiert mit Wahrscheinlichkeit  $\frac{\lambda}{n}$ . Bezeichne  $C_{(1)}^*$  die Zusammenhangskomponente des Knotens 1, eingeschränkt auf die ersten  $n-l$  Knoten, so ergibt sich

$$\begin{aligned}
 P_{n-l,\lambda}(|C_{(1)}| \leq k) - P_{n,\lambda}(|C_{(1)}| \leq k) &= P_\lambda(|C_{(1)}^*| \leq k) - P_\lambda(|C_{(1)}| \leq k) \\
 &= P_\lambda(|C_{(1)}^*| \leq k, |C_{(1)}| \geq k).
 \end{aligned}$$

Dieses Ereignis kann mit der obigen Konstruktion nur auftreten, wenn einer der

$l$  hinzugefügten Knoten mit einem Knoten aus  $C_{(1)}^*$  verbunden wird, was mit einer Wahrscheinlichkeit kleiner als  $\frac{k l \lambda}{n}$  auftritt, so dass

$$P_\lambda(|C_{(j)}| \leq k \mid |C_{(i)}| = l, i \not\leftrightarrow j) \leq \frac{k l \lambda}{n} + P_\lambda(|C_{(j)}| \leq k).$$

Zusammengesetzt ergibt sich

$$\begin{aligned} & \sum_{i,j=1}^n P_\lambda(|C_{(i)}| \leq k, |C_{(j)}| \leq k, i \not\leftrightarrow j) - P_\lambda(|C_{(i)}| \leq k)P_\lambda(|C_{(j)}| \leq k) \\ &= \sum_{i,j=1}^n \sum_{l=1}^{k-1} P_\lambda(|C_{(i)}| = l, |C_{(j)}| \leq k, i \not\leftrightarrow j) - P_\lambda(|C_{(i)}| \leq k)P_\lambda(|C_{(j)}| \leq k) \\ &\leq \sum_{i,j=1}^n \sum_{l=1}^{k-1} P_\lambda(|C_{(i)}| = l) \left( \frac{k l \lambda}{n} + P_\lambda(|C_{(j)}| \leq k) \right) - P_\lambda(|C_{(i)}| \leq k)P_\lambda(|C_{(j)}| \leq k) \\ &= \sum_{i,j=1}^n \sum_{l=1}^{k-1} \frac{k l \lambda}{n} P_\lambda(|C_{(i)}| = l) + \sum_{l=1}^{k-1} P(|C_{(i)}| = l)P_\lambda(|C_{(j)}| \leq k) \\ &\quad - P_\lambda(|C_{(i)}| \leq k)P_\lambda(|C_{(j)}| \leq k) \\ &= \frac{k \lambda}{n} \sum_{l=1}^{k-1} l \sum_{i,j=1}^n P_\lambda(|C_{(1)}| = l) \\ &= k \lambda j \sum_{l=1}^{k-1} l P_\lambda(|C_{(1)}| = l) \\ &= k \lambda n k \end{aligned}$$

und damit

$$\text{Var } Z_{\leq k} \leq (\lambda k + 1)nk,$$

was den Beweis abschließt.  $\square$

**Beweis von Satz 3.6:** Zunächst wird, wie angekündigt, ein schwaches Gesetz der großen Zahlen für  $Z_{>k_n}$  gezeigt. Sei also  $\varepsilon > 0$  und  $k_n = K \ln n$  für ein

hinreichend großes  $K$ . Lemma 2.28 liefert, wenn  $n$  so groß gewählt wird, dass  $\frac{k_n}{n} < \varepsilon$ , mit Hilfe der Tschebyschow-Ungleichung und Lemma 3.10:

$$\begin{aligned} P_\lambda\left(\frac{1}{n}|Z_{>k_n} - \zeta_\lambda| > 2\varepsilon\right) &\leq P_\lambda\left(\frac{1}{n}|Z_{>k_n} - \mathbb{E}Z_{>k_n}| > \varepsilon\right) \\ &\leq \frac{1}{n^2\varepsilon} \text{Var } Z_{>k_n} \\ &\leq \frac{1}{n\varepsilon}(\lambda k_n + 1)k_n, \end{aligned}$$

was für  $n$  gegen  $\infty$  wie gewünscht gegen 0 strebt.

Lässt sich jetzt zeigen, dass  $\lim_{n \rightarrow \infty} P_\lambda(Z_{>k_n} \neq |C_{\max}|) = 0$  gilt, so ergibt sich wegen

$$\begin{aligned} P_\lambda(|C_{\max}| > \varepsilon) &= P_\lambda(|C_{\max}| > \varepsilon, Z_{>k_n} \neq |C_{\max}|) + P_\lambda(Z_{>k_n} > \varepsilon, Z_{>k_n} = |C_{\max}|) \\ &\leq P_\lambda(Z_{>k_n} \neq |C_{\max}|) + P_\lambda(Z_{>k_n} > \varepsilon) \end{aligned}$$

bereits die  $P$ -stochastische Konvergenz von  $|C_{\max}|$  gegen  $\zeta_\lambda$ , die hier bewiesen werden soll. Dafür betrachtet man für ein kleines  $\varepsilon > 0$

$$P_\lambda(Z_{>k_n} \neq |C_{\max}|) \leq P_\lambda(Z_{>k_n} \neq |C_{\max}|, \frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon) + o(1).$$

Auf der Menge  $\{\frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon\}$  ist  $Z_{>k_n}$  sicher größer als  $k_n$  und deswegen ist hier  $|C_{\max}| \leq Z_{>k_n}$ , was die Umformung

$$P_\lambda(Z_{>k_n} \neq |C_{\max}|, \frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon) = P_\lambda(Z_{>k_n} > |C_{\max}|, \frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon)$$

erlaubt. Damit  $|C_{\max}| < Z_{>k_n}$  gelten kann, muss es mindestens zwei Komponenten mit Größe über  $k_n$  geben. Da durch Korollar 3.9 gesichert ist, dass es mit hoher Wahrscheinlichkeit keine Zusammenhangskomponenten der Größe



zwischen  $k_n$  und  $\alpha n$  für ein  $\alpha < \zeta_\lambda$ ,  $2\alpha > \zeta_\lambda + \varepsilon$  gibt, gilt

$$\begin{aligned} & P_\lambda(Z_{>k_n} > |C_{\max}|, \frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon) \\ &= P_\lambda(Z_{>k_n} > 2\alpha n, \frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon) + o(1) \\ &\leq P_\lambda(Z_{>k_n} > \zeta_\lambda + \varepsilon, \frac{1}{n}|Z_{>k_n} - \zeta_\lambda| \leq \varepsilon) + o(1) \\ &= o(1) \end{aligned}$$

und damit

$$P_\lambda(Z_{>k_n} \neq |C_{\max}|) = o(1),$$

was zu zeigen war. □

### 3.2.1 Die Größe der zweitgrößten Komponente

Das Dualitätsprinzip (Abschnitt 2.1.3) erlaubt es jetzt zusammen mit den Resultaten der beiden vorhergehenden Abschnitte das Verhalten der zweitgrößten Komponente im superkritischen Fall zu untersuchen. Entfernt man nämlich die größte Komponente aus dem Graphen, so ist der Rest ähnlich wie ein subkritischer Graph mit Parameter  $\lambda(1 - \zeta_\lambda)$  verteilt und dementsprechend gilt:

**Satz 3.11.** *Bezeichne  $C'$  die zweitgrößte Komponente eines zufälligen Graphen und sei  $\lambda > 1$ . Dann ist*

$$\frac{|C'|}{\ln n} \rightarrow I_{\lambda(1-\zeta_\lambda)}^{-1} \quad \text{stochastisch.}$$

Zunächst ist zu klären, dass  $\lambda(1 - \zeta_\lambda)$  tatsächlich einen subkritischen Graphen beschreibt.

**Lemma 3.12.** *Für  $\lambda > 1$  ist*

$$\lambda(1 - \zeta_\lambda) < 1.$$

**Beweis:** Sei

$$f(t) = \lambda e^{-\lambda t} - 1.$$

Wegen  $1 - \zeta_\lambda = e^{-\lambda \zeta_\lambda}$  ist

$$f(\zeta_\lambda) = \lambda(1 - \zeta_\lambda) - 1.$$

Es ist also nur zu zeigen, dass  $f(\zeta_\lambda) < 0$ . Dafür stellt man fest, dass

$$\frac{d}{d\lambda} f(t) = (1 - \lambda^2)e^{-\lambda t} < 0 \text{ für alle } t \in \mathbb{R} \text{ und } \lambda > 1.$$

Da weiterhin  $\lim_{\lambda \rightarrow 1} f(\zeta_\lambda) = 0$  wegen  $\lim_{\lambda \rightarrow 1} \zeta_\lambda = 0$ , ist  $\lambda(1 - \zeta_\lambda) < 1$  für alle  $\lambda > 1$ .  $\square$

**Beweis von Satz 3.11:** Die zweitgrößte Komponente  $C'$  ist im Graphen, aus dem  $C_{\max}$  entfernt wurde, offenbar die größte Komponente. Eine Aussage über  $C'$  erlaubt dann das Dualitätsprinzip (Bemerkung 2.8) zusammen mit den bislang in diesem Kapitel gemachten Aussagen. Sei  $\mu_\lambda = \lambda(1 - \zeta_\lambda)$ . Für alle  $\varepsilon, \varepsilon' > 0$  gilt dann

$$\begin{aligned} & P_{n,\lambda} \left( \left| \frac{|C'|}{\ln n} - I_{\mu_\lambda}^{-1} \right| > \varepsilon \right) \\ &= \sum_{l=0}^{\frac{n(n-1)}{2}} P_{n-l,\lambda \frac{n-l}{n}} \left( \left| \frac{|C_{\max}|}{\ln n} - I_{\mu_\lambda}^{-1} \right| > \varepsilon \right) P_{n,\lambda}(|C_{\max}| = l) + o(1) \\ &\leq \sum_{l=(\zeta_\lambda - \varepsilon')n}^{(\zeta_\lambda + \varepsilon')n} P_{n-l,\lambda \frac{n-l}{n}} \left( \left| \frac{|C_{\max}|}{\ln n} - I_{\mu_\lambda}^{-1} \right| > \varepsilon \right) P_{n,\lambda}(|C_{\max}| = l) \\ &\quad + P_{n,\lambda}(|C_{\max}| - \zeta_\lambda n| > \varepsilon n) + o(1). \end{aligned}$$

Der letzte Term verschwindet für wachsendes  $n$  nach Satz 3.6. Da die größte Komponente mit einer wachsenden Anzahl von Knoten oder Kanten höchstens größer wird, ist für  $l \in [(\zeta_\lambda - \varepsilon')n, (\zeta_\lambda + \varepsilon')n]$

$$P_{n-l,\lambda \frac{n-l}{n}} \left( \left| \frac{|C_{\max}|}{\ln n} - I_{\mu_\lambda}^{-1} \right| > \varepsilon \right) \leq P_{n(1-\zeta_\lambda+\varepsilon'),\lambda(1-\zeta_\lambda+\varepsilon')} \left( \left| \frac{|C_{\max}|}{\ln n} - I_{\mu_\lambda}^{-1} \right| > \varepsilon \right)$$

und

$$P_{n-l, \lambda \frac{n-l}{n}} \left( -\frac{|C_{\max}|}{\ln n} + I_{\mu_\lambda}^{-1} < \varepsilon \right) \leq P_{n(1-\zeta_\lambda-\varepsilon'), \lambda(1-\zeta_\lambda-\varepsilon')} \left( -\frac{|C_{\max}|}{\ln n} + I_{\mu_\lambda}^{-1} < \varepsilon \right).$$

Die Funktion  $I_\lambda = \ln \lambda - \lambda + 1$  ist stetig in  $\lambda$ , also kann man  $\varepsilon'$  so wählen, dass

$$|I_{\mu_\lambda} - I_{\lambda(1-\zeta_\lambda \pm \varepsilon')}| < \varepsilon.$$

Wegen Lemma 3.12 zeigt Satz 3.1 dann, dass die beiden rechten Seiten der obenstehenden Ungleichungen mit wachsendem  $n$  gegen 0 streben. Insgesamt ist dann

$$\begin{aligned} & P_{n, \lambda} \left( \left| \frac{|C'|}{\ln n} - I_{\mu_\lambda}^{-1} \right| > \varepsilon \right) \\ & \leq \sum_{l=(\zeta_\lambda-\varepsilon')n}^{(\zeta_\lambda+\varepsilon')n} P_{n, \lambda}(|C_{\max}| = l) \left( P_{n(1-\zeta_\lambda+\varepsilon'), \lambda(1-\zeta_\lambda+\varepsilon')} \left( \frac{|C_{\max}|}{\ln n} - I_{\lambda(1-\zeta_\lambda+\varepsilon')}^{-1} > 2\varepsilon \right) \right. \\ & \quad \left. + P_{n(1-\zeta_\lambda-\varepsilon'), \lambda(1-\zeta_\lambda-\varepsilon')} \left( -\frac{|C_{\max}|}{\ln n} + I_{\lambda(1-\zeta_\lambda-\varepsilon')}^{-1} < 2\varepsilon \right) \right) \\ & \quad + P_{n, \lambda}(|C_{\max}| - \zeta_\lambda n| > \varepsilon n) + o(1) \\ & \longrightarrow 0, \end{aligned}$$

was zu zeigen war. □

### 3.3 Der kritische Fall: $\lambda = 1$

Zum Schluss dieses Kapitels wird das Verhalten der Größe der größten Komponente noch für  $\lambda = 1$  untersucht. Interessanterweise unterscheidet sich dieses Verhalten signifikant vom dem in den anderen Fällen. Tatsächlich ist hier kein nicht triviales schwachen Gesetz der großen Zahlen beweisbar ([Ald97]), so dass an dieser Stelle nur eine für  $\lambda \neq 1$  nicht auftretende, positive Wahrscheinlichkeit für eine Größe der Größenordnung  $n^{\frac{2}{3}}$  gezeigt wird.

**Satz 3.13.** *Es gibt eine Konstante  $c > 0$ , so dass, falls  $n$  groß genug ist, für jedes  $a \geq 1$  gilt:*

$$P_1(a^{-1}n^{\frac{2}{3}} \leq |C_{\max}| \leq an^{\frac{2}{3}}) \geq 1 - \frac{c}{a}.$$

**Beweis:** Der Beweis dieses Satzes stützt sich auf die in Abschnitt 2.3.3 behandelten Abschätzungen. Analog zum Beweis des Lemmas 3.2 verwendet man hier die Aussage, dass

$$|C_{\max}| > k \iff Z_{>k} > k.$$

Zunächst liefert die Markow-Ungleichung zusammen mit Lemma 2.33

$$\begin{aligned} P_1(|C_{\max}| > an^{\frac{2}{3}}) &= P_1(Z_{>an^{\frac{2}{3}}} > an^{\frac{2}{3}}) \\ &\leq a^{-1}n^{-\frac{2}{3}}\mathbb{E}_1Z_{>an^{\frac{2}{3}}} \\ &= a^{-1}n^{-\frac{2}{3}}nP_1(|C_{(1)}| > an^{\frac{2}{3}}) \\ &\leq a^{-1}\frac{c_2}{\sqrt{a}} \end{aligned}$$

mit  $c_2 > 0$ . Die zweite nötige Abschätzung ist

$$\begin{aligned} P_1(|C_{\max}| < a^{-1}n^{\frac{2}{3}}) &= P_1(Z_{>a^{-1}n^{\frac{2}{3}}} = 0) \\ &\leq P_1(|Z_{>a^{-1}n^{\frac{2}{3}}} - \mathbb{E}_1Z_{>a^{-1}n^{\frac{2}{3}}}| \geq \mathbb{E}_1Z_{>a^{-1}n^{\frac{2}{3}}}) \\ &\leq \frac{\text{Var}_1Z_{>a^{-1}n^{\frac{2}{3}}}}{\mathbb{E}_1^2Z_{>a^{-1}n^{\frac{2}{3}}}}. \end{aligned}$$

Ausgehend hiervon sieht man mit Hilfe von Lemma 2.33, dass

$$\begin{aligned} \mathbb{E}_1Z_{>a^{-1}n^{\frac{2}{3}}} &= nP_1(|C_{(1)}| > a^{-1}n^{\frac{2}{3}}) \\ &\geq c_1\sqrt{an^{\frac{2}{3}}} \end{aligned}$$

für ein  $c_1 > 0$ , und Lemma 3.5 zusammen mit Lemma 2.34 liefern

$$\begin{aligned} \text{Var}_1Z_{>a^{-1}n^{\frac{2}{3}}} &\leq nE_1|C_{(1)}|\mathbb{1}_{\{|C_{(1)}|>a^{-1}n^{\frac{2}{3}}\}} \\ &\leq nE_1|C_{(1)}| \\ &\leq Kn^{\frac{4}{3}} \end{aligned}$$

für ein  $K > 0$ , so dass

$$P_1(|C_{\max}| < a^{-1}n^{\frac{2}{3}}) \leq \frac{K}{c_1^2}.$$

Zusammen ergibt sich mit  $c > c_2 + \frac{K}{c_1^2}$  die gewünschte Abschätzung über

$$\begin{aligned} P_1(a^{-1}n^{\frac{2}{3}} \leq |C_{\max}| \leq an^{\frac{2}{3}}) &\geq 1 - P_1(a^{-1}n^{\frac{2}{3}} \geq |C_{\max}|) - P_1(|C_{\max}| \geq an^{\frac{2}{3}}) \\ &\geq 1 - \frac{c_2}{a^{\frac{3}{2}}} - \frac{K}{ac_1^2} \\ &\geq 1 - \frac{c}{a}. \end{aligned}$$

□



## 4 Ein Prinzip moderater Abweichungen

Ziel dieses Kapitels ist die Herleitung eines Prinzips moderater Abweichungen für die größte Komponente eines zufälligen Graphen unter der Wahrscheinlichkeitsverteilung  $P_\lambda$  für  $\lambda > 1$ . Der hier geführte Beweis verwendet Methoden ähnlich denen im Beweis des zentralen Grenzwertsatzes aus [Hof09] (vergleiche Kapitel 5).

**Satz 4.1.** *Sei  $\lambda > 1$  und  $C_{\max}$  die größte Komponente eines zufälligen Graphen mit Verteilung  $P_\lambda$ . Dann genügt  $|C_{\max}|$  einem Prinzip moderater Abweichungen der Form*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (||C_{\max}| - \zeta_\lambda n| > xn^\alpha) = -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}$$

für alle  $\alpha \in (\frac{1}{2}, 1)$  und mit  $\zeta_\lambda = 1 - e^{-\lambda\zeta_\lambda}$ .

Kernidee des Beweises ist die Verwendung der Größe des von den ersten  $k_n$  Knoten aufgespannten Teilgraphen  $C_{\leq k_n}$  an Stelle von  $|C_{\max}|$ . Für eine geeignete Einstellung von  $k_n$  kann sichergestellt werden, dass

$$|C_{\leq k_n}| \geq |C_{\max}| \geq |C_{\leq k_n}| - r_n$$

für einen gegen  $n^\alpha$  verschwindenden Restterm  $r_n$  und im Sinne des Prinzips moderater Abweichungen mit hinreichend großer Wahrscheinlichkeit. Das ist nicht überraschend, da, wie in Abschnitt 3.2 gezeigt, die größte Komponente einen wesentlichen Teil des Graphen einnimmt, also mit hoher Wahrscheinlichkeit

mindestens einen der  $k_n$  ersten Knoten enthält. Andererseits sind die durch die übrigen Knoten aufgespannten Komponenten, wie in Abschnitt 3.2.1 gezeigt, im Vergleich nur von der Ordnung  $\ln n$ , liefern also auf einer  $n^\alpha$ -Skala keinen Beitrag.

## 4.1 Ein Prinzip moderater Abweichungen für Binomialverteilungen

Mit diesem Ansatz und dem bereits in Abschnitt 2.1.4 erarbeiteten Zusammenhang zwischen  $|C_{\leq k}|$  und dem Explorationsprozess  $S_t$ , sowie den Verteilungsaussagen zu  $S_t$  lässt sich ein Prinzip moderater Abweichungen für Binomialverteilungen auf  $|C_{\max}|$  übertragen.

**Satz 4.2.** *Seien  $(X_n)_{n \in \mathbb{N}}$  eine Folge von unter  $P$  binomialverteilten Zufallsgrößen mit Parametern  $(a_n, p_n)$  für die  $\lim_{n \rightarrow \infty} a_n = \infty$  und  $\lim_{n \rightarrow \infty} (1 - p_n)p_n = c \in (0, 1)$  gelten,  $\alpha \in (\frac{1}{2}, 1)$  und*

$$\mu_n := P^{\frac{X_n - a_n p_n}{n^\alpha}}.$$

*Dann genügt die Folge  $(\mu_n)_{n \in \mathbb{N}}$  einem Prinzip großer Abweichungen mit Rate  $a_n^{2\alpha-1}$  und Ratenfunktion  $\frac{x^2}{2c}$ , das heißt, es gilt*

$$\liminf_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln \mu_n(U) \geq - \inf_{x \in U} \frac{x^2}{2c}$$

*für jede offene Menge  $U \subset \mathbb{R}$  und*

$$\limsup_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln \mu_n(A) \leq - \inf_{x \in A} \frac{x^2}{2c}$$

*für jede abgeschlossene Menge  $A \subset \mathbb{R}$ .*



Der Beweis dieses Prinzips moderater Abweichungen erfolgt mit Hilfe des Gärtner-Ellis-Theorems. Hier wird eine Version dieses Theorems gegeben, deren Voraussetzungen an die vorliegende Situation angepasst sind.

**Satz 4.3** (Gärtner-Ellis-Theorem, [OV05, Theorem 3.4]). *Sei  $(\mu_n)_{n \in \mathbb{N}}$  eine Folge von Wahrscheinlichkeitsmaßen und habe  $\mu_n$  die momentenerzeugende Funktion  $\varphi_n$ . Sei weiter  $(a_n)_{n \in \mathbb{N}}$  eine wachsende divergente Folge, so dass für jedes  $t \in \mathbb{R}$  der Grenzwert*

$$\psi(t) := \lim_{n \rightarrow \infty} \frac{1}{a_n} \ln \varphi(a_n t)$$

*existiert und die Funktion  $\psi$  differenzierbar ist. Mit der Bezeichnung*

$$\psi^*(x) := \sup_{t \in \mathbb{R}} \{xt - \psi(t)\}$$

*gilt dann*

$$\liminf_{n \rightarrow \infty} \frac{1}{a_n} \ln \mu_n(U) \geq - \inf_{x \in U} \psi^*(x)$$

*für jede offene Menge  $U \subset \mathbb{R}$  und*

$$\limsup_{n \rightarrow \infty} \frac{1}{a_n} \ln \mu_n(A) \leq - \inf_{x \in A} \psi^*(x)$$

*für jede abgeschlossene Menge  $A \subset \mathbb{R}$ .*

**Beweis von Satz 4.2:** Für die Anwendung des Gärtner-Ellis-Theorems ist die Untersuchung der momentenerzeugenden Funktion  $\varphi_n$  der Zufallsgröße

$$a_n^{-\alpha} (X_n - a_n p_n)$$

notwendig. So ist eine der Voraussetzungen des Theorems die Existenz des Ausdrucks

$$\psi(t) := \lim_{n \rightarrow \infty} a_n^{-2\alpha+1} \ln \varphi_n(t a_n^{2\alpha-1}).$$

Ist dafür  $Y_n$  eine Bernoulli-Variable mit Erfolgswahrscheinlichkeit  $p_n$  mit momentenerzeugender Funktion  $\varphi_{Y_n}$ , so ist

$$\begin{aligned}\varphi_n(a_n^{2\alpha-1}t) &= \mathbb{E} \exp(ta_n^{\alpha-1}(X_n - a_n p_n)) \\ &= \left( \mathbb{E} \exp(ta_n^{\alpha-1}(Y_n - p_n)) \right)^{a_n} \\ &= \left( \varphi_{Y_n}(ta_n^{\alpha-1}) \right)^{a_n}.\end{aligned}$$

Erinnert man sich daran, dass die Ableitungen der momentenerzeugenden Funktion über die Momente der zugrundeliegenden Verteilung angegeben werden können (z. B. [Als00, Kapitel 40]), so führt eine Taylorentwicklung um  $t = 0$  zu

$$\begin{aligned}a_n^{1-2\alpha} \ln \varphi_n(ta_n^{2\alpha-1}) &= a_n^{2-2\alpha} \ln \varphi_{Y_n}(ta_n^{\alpha-1}) \\ &= a_n^{2-2\alpha} \left( \left. \frac{\varphi'_{Y_n}(ta_n^{\alpha-1})}{\varphi_{Y_n}(ta_n^{\alpha-1})} \right|_{t=0} ta_n^{\alpha-1} \right. \\ &\quad \left. + \frac{\varphi_{Y_n}(ta_n^{\alpha-1})\varphi''_{Y_n}(ta_n^{\alpha-1}) - (\varphi_{Y_n}(ta_n^{\alpha-1}))^2}{\varphi_{Y_n}^2(ta_n^{\alpha-1})} \right|_{t=0} \frac{t^2 a_n^{2\alpha-2}}{2} \\ &\quad \left. + \mathcal{O}(t^3 a_n^{3\alpha-3}) \right) \\ &= \frac{\text{Var } Y_n}{2} t^2 + \mathcal{O}(t^3 a_n^{\alpha-1}) \\ &= \frac{p_n(1-p_n)}{2} t^2 + \mathcal{O}(t^3 a_n^{\alpha-1}).\end{aligned}$$

Also ergibt sich  $\psi(t) = \frac{c}{2}t^2$ . Damit ergibt sich die Ratenfunktion zu

$$\varphi^*(x) = \sup_{t \in \mathbb{R}} \left\{ xt - \frac{c}{2}t^2 \right\} = \frac{x^2}{2c}$$

und ist stetig und auf ganz  $\mathbb{R}$  definiert, so dass die Behauptung dieses Satzes aus dem Gärtner-Ellis-Theorem folgt.  $\square$

Im Beweis von Satz 4.1 wird dieser Satz nur für Mengen von spezieller Form benötigt.

**Korollar 4.4.** Insbesondere ist für jedes  $x \in \mathbb{R}^+$

$$\lim_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P(X_n - a_n p_n > a_n^\alpha x) = \lim_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P(X_n - a_n p_n \geq a_n^\alpha x) = -\frac{x^2}{2c}$$

und für jedes  $x \in \mathbb{R}^-$

$$\lim_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P(X_n - a_n p_n < a_n^\alpha x) = \lim_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P(X_n - a_n p_n \leq a_n^\alpha x) = -\frac{x^2}{2c}.$$

**Beweis:** Dieses Korollar folgt direkt aus Satz 4.2 mit der Beobachtung, dass für jede offene Menge  $U$ , deren Abschluss hier mit  $\bar{U}$  bezeichnet wird, die Ungleichungskette

$$\begin{aligned} -\inf_{x \in U} \frac{x^2}{2c} &\leq \liminf_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P\left(\frac{X_n - a_n p_n}{a_n^\alpha} \in U\right) \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P\left(\frac{X_n - a_n p_n}{a_n^\alpha} \in \bar{U}\right) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{a_n^{2\alpha-1}} \ln P\left(\frac{X_n - a_n p_n}{a_n^\alpha} \in \bar{U}\right) \\ &\leq -\inf_{x \in \bar{U}} \frac{x^2}{2c} \\ &= -\inf_{x \in U} \frac{x^2}{2c} \end{aligned}$$

wahr ist, wenn man  $U$  als  $(x, \infty)$  beziehungsweise  $(-\infty, x)$  wählt und die Monotonieeigenschaften der Funktion  $x \rightarrow x^2$  beachtet.  $\square$

## 4.2 Weitere Untersuchungen des Explorationsprozesses $S_t$

Abschnitt 3.2 hat gezeigt, dass sich die Größe der größten Komponente  $C_{\max}$  für  $\lambda > 1$  asymptotisch wie  $\zeta_\lambda n$  verhält, der Explorationsprozess  $S_t$  also beim

Erforschen von  $C_{\max}$  nach etwa  $\zeta_\lambda n$  Schritten stoppt. Um den Prozess  $S_t$  für Zeiten um  $\beta n$  für  $\beta \in (0, 1)$  mit Hilfe des Prinzips moderater Abweichungen für Binomialverteilungen zu untersuchen, werden die asymptotischen ersten und zweiten Momente verwendet, die in diesem Abschnitt hergeleitet werden sollen. Im Folgenden startet der Explorationsprozess immer in den ersten  $k_n$  Knoten, wobei die Werte der Folge  $(k_n)_{n \in \mathbb{N}}$  im Laufe der nächsten Abschnitte passend gewählt werden.

#### 4.2.1 Die Konvergenzgeschwindigkeit der Folge $(1 + \frac{x}{n})^n$

Erinnert man sich an die in Abschnitt 2.1.4 gezeigte Verteilung des Explorationsprozesses

$$S_t + t - k \sim \mathfrak{B}_{n-k, 1-(1-\lambda n^{-1})^t},$$

so liegt es nahe, die Erfolgswahrscheinlichkeit der Binomialverteilung mit der Approximation

$$\left(1 + \frac{x}{n}\right)^n \rightarrow e^x$$

abzuschätzen. Um sicherzustellen, dass der dabei auftretende Fehler auf der Skala des Prinzips moderater Abweichungen verschwindet, benötigt man die Geschwindigkeit dieser Konvergenz.

**Lemma 4.5.** *Sei  $x \in \mathbb{R}$  und  $\lim_{n \rightarrow \infty} b_n = \infty$ . Dann ist*

$$\left(1 + \frac{x}{b_n}\right)^{b_n} - e^x = -\frac{1}{2}e^x x^2 \frac{1}{b_n} + \mathcal{O}(b_n^{-2}).$$

**Beweis:**

$$\begin{aligned}
 \left(1 + \frac{x}{b_n}\right)^{b_n} - e^x &= e^{b_n \ln(1 + \frac{x}{b_n})} - e^x \\
 &= e^{b_n \left(\frac{x}{b_n} - \frac{x^2}{2b_n^2} + \mathcal{O}(b_n^{-3})\right)} - e^x \\
 &= e^x \left( e^{-\frac{x^2}{2b_n} + \mathcal{O}(b_n^{-2})} - 1 \right) \\
 &= e^x \left( -\frac{x^2}{2b_n} + \mathcal{O}(b_n^{-2}) \right),
 \end{aligned}$$

was zu zeigen war. □

**Lemma 4.6.** Sei  $\lim_{n \rightarrow \infty} a_n = 0$  und  $\lim_{n \rightarrow \infty} b_n = \infty$ . Dann ist

$$\left(1 + \frac{a_n}{b_n}\right)^{b_n} - 1 = a_n + \mathcal{O}(a_n^2).$$

**Beweis:**

$$\begin{aligned}
 \left(1 + \frac{a_n}{b_n}\right)^{b_n} - 1 &= e^{b_n \ln(1 + \frac{a_n}{b_n})} - 1 \\
 &= e^{a_n + \mathcal{O}(a_n^2)} - 1 \\
 &= a_n + \mathcal{O}(a_n^2),
 \end{aligned}$$

was zu zeigen war. □

#### 4.2.2 Asymptotische Momente von $S_{\lfloor \beta n \rfloor}$

**Lemma 4.7.** Der asymptotische Erwartungswert von  $S_{\lfloor \beta n \rfloor}$  ist

$$\mathbb{E}S_{\lfloor \beta n \rfloor} = n\mu_\beta + \mathcal{O}(k_n)$$

mit  $\mu_\beta = 1 - \beta - e^{-\lambda\beta}$ .

**Beweis:** Mit Hilfe der Lemmata 4.5 und 4.6 ergibt sich

$$\begin{aligned}
 \mathbb{E}S_{\lfloor \beta n \rfloor} - n\mu_t &= (n - k_n) \left( 1 - \left( 1 - \frac{\lambda}{n} \right)^{\lfloor \beta n \rfloor + k_n - 1} \right) - \lfloor \beta n \rfloor - k_n + 1 - n\mu_t \\
 &= n \left( 1 - \left( 1 - \frac{\lambda\beta}{\beta n} \right)^{\lfloor \beta n \rfloor} \left( 1 - \frac{\lambda \frac{k_n - 1}{n}}{k_n - 1} \right)^{k_n - 1} - t - \mu_t \right) + \mathcal{O}(k_n) \\
 &= n \left( e^{-\lambda} - \left( 1 - \frac{\lambda\beta}{\beta n} \right)^{\lfloor \beta n \rfloor} \left( 1 + \mathcal{O}\left(\frac{k_n}{n}\right) \right) \right) + \mathcal{O}(k_n) \\
 &= \mathcal{O}(k_n),
 \end{aligned}$$

wie zu zeigen war. □

**Lemma 4.8.** Die asymptotische Varianz von  $S_{\lfloor \beta n \rfloor}$  ist

$$\text{Var } S_{\lfloor \beta n \rfloor} = n\nu_\beta + \mathcal{O}(k_n)$$

mit  $\nu_\beta = e^{-\lambda\beta} (1 - e^{-\lambda\beta})$ .

**Beweis:** Analog zum Beweis des Lemmas 4.7 ergibt sich

$$\begin{aligned}
 \text{Var } S_{\lfloor \beta n \rfloor} &= (n - k_n) \left( 1 - \left( 1 - \frac{\lambda}{n} \right)^{\lfloor \beta n \rfloor} \right) \left( 1 - \frac{\lambda}{n} \right)^{\lfloor \beta n \rfloor} \\
 &= (n - k_n) (1 - e^{-\lambda\beta} + \mathcal{O}(n^{-1})) (e^{-\lambda\beta} + \mathcal{O}(n^{-1})) \\
 &= n\nu_\beta + \mathcal{O}(k_n).
 \end{aligned}$$

□

Spezielle Werte für  $\beta n$ , die in den folgenden Ausführungen Verwendung finden, sind

$$m_{x^+} := \lfloor \zeta_\lambda n + xn^\alpha \rfloor \text{ und } m_{x^-} := \lfloor \zeta_\lambda n - xn^\alpha \rfloor.$$

**Korollar 4.9.** *Es gilt*

$$\mathbb{E}S_{m_{x\pm}} = \pm (\lambda(1 - \zeta_\lambda) - 1) n^\alpha x + \mathcal{O}(k_n + n^{2\alpha-1}).$$

**Beweis:** Schreibt man  $\mathbb{E}S_{m_{x\pm}} = \mathbb{E}S_{\lfloor n\zeta_\lambda \pm xn^\alpha \rfloor}$  mit Hilfe des Resultats aus Lemma 4.7 aus und entwickelt  $\mu_{\zeta_\lambda \pm xn^{\alpha-1}}$  dann um  $x = 0$ , so folgt

$$\begin{aligned} \mathbb{E}S_{m_{x\pm}} &= n\mu_{\zeta_\lambda \pm xn^{\alpha-1}} + \mathcal{O}(k_n) \\ &= n(\mu_{\zeta_\lambda} \pm \mu'_{\zeta_\lambda} xn^{\alpha-1} + \mathcal{O}(n^{2\alpha-2})) + \mathcal{O}(k_n) \\ &= n\mu_{\zeta_\lambda} \pm \mu'_{\zeta_\lambda} xn^\alpha + \mathcal{O}(n^{2\alpha-1} + k_n). \end{aligned}$$

Nun ist  $\zeta_\lambda$  definiert durch die Gleichung  $\zeta_\lambda = 1 - e^{-\lambda\zeta_\lambda}$ , so dass  $\mu_{\zeta_\lambda} = 0$  und

$$\begin{aligned} \mu'_{\zeta_\lambda} &= -1 + \lambda e^{-\lambda\zeta_\lambda} \\ &= \lambda(1 - \zeta_\lambda) - 1, \end{aligned}$$

womit die Behauptung folgt. □

### 4.3 Das Prinzip moderater Abweichungen für $|C_{\leq k_n}|$

**Lemma 4.10.** *Für  $\lambda > 1$  und  $k_n = K \ln n$  für ein geeignetes  $K > 0$  ist*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (||C_{\leq k_n}| - n\zeta_\lambda| > xn^\alpha) = -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}$$

für alle  $\alpha \in (\frac{1}{2}, 1)$ .

Dieser Abschnitt verwendet den in Bemerkung 2.9 festgehaltenen Zusammenhang

$$|C_{\leq k_n}| = \min_{t \in \mathbb{N}} \{S_t = 0\},$$

um mit Hilfe der äquivalenten Darstellung

$$|C_{\leq k_n}| > l \iff S_m > 0 \text{ für alle } m \leq l$$

und den Aussagen zur Verteilung von  $S_t$  in Abschnitt 2.1.4 ein Prinzip moderater Abweichungen von Satz 4.2 auf  $|C_{\leq k_n}|$  zu übertragen. Dafür betrachtet man die Abweichungen nach oben und nach unten separat, was eine einfachere Verwendung dieser Darstellung erlaubt. Der dabei entstehende Fehler

$$\begin{aligned} & P_\lambda (||C_{\leq k_n}| - n\zeta_\lambda| > xn^\alpha) \\ & - (P_\lambda (|C_{\leq k_n}| > n\zeta_\lambda + xn^\alpha) + P_\lambda (|C_{\leq k_n}| < n\zeta_\lambda - xn^\alpha)) \end{aligned}$$

verschwindet auf der Skala der moderaten Abweichungen.

## Die oberen Schranken

**Lemma 4.11.** *In der Situation von Lemma 4.10 gilt für alle  $\alpha \in (\frac{1}{2}, 1)$*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (|C_{\leq k_n}| > n\zeta_\lambda + xn^\alpha) \leq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

**Beweis:** Mit der oben vorgestellten Darstellung von  $|C_{\leq k_n}|$  gilt für alle  $l \in \mathbb{N}$

$$P_\lambda(|C_{\leq k_n}| > l) = P_\lambda(\forall m \leq l : S_m > 0).$$

Erinnert man sich an das in Abschnitt 4.2.2 definierte  $m_{x^+}$ , so erhält man

$$\begin{aligned} P_\lambda(|C_{\leq k_n}| > m_{x^+}) &= P_\lambda(\forall m \leq m_{x^+} : S_m > 0) \\ &\leq P_\lambda(S_{m_{x^+}} > 0) \\ &= P_\lambda\left(\frac{S_{m_{x^+}} - \mathbb{E}S_{m_{x^+}}}{n^\alpha} > -\frac{\mathbb{E}S_{m_{x^+}}}{n^\alpha}\right) \\ &= P_\lambda\left(\frac{S_{m_{x^+}} - \mathbb{E}S_{m_{x^+}}}{n^\alpha} > x(1 - \lambda(1 - \zeta_\lambda)) + o(1)\right). \end{aligned}$$



Für die letzte Umformung fand dabei Korollar 4.9 Verwendung. Die Verteilung von  $S_{m_{x^+}}$  ist bekannt, und Satz 4.2 zusammen mit Lemma 4.8 liefern

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} P_\lambda (S_{m_{x^+}} > 0) = -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

□

**Lemma 4.12.** *In der Situation von Lemma 4.10 gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (|C_{\leq k_n}| < n\zeta_\lambda - xn^\alpha) \leq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

Für diese Abschätzung muss zunächst festgehalten werden, dass die Wahrscheinlichkeit für einen Nulldurchgang des Explorationsprozesses deutlich vor  $n\zeta_\lambda$  auf der Skala der moderaten Abweichungen verschwindet, um die Untersuchung auf eine kleine Umgebung von  $n\zeta_\lambda$  beschränken zu können.

**Korollar 4.13.** *Für jedes  $\varepsilon > 0$  ist*

$$\lim_{n \rightarrow \infty} \ln P_\lambda (\exists m < (\zeta_\lambda - \varepsilon)n : S_m = 0) = 0.$$

**Beweis:** Diese Aussage folgt aus Lemma 3.8, mit  $\alpha = \zeta - \varepsilon$ , wenn  $k_n$  mindestens wie  $\ln n$  wächst. □

**Beweis von Lemma 4.12:** Analog zum Vorgehen im vorhergehenden Beweis findet man

$$\begin{aligned} P_\lambda (|C_{\leq k_n}| < m_{x^-}) &= P_\lambda (\exists m < m_{x^-} : S_m = 0) \\ &\leq \sum_{m=(\zeta_\lambda - \varepsilon)n}^{m_{x^-}} P_\lambda (S_m = 0) + P_\lambda (\exists m < (\zeta_\lambda - \varepsilon)n : S_m = 0) \end{aligned}$$

für jedes  $\varepsilon > 0$ . Der zweite Term fällt wegen Korollar 4.13 schnell genug, um keinen Beitrag zum Resultat zu liefern. Bezeichnet man

$$m_\delta := \lfloor n(\zeta_\lambda - \delta) - xn^\alpha \rfloor,$$

so lässt sich für jeden Summanden  $S_m$  des ersten Terms ein  $\delta \in [0, \varepsilon]$  finden mit  $m = m_\delta$ . Korollar 4.9 untersucht  $\mathbb{E}_\lambda S_{m_\delta}$  für  $\delta = 0$ . Der Beweis des Korollars lässt aber auch eine Untersuchung für  $\delta > 0$  zu. Dazu bemerkt man, dass wegen Lemma 3.12

$$\begin{aligned} \frac{d}{dt} \mu_t \Big|_{t=\zeta_\lambda} &= \lambda e^{-\lambda \zeta_\lambda} - 1 \\ &= \lambda(1 - \zeta_\lambda) - 1 < 0. \end{aligned}$$

Zusammen mit  $\mu_{\zeta_\lambda} = 0$  kann man  $\varepsilon$  so wählen, dass für alle  $0 < \delta \leq \varepsilon$  gilt

$$\mu_{\zeta_\lambda - \delta} > 0,$$

woraus man für diese  $\delta$  ablesen kann, dass

$$\frac{1}{n^\alpha} \mathbb{E}_\lambda S_{m_\delta} \xrightarrow{n \rightarrow \infty} \infty.$$

Mit Satz 4.2 ergibt sich dann für  $\delta \geq 0$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda(S_{m_\delta} = 0) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda\left(\frac{S_{m_\delta} - \mathbb{E}_\lambda S_{m_\delta}}{n^\alpha} = -\frac{\mathbb{E}_\lambda S_{m_\delta}}{n^\alpha}\right) \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda\left(\frac{S_{m_\delta} - \mathbb{E}_\lambda S_{m_\delta}}{n^\alpha} \leq -x(1 - \lambda(1 - \zeta_\lambda)) + o(1)\right) \\ &\leq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}. \end{aligned}$$

Deswegen kann man folgern

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda(|C_{\leq k_n}| < m_{x^-}) \leq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)},$$

was den Beweis dieser Schranke abschließt. □

## Die unteren Schranken

**Lemma 4.14.** *Es gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (|C_{\leq k_n}| > n\zeta_\lambda + xn^\alpha) \geq -\frac{x^2(1-\lambda(1-\zeta_\lambda))^2}{2\zeta_\lambda(1-\zeta_\lambda)}.$$

Analog zur oberen Schranke in diesem Fall, wird eine Untersuchung der Verteilung des Explorationsprozesses durch einen Vergleich mit einer Binomialverteilung mit passenden Parametern vereinfacht. Dafür ist folgende Bemerkung hilfreich.

**Bemerkung 4.15.** Sei  $0 \leq a \leq 1$  und  $b \geq 1$ , dann ist

$$1 - (1-a)^b \leq ab.$$

**Beweis:** Diese Ungleichung gilt für  $a = 0$  und  $b = 1$  und außerdem ist

$$\frac{d}{da} ab - (1 - (1-a)^b) = b - b(1-a)^b \geq 0$$

und

$$\begin{aligned} \frac{d}{db} ab - (1 - (1-a)^b) &= a + (1-a)^b \ln(1-a) \\ &\geq a + (1-a) \ln(1-a). \end{aligned}$$

Nach de L'Hospital gilt

$$\lim_{a \rightarrow 0} a + (1-a) \ln(1-a) = 0$$

und schließlich ist

$$\frac{d}{da} a + (1-a) \ln(1-a) = -\ln(1-a) \geq 0.$$

Damit folgt die Ungleichung für alle  $0 \leq a \leq 1$  und  $b \geq 1$ . □

**Beweis von Lemma 4.14:** Wie für die obere Schranke betrachtet man hier

$$\begin{aligned} P_\lambda(|C_{\leq k_n}| > m_x) &= P_\lambda(\forall m \leq m_x : S_m > 0) \\ &\geq P_\lambda(\forall m \leq m_x : S_m > 0, S_{m_x} > \varepsilon n^\alpha) \\ &= P_\lambda(S_{m_x} > \varepsilon n^\alpha) - P_\lambda(S_{m_x} > \varepsilon n^\alpha, \exists m \leq m_x : S_m = 0) \end{aligned}$$

für ein später gewähltes kleines  $\varepsilon > 0$  und erhält für den vorderen Term mit Hilfe von Lemma 2.12 und Korollar 4.4

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} P_\lambda(S_{m_x} > \varepsilon n^\alpha) = -\frac{x^2(1 - \lambda(1 - \zeta_\lambda) + \varepsilon)^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

Zu zeigen bleibt also nur, dass

$$\begin{aligned} &e^{n^{2\alpha-1}} P_\lambda(S_{m_x} > \varepsilon n^\alpha, \exists m \leq m_x : S_m = 0) \\ &\leq e^{n^{2\alpha-1}} \sum_{m=0}^{m_x} P_\lambda(S_m = 0, S_{m_x} > \varepsilon n^\alpha) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Dies geschieht mit Hilfe einer Unterscheidung in zwei Fälle. Sei dazu  $\beta = \zeta_\lambda - \varepsilon$ . Für den Fall  $m \leq \beta n$  ist die nötige Aussage bereits in Korollar 4.13 gezeigt. Sei also  $m \geq \beta n$ , dann liefert Lemma 2.12

$$\begin{aligned} P_\lambda(S_m = 0, S_{m_x} > \varepsilon n^\alpha) &\leq P_\lambda(S_{m_x} > \varepsilon n^\alpha \mid S_m = 0) \\ &= P_\lambda(X > (m_x - m) + \varepsilon n^\alpha), \end{aligned}$$

mit  $X \sim \mathcal{B}(n - (m + k_n - 1), (1 - (1 - \frac{\lambda}{n})^{m_x - m}))$ . Eine Abschätzung des Erwar-

tungswertes von  $X$  mit Hilfe der Bemerkung 4.15 ergibt

$$\begin{aligned}
 \mathbb{E}X &= (n - (m + k_n - 1)) \left( 1 - \left( 1 - \frac{\lambda}{n} \right)^{m_x - m} \right) \\
 &\leq (n - (m + k_n - 1)) (m_x - m) \frac{\lambda}{n} \\
 &= (m_x - m) \lambda \left( 1 - \frac{m + k_n - 1}{n} \right) \\
 &\leq (m_x - m) \lambda (1 - (\zeta_\lambda - \varepsilon)).
 \end{aligned}$$

In Lemma 3.12 wurde gezeigt, dass  $\lambda(1 - \zeta_\lambda) < 1$ , also gibt es ein  $\varepsilon$ , so dass

$$\lambda(1 - (\zeta_\lambda - \varepsilon)) < 1 - \varepsilon.$$

Sei  $\varepsilon$  von nun an mit dieser Eigenschaft gewählt. Es folgt

$$\mathbb{E}X \leq (1 - \varepsilon)(m_x - m)$$

und damit

$$\begin{aligned}
 P_\lambda(S_m = 0, S_{m_x} > \varepsilon n^\alpha) &\leq P_\lambda(X > (m_x - m) + \varepsilon n^\alpha) \\
 &\leq P_\lambda(X > \mathbb{E}X + (m_x - m) + \varepsilon((m_x - m) + n^\alpha)).
 \end{aligned}$$

Die Abschätzung von Chernoff wiederum (z. B. [Hof09, Theorem 2.18]) liefert

dann

$$\begin{aligned}
 P_\lambda(S_m = 0, S_{m_x} > \varepsilon n^\alpha) &\leq P_\lambda(X > \mathbb{E}X + (m_x - m) + \varepsilon((m_x - m) + n^\alpha)) \\
 &\leq \exp\left(-\frac{\varepsilon^2(m_x - m + n^\alpha)^2}{2\left((1 - \varepsilon)(m_x - m) + \frac{\varepsilon(m_x - m + n^\alpha)}{3}\right)}\right) \\
 &\leq \exp\left(-\frac{\varepsilon^2(m_x - m + n^\alpha)^2}{2\left((m_x - m) + \frac{\varepsilon n^\alpha}{3}\right)}\right) \\
 &\leq \exp\left(-\frac{\varepsilon^2}{2}(m_x - m + n^\alpha)\right) \\
 &\leq \exp\left(-\frac{\varepsilon^2}{2}n^\alpha\right),
 \end{aligned}$$

was wegen  $n^\alpha n^{-(2\alpha-1)} = n^{1-\alpha} \xrightarrow{n \rightarrow \infty} \infty$  hinreichend schnell fällt.  $\square$

**Lemma 4.16.** *In der Situation von Lemma 4.10 gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda(|C_{\leq k_n}| < n\zeta_\lambda - xn^\alpha) \geq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

**Beweis:** Man betrachtet

$$\begin{aligned}
 P_\lambda(|C_{\leq k_n}| < n\zeta_\lambda - xn^\alpha) &= P_\lambda(\exists m \leq m_{x^-} : S_m = 0) \\
 &\geq P_\lambda(S_{m_{x^-}} \leq 0) \\
 &= P_\lambda\left(\frac{S_{m_{x^-}} - \mathbb{E}_\lambda S_{m_{x^-}}}{n^\alpha} \leq -\frac{\mathbb{E}_\lambda S_{m_{x^-}}}{n^\alpha}\right).
 \end{aligned}$$

Korollar 4.9 zusammen mit Korollar 4.4 liefern hier

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda\left(\frac{S_{m_{x^-}} - \mathbb{E}_\lambda S_{m_{x^-}}}{n^\alpha} \leq -\frac{\mathbb{E}_\lambda S_{m_{x^-}}}{n^\alpha}\right) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda\left(\frac{S_{m_{x^-}} - \mathbb{E}_\lambda S_{m_{x^-}}}{n^\alpha} \leq -x(1 - \lambda(1 - \zeta_\lambda) + o(1))\right) \\
 &= -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)},
 \end{aligned}$$

was zu zeigen war. □

**Beweis von Lemma 4.10:** Der Beweis dieses Lemmas ist mit den vier Ungleichungen aus den Lemmata 4.11 bis 4.16 erbracht. □

## 4.4 Das Prinzip moderater Abweichungen für $|C_{\max}|$

Schließlich gilt es zu beweisen, dass aus dem vorher gezeigten Prinzip moderater Abweichungen für die Größe der Komponente  $C_{\leq k_n}$  ein Prinzip moderater Abweichungen für die größte Komponente  $C_{\max}$  folgt. Um zu schließen, dass  $C_{\leq k_n}$  mit  $C_{\max}$  im Wesentlichen übereinstimmt, sind genaue Aussagen über die Konvergenzgeschwindigkeit der Konvergenz

$$\frac{|C_{\max}|}{n} \rightarrow \zeta_\lambda$$

nötig.

**Bemerkung 4.17.** O’Connel zeigt in [O’C98] für diese Situation ein Prinzip großer Abweichungen, welches es gestattet für jedes  $\varepsilon > 0$  ein  $K_\varepsilon > 0$  anzugeben mit

$$P_\lambda(|C_{\max}| - n\zeta_\lambda| > \varepsilon n) \leq e^{-nK_\varepsilon}.$$

Zunächst ist  $C_{\max}$  nur dann größer als  $C_{\leq k_n}$ , wenn keiner der ersten  $k_n$  Knoten des Graphen in der größten Komponente liegt, was mit wachsendem  $k_n$  zunehmend unwahrscheinlicher wird.

**Lemma 4.18.** *Es gibt ein  $C > 0$ , so dass*

$$P_\lambda(|C_{\max}| > |C_{\leq k_n}|) \leq e^{-Ck_n}.$$

**Beweis:** Einfache Umformungen ergeben

$$\begin{aligned}
 & P_\lambda(|C_{\max}| > |C_{\leq k_n}|) \\
 &= P_\lambda(\forall 1 \leq i \leq k_n : i \notin C_{\max}) \\
 &= P_\lambda(1 \notin C_{\max})P_\lambda(2 \notin C_{\max} \mid 1 \notin C_{\max}) \cdots P_\lambda(k_n \notin C_{\max} \mid 1, \dots, k_n - 1 \notin C_{\max}).
 \end{aligned}$$

Da die Größe der größten Komponente invariant gegenüber Umbenennung der Knoten ist, ist für jedes  $\varepsilon > 0$

$$\begin{aligned}
 P_\lambda(1 \notin C_{\max}) &\leq 1 - \frac{\zeta_\lambda - \varepsilon}{n} + e^{-nK_\varepsilon}, \\
 P_\lambda(j \notin C_{\max} \mid 1, \dots, j-1 \notin C_{\max}) &\leq P_\lambda(1 \notin C_{\max}),
 \end{aligned}$$

und damit gibt es ein  $C > 0$ , so dass

$$\begin{aligned}
 P_\lambda(|C_{\max}| > |C_{\leq k_n}|) &\leq \left(1 - \frac{\zeta_\lambda - \varepsilon}{n} + e^{-nK_\varepsilon}\right)^{k_n} \\
 &\leq e^{-Ck_n},
 \end{aligned}$$

wobei die Aussage aus Bemerkung 4.17 Verwendung findet. □

Für die Abschätzung in die andere Richtung nutzt man aus, dass, wie in Satz 3.11 angedeutet wurde, die Komponenten, die  $C_{\leq k_n}$  zusätzlich zur größten Komponente enthält, deutlich zu klein sind, um auf der Skala  $n^\alpha$  einen wesentlichen Beitrag zur Gesamtgröße zu liefern.

**Lemma 4.19.** *Es gibt einen Rest  $r_n \in \mathcal{O}(\sqrt{n})$  und ein  $C > 0$ , so dass*

$$P_\lambda(|C_{\max}| + r_n < |C_{\leq k_n}|) \leq e^{-Ck_n}.$$

**Beweis:** Man teilt das Ereignis  $\{|C_{\max}| + r_n < |C_{\leq k_n}|\}$  nach der Anzahl  $j$  der Knoten in den ersten  $k_n$ , die nicht in der größten Komponente liegen. Benennt



man diese Knoten mit  $1, \dots, j$ , so ergibt sich

$$\begin{aligned}
 & P_\lambda(|C_{\max}| + r_n < |C_{\leq k_n}|) \\
 & \leq \sum_{j=0}^{k_n} \binom{k_n}{j} P_\lambda(|C_{\max}| + r_n < |C_{\max}| + |C_{\leq j}|, 1, \dots, j \notin C_{\max}, j+1, \dots, k_n \in C_{\max}) \\
 & \leq e^{k_n} \sum_{j=0}^{k_n} P_\lambda(|C_{\max}| + r_n < |C_{\max}| + |C_{\leq j}|, 1, \dots, j \notin C_{\max}) \\
 & = e^{k_n} \sum_{j=0}^{k_n} P_\lambda(|C_{\leq j}| > r_n, 1, \dots, j \notin C_{\max})
 \end{aligned}$$

Wie im Beweis von Satz 3.11 wird jetzt Gebrauch vom Prinzip großer Abweichungen für die größte Komponente (Bemerkung 4.17) gemacht, um mit Hilfe des Dualitätsprinzips eine Abschätzung dieser Summanden zu bekommen. Das Prinzip großer Abweichungen gewährleistet eine relativ präzise Kontrolle der Abweichung von  $|C_{\max}|$  zu  $\zeta_\lambda n$ . So gibt es für jedes  $\varepsilon > 0$  ein  $K > 0$ , so dass

$$P_\lambda(|C_{\max}| \leq (\zeta_\lambda - \varepsilon)n) \leq e^{-Kn}.$$

Daher ist

$$\begin{aligned}
 & P_\lambda(|C_{\leq j}| > r_n, 1, \dots, j \notin C_{\max}) \\
 & \leq P_\lambda(|C_{\leq j}| > r_n, 1, \dots, j \notin C_{\max}, |C_{\max}| > (\zeta_\lambda - \varepsilon)n) + P_\lambda(|C_{\max}| \leq (\zeta_\lambda - \varepsilon)n) \\
 & \leq P_\lambda(|C_{\leq j}| > r_n, 1, \dots, j \notin C_{\max}, |C_{\max}| > (\zeta_\lambda - \varepsilon)n) + e^{-Kn}
 \end{aligned}$$

für ein  $K > 0$ . Betrachtet man

$$\begin{aligned}
 & P_\lambda(C_{\leq j} > r_n, 1, \dots, j \notin C_{\max}, C_{\max} > (\zeta_\lambda - \varepsilon)n) \\
 & = P_\lambda(C_{\leq j} > r_n \mid 1, \dots, j \notin C_{\max}, C_{\max} > (\zeta_\lambda - \varepsilon)n) \\
 & \quad \times P_\lambda(1, \dots, j \notin C_{\max}, C_{\max} > (\zeta_\lambda - \varepsilon)n) \\
 & \leq P_\lambda(C_{\leq j} > r_n \mid 1, \dots, j \notin C_{\max}, C_{\max} > (\zeta_\lambda - \varepsilon)n),
 \end{aligned}$$

so ergibt sich wie angekündigt eine Abschätzung mit Hilfe des Dualitätsprinzips:

$$\begin{aligned} & P_\lambda(|C_{\leq j}| > r_n \mid 1, \dots, j \notin C_{\max}, |C_{\max}| > (\zeta_\lambda - \varepsilon)n) \\ & \leq P_{\lambda(1-\zeta_\lambda+\varepsilon)}(|C_{\leq j}| > r_n) \\ & \leq P_{\lambda(1-\zeta_\lambda+\varepsilon)}(|C_{\leq k_n}| > r_n). \end{aligned}$$

Da wegen Lemma 3.12 der Ausdruck  $\lambda(1 - \zeta_\lambda)$  kleiner als 1 ist, lässt sich auch ein  $\varepsilon$  finden mit  $\lambda(1 - \zeta_\lambda + \varepsilon) < 1$ . Das Lemma 2.30 liefert dann die erforderliche Abschätzung:

$$P_{\lambda(1-\zeta_\lambda+\varepsilon)}(|C_{\leq k_n-1}| > r_n) \leq e^{-r_n I_{\lambda(1-\zeta_\lambda+\varepsilon)} - (k_n-1) \ln \lambda}$$

und deswegen

$$P_\lambda(|C_{\max}| + r_n < |C_{\leq k_n}|) \leq (k_n + 1)e^{k_n} (e^{-r_n I_{\lambda(1-\zeta_\lambda+\varepsilon)} - (k_n-1) \log \lambda} + e^{-Kn}).$$

Die Restfolge  $r_n$  kann wegen  $k_n \in \mathcal{O}(\ln n)$  jetzt wie erforderlich so gewählt werden, dass  $k_n \in \mathcal{o}(r_n)$  und  $r_n \in \mathcal{o}(n^\alpha)$ .  $\square$

Mit diesen Hilfsmitteln lässt sich nun der Beweis des Prinzips moderater Abweichungen für  $|C_{\max}|$  zusammensetzen:

**Beweis von Theorem 4.1:** Man betrachtet für die Folge  $r_n$  aus Lemma 4.19 und ein geeignetes  $C > 0$

$$\begin{aligned} & P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha) \\ & = P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha, ||C_{\max}| - |C_{\leq k_n}|| \leq r_n) \\ & \quad + P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha, ||C_{\max}| - |C_{\leq k_n}|| > r_n) \\ & \leq P_\lambda (||C_{\leq k_n}| - n\zeta_\lambda| > xn^\alpha + \mathcal{o}(n^\alpha)) + e^{-Ck_n}. \end{aligned}$$

Damit lässt sich mit dem Prinzip moderater Abweichungen von  $C_{\leq k_n}$  eine obere Schranke herleiten, wenn  $k_n = K \ln n$  mit  $CK > 2\alpha - 1$  gewählt ist:

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha) \leq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

Analog gilt natürlich

$$\begin{aligned} & P_\lambda (||C_{\leq k_n}| - n\zeta_\lambda| > xn^\alpha) \\ &= P_\lambda (||C_{\leq k_n}| - n\zeta_\lambda| > xn^\alpha, ||C_{\max}| - |C_{\leq k_n}|| \leq r_n) \\ &\quad + P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha, ||C_{\max}| - |C_{\leq k_n}|| > r_n) \\ &\leq P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha + o(n^\alpha)) + e^{-Ck_n} \end{aligned}$$

und damit

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_\lambda (||C_{\max}| - n\zeta_\lambda| > xn^\alpha) \geq -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)}.$$

Damit ist der Beweis des Prinzips moderater Abweichungen für die größte Komponente eines Erdős-Rényi-Zufallsgraphen im superkritischen Fall vollständig.

□

## 4.5 Übertragung auf $P_{n,M}$

**Satz 4.20.** *Sei  $(M_n)_{n \in \mathbb{N}}$  eine Folge natürlicher Zahlen mit*

$$\frac{2M_n}{n} \rightarrow \lambda$$

*für  $\lambda > 1$ . Dann gilt für alle  $\alpha \in (\frac{1}{2}, 1)$*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_{n,M_n} (||C_{\max}| - \zeta_\lambda n| > xn^\alpha) = -\frac{x^2(1 - \lambda(1 - \zeta_\lambda))^2}{2\zeta_\lambda(1 - \zeta_\lambda)},$$

*wenn  $\zeta_\lambda = 1 - e^{-\lambda\zeta_\lambda}$ .*

**Beweis:** Zum Beweis dieses Satzes verwendet man die in Abschnitt 2.1.2 vorgestellten Beziehungen zwischen dem bereits in Satz 4.1 untersuchten Modell  $G(n, p)$  und  $G(n, M)$ . So lässt sich die Wahrscheinlichkeit für ein Ereignis  $A$  unter  $P_{n,\lambda}$  darstellen durch

$$P_{n,\lambda}(A) = \sum_{M=0}^{\frac{n(n-1)}{2}} P_{n,M}(A) \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{M\}).$$

Für  $a_n \in \mathcal{O}(\exp(n))$  und  $b_{i,n} > 0$  gilt wegen

$$\begin{aligned} \frac{1}{n} \ln \sum_{i=1}^{a_n} b_{i,n} &\geq \frac{1}{n} \ln \max_{1 \leq i \leq a_n} b_{i,n} \\ \text{und } \frac{1}{n} \ln \sum_{i=1}^{a_n} b_{i,n} &\leq \frac{1}{n} \ln a_n \max_{1 \leq i \leq a_n} b_{i,n} \\ &\leq \frac{\ln a_n}{n} + \frac{1}{n} \ln \max_{1 \leq i \leq a_n} b_{i,n} \end{aligned}$$

so lange  $\frac{1}{n} \ln \max_{1 \leq i \leq a_n} b_{i,n}$  in  $\mathbb{R}^+$  konvergiert,

$$\frac{1}{n} \ln \sum_{i=1}^{a_n} b_{i,n} \sim \frac{1}{n} \ln \max_{1 \leq i \leq a_n} b_{i,n}$$

(Prinzip des größten Terms). Deswegen genügt es, sich bei der Untersuchung der rechten Summe auf den maximalen Summanden zu beschränken. Weiterhin erwartet man, dass das maximale Element bei  $M \sim \frac{\lambda}{2}n$  liegt. Eine Umindizierung der Summe ergibt dann

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_{n,\lambda}(A) &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln \sum_{M=0}^{\frac{n(n-1)}{2}} P_{n,M}(A) \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{M\}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln \sum_{c \in \mathbb{R}^+, cn \in \{0, \dots, \frac{n(n-1)}{2}\}} P_{n,cn}(A) \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{cn\}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln \max_{c \in \mathbb{R}^+} P_{n,cn}(A) \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{cn\}). \end{aligned}$$

Um festzustellen, für welches  $c$  dieser Ausdruck maximal wird, bietet sich zunächst

für beliebiges  $c \in \mathbb{R}^+$  eine Untersuchung des Terms

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{cn\}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln \left( \frac{\binom{n(n-1)}{2}}{cn} \right) \left( \frac{\lambda}{n} \right)^{cn} \left( 1 - \frac{\lambda}{n} \right)^{\frac{n(n-1)}{2} - cn} \end{aligned}$$

an. Verwendet man die Stirling-Approximation  $\ln n! = n \ln n - n + \mathcal{O}(\ln n)$  sowie die Taylor-Entwicklung des natürlichen Logarithmus  $\ln(1+x) = x + \mathcal{O}(x^2)$  und die daraus folgende Beobachtung  $n^2 \ln(n-c) = n^2 \ln n - cn + \mathcal{O}(1)$ , so folgt

$$\begin{aligned} \ln \left( \frac{\binom{n(n-1)}{2}}{cn} \right) &= \ln \left( \frac{n(n-1)}{2} \right)! - \ln(cn)! - \ln \left( \frac{n(n-1)}{2} - cn \right)! \\ &= \frac{n^2}{2} \ln \frac{n^2}{2} - \frac{n^2}{2} - cn \ln cn + cn \\ &\quad - \left( \frac{n^2}{2} - cn \right) \ln \left( \frac{n^2}{2} - cn \right) + \frac{n^2}{2} - cn + \mathcal{O}(n^{2\alpha-1}) \\ &= n^2 \ln n - n^2 \ln 2 - cn \ln cn - \frac{n^2}{2} \ln \left( \frac{n}{2}(n-c) \right) \\ &\quad + cn \ln \left( \frac{n}{2}(n-c) \right) + \mathcal{O}(n^{2\alpha-1}) \\ &= n^2 \ln n - n^2 \ln 2 - cn \ln cn - \frac{n^2}{2} \ln n + \frac{n^2}{2} \ln 2 \\ &\quad - \frac{n^2}{2} \ln(n-c) + cn \ln n + cn \ln 2 - cn \ln(n-c) + \mathcal{O}(n^{2\alpha-1}) \\ &= -cn(\ln c + \ln 2 - \ln n - 1) + \mathcal{O}(n^{2\alpha-1}), \end{aligned}$$

$$\ln \left( \frac{\lambda}{n} \right)^{cn} = cn \ln \lambda - cn \ln n$$

und

$$\begin{aligned} \ln \left( 1 - \frac{\lambda}{n} \right)^{\frac{n(n-1)}{2} - cn} &= \left( \frac{n(n-1)}{2} - cn \right) \ln \left( 1 - \frac{\lambda}{n} \right) \\ &= \left( \frac{n(n-1)}{2} - cn \right) \left( -\frac{\lambda}{n} \right) + \mathcal{O}(n^{2\alpha-1}) \\ &= -\frac{n}{2} \lambda + \mathcal{O}(n^{2\alpha-1}). \end{aligned}$$

Insgesamt ergibt sich

$$\begin{aligned} & \ln \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{cn\}) \\ &= -cn(\ln c + \ln 2 - \ln n - 1) + cn \ln \lambda - cn \ln n - \lambda \frac{n}{2} + \mathcal{O}(n^{2\alpha-1}), \end{aligned}$$

woraus sich für den Grenzübergang folgern lässt

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln \mathfrak{B}_{\frac{n(n-1)}{2}, \frac{\lambda}{n}}(\{cn\}) = \begin{cases} 0 & \text{für } c = \frac{\lambda}{2} \\ -\infty & \text{sonst.} \end{cases}$$

Damit ist

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_{n,\lambda}(|C_{\max}| - \zeta_{\lambda} n > xn^{\alpha}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^{2\alpha-1}} \ln P_{n, \lfloor \frac{\lambda}{2} n \rfloor}(|C_{\max}| - \zeta_{\lambda} n > xn^{\alpha}), \end{aligned}$$

was die Behauptung zeigt. □

## 5 Ein Zentraler Grenzwertsatz

Dieses Kapitel beschäftigt sich mit dem Beweis des Zentralen Grenzwertsatzes für die Größe der größten Komponente eines zufälligen Graphen aus [Hof09, Kapitel 4.5], der den Beweis des Prinzips moderater Abweichungen in Kapitel 4 inspiriert hat.

**Satz 5.1.** *Für  $\lambda > 1$  ist*

$$\frac{|C_{\max}| - n\zeta_\lambda}{\sqrt{n}} \rightarrow \mathfrak{N}\left(0, \frac{\zeta_\lambda(1 - \zeta_\lambda)}{(1 - \lambda + \lambda\zeta_\lambda)^2}\right) \quad \text{in Verteilung.}$$

Analog zum Beweis von Satz 4.1 wird zunächst ein Zentraler Grenzwertsatz für allgemeinere Folgen von binomialverteilten Zufallsgrößen angegeben. Über die Darstellung der Verteilung des Explorationsprozesses durch eine Folge von solchen Binomialverteilungen in Lemma 2.12 lässt sich dann ein Zentraler Grenzwertsatz für verschiedene Teilfolgen des Explorationsprozesses zeigen. Ausgehend davon wird dann ein Zentraler Grenzwertsatz für die Größe des von den ersten  $k_n$  Knoten aufgespannten Teilgraphen gezeigt und auf die Größe der größten Komponente über den Vergleich dieser Größen aus Abschnitt 4.4 übertragen.

**Lemma 5.2.** *Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge unabhängiger Zufallsgrößen mit*

$$X_n \sim \mathfrak{B}_{a_n, p_n}$$

*und  $a_n p_n(1 - p_n) \rightarrow \infty$ . Dann gilt*

$$\frac{X_n - a_n p_n}{\sqrt{a_n p_n(1 - p_n)}} \rightarrow \mathfrak{N}(0, 1) \quad \text{in Verteilung.}$$

**Beweis:** Diese Aussage wird genau wie der bekannte Zentrale Grenzwertsatz für Summen unabhängiger, identisch verteilter Zufallsgrößen aus dem Stetigkeitssatz für Fouriertransformierte (z. B. [Als00, S. 45.2]) hergeleitet. Dafür sei  $(Y_{i,n})_{i \in N}$  für jedes  $n \in N$  eine Folge unabhängiger,  $\mathfrak{B}_{1,p_n}$ -verteilter Zufallsgrößen, so dass

$$X_n = \sum_{i=1}^{a_n} Y_{i,n}.$$

Damit ist

$$\begin{aligned} \varphi_{\frac{X_n - a_n p_n}{\sqrt{a_n p_n (1-p_n)}}}(t) &= \mathbb{E} e^{it \frac{X_n - a_n p_n}{\sqrt{a_n p_n (1-p_n)}}} \\ &= \prod_{i=1}^{a_n} \mathbb{E} e^{it \frac{Y_{i,n} - p_n}{\sqrt{a_n p_n (1-p_n)}}} \\ &= \left( \mathbb{E} \left( 1 + it \frac{Y_{i,n} - p_n}{\sqrt{a_n p_n (1-p_n)}} - t^2 \frac{(Y_{i,n} - p_n)^2}{2a_n p_n (1-p_n)} + \mathcal{O}\left(a_n^{-\frac{3}{2}}\right) \right) \right)^{a_n} \\ &= \left( 1 - \frac{t^2 + \mathcal{O}(1)}{2a_n} \right)^{a_n} \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}}, \end{aligned}$$

woraus die Behauptung mit Hilfe des Stetigkeitssatzes folgt. □

Aus diesem Lemma lässt sich jetzt zunächst für jedes  $t \in \mathbb{R}^+$  ein Zentraler Grenzwertsatz für den Prozess  $S_{\lfloor nt \rfloor}$  herleiten.

**Lemma 5.3.** *Für den Explorationsprozess  $(S_t)_{t \in N}$  mit Start in  $k_n \in \mathcal{O}(\sqrt{n})$  gilt für jedes  $t \in \mathbb{R}^+$*

$$\frac{S_{\lfloor nt \rfloor} - n\mu_t}{\sqrt{n\nu_t}} \rightarrow \mathfrak{N}(0, 1) \quad \text{in Verteilung.}$$

**Beweis:** Es gilt

$$\frac{S_{\lfloor nt \rfloor} - n\mu_t}{\sqrt{n\nu_t}} = \sqrt{\frac{\text{Var } S_{\lfloor nt \rfloor}}{n\nu_t}} \frac{S_{\lfloor nt \rfloor} - \mathbb{E}S_{\lfloor nt \rfloor}}{\sqrt{\text{Var } S_{\lfloor nt \rfloor}}} + \frac{S_{\lfloor nt \rfloor} - n\mu_t}{\sqrt{\text{Var } S_{\lfloor nt \rfloor}}}.$$



---

Da der führende Faktor wegen Lemma 4.8 gegen 1 und der letzte Summand wegen Lemma 4.7 gegen 0 konvergieren, liefert der Satz von Slutsky (z. B. [Als00, Satz 36.12]) zusammen mit dem zuvor gezeigten Zentralen Grenzwertsatz für Binomialverteilungen (Lemma 5.2) die Behauptung.  $\square$

**Lemma 5.4.** *Sei  $k_n = \ln n$ . Dann gilt*

$$\frac{|C_{\leq k_n}| - \zeta_\lambda n}{\sqrt{n}} \rightarrow \mathfrak{N}(0, \sigma_\lambda^2) \quad \text{in Verteilung}$$

mit

$$\sigma_\lambda^2 = \frac{\zeta_\lambda(1 - \zeta_\lambda)}{(1 - \lambda - \zeta_\lambda)^2}.$$

**Beweis:** Hier benötigt man obere und untere Schranken für die Wahrscheinlichkeit

$$\begin{aligned} & P_\lambda \left( \frac{|C_{\leq k_n}| - \zeta_\lambda n}{\sqrt{n}} > x \right) \\ &= P_\lambda(|C_{\leq k_n}| > m_x) \end{aligned}$$

mit  $m_x = \lfloor n\zeta_\lambda + x\sqrt{n} \rfloor$ . Für die obere Schranke sieht man analog zum Beweis von Lemma 4.11

$$\begin{aligned} P_\lambda(|C_{\leq k_n}| > m_x) &\leq P_\lambda(S_{m_x} > 0) \\ &= P_\lambda \left( \frac{S_{m_x} - \mathbb{E}_\lambda S_{m_x}}{\sqrt{\text{Var}_\lambda S_{m_x}}} > -\frac{\mathbb{E}_\lambda S_{m_x}}{\sqrt{\text{Var}_\lambda S_{m_x}}} \right). \end{aligned}$$

In Lemma 4.7 wurde

$$\mathbb{E}_\lambda S_{m_x} = (\lambda(1 - \zeta_\lambda) - 1)\sqrt{n}x$$

gezeigt und da  $\nu_t$  in  $t$  differenzierbar ist, ist

$$\nu_{\zeta_\lambda + xn}^{-\frac{1}{2}} = \nu_{\zeta_\lambda} + \mathcal{O}(1)$$

und damit mit Hilfe von Lemma 4.8

$$\text{Var } {}_{\lambda}S_{m_x} = n\nu_{\zeta_{\lambda}} + o(n),$$

so dass

$$\begin{aligned} & P_{\lambda} \left( \frac{S_{m_x} - \mathbb{E}_{\lambda} S_{m_x}}{\sqrt{\text{Var } {}_{\lambda}S_{m_x}}} > -\frac{\mathbb{E}_{\lambda} S_{m_x}}{\sqrt{\text{Var } {}_{\lambda}S_{m_x}}} \right) \\ &= P_{\lambda} \left( \frac{S_{m_x} - \mathbb{E}_{\lambda} S_{m_x}}{\sqrt{\text{Var } {}_{\lambda}S_{m_x}}} > -\frac{(\lambda(1 - \zeta_{\lambda}) - 1)x}{\sqrt{\nu_{\zeta_{\lambda}}}} + o(1) \right). \end{aligned}$$

Da die Zufallsgröße auf der linken Seite nach Lemma 5.3 in Verteilung gegen eine standardnormalverteilte Zufallsgröße konvergiert, ist schließlich

$$\lim_{n \rightarrow \infty} P_{\lambda} \left( \frac{|C_{\leq k_n}| - \zeta_{\lambda} n}{\sqrt{n}} > x \right) \leq P(Z > x),$$

mit  $Z \sim \mathfrak{N}(0, \sigma_{\lambda}^2)$ .

Für die untere Schranke greift man auf die im Beweis von Lemma 4.14 geleistete Vorarbeit zurück und erhält für jedes  $\varepsilon > 0$  ein  $\gamma < \zeta_{\lambda}$  und  $\delta > 0$ , so dass

$$\begin{aligned} P_{\lambda}(|C_{\leq k_n}| > m_x) &\geq P_{\lambda}(S_{m_x} > \varepsilon\sqrt{n}) - P_{\lambda}(S_{m_x} > \varepsilon\sqrt{n}, \exists m \leq m_x : S_m = 0) \\ &\geq P_{\lambda}(S_{m_x} > \varepsilon\sqrt{n}) - \sum_{m=0}^{\gamma n} e^{-\frac{\delta^2(m+k_n-1)}{2\lambda}} - \sum_{m=\gamma n}^{m_x} e^{-\frac{\varepsilon^2}{2}\sqrt{n}}. \end{aligned}$$

Wie im Fall der moderaten Abweichungen sind die Restterme hinreichend klein, wenn  $k_n = K \ln n$  mit hinreichend großem  $K$ , denn

$$\begin{aligned} e^{-\frac{\delta^2(m+K \ln n-1)}{2\lambda}} &\leq n^{-\frac{\delta^2 K}{2\lambda}} \in o(n^{-1}) \\ \text{und } e^{-\frac{\varepsilon^2}{2}\sqrt{n}} &\in o(n^{-1}) \end{aligned}$$

und somit

$$\sum_{m=0}^{\gamma n} e^{-\frac{\delta^2(m+k_n-1)}{2\lambda}} + \sum_{m=\gamma n}^{m_x} e^{-\frac{\varepsilon^2}{2}\sqrt{n}} \in o(1).$$

Analog zu den Schritten, die zu der oberen Schranke geführt haben, sieht man, dass

$$\lim_{n \rightarrow \infty} P_\lambda(S_{m_x} > \varepsilon \sqrt{n}) = P\left(Z > x + \frac{\varepsilon}{\sqrt{\nu_{\zeta_\lambda}}}\right).$$

Deswegen ist

$$\begin{aligned} & \lim_{n \rightarrow \infty} P_\lambda(|C_{\leq k_n}| > m_x) \\ & \geq \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \left( P_\lambda(S_{m_x} > \varepsilon \sqrt{n}) - \sum_{m=0}^{\gamma n} e^{-\frac{\delta^2(m+k_n-1)}{2\lambda}} - \sum_{m=\gamma n}^{m_x} e^{-\frac{\varepsilon^2}{2}\sqrt{n}} \right) \\ & = P(Z > x), \end{aligned}$$

was die Behauptung zeigt. □

**Beweis von Satz 5.1:** Der Beweis des zentralen Grenzwertsatzes für die Größe der größten Komponente ergibt sich jetzt aus Lemma 5.4 mit Hilfe von Lemma 4.18, denn

$$\begin{aligned} P_\lambda(|C_{\max}| > m_x) &= P_\lambda(|C_{\max}| - |C_{\leq k_n}| + |C_{\leq k_n}| > m_x). \\ &\leq P_\lambda(|C_{\max}| - |C_{\leq k_n}| + |C_{\leq k_n}| > m_x, |C_{\max}| < |C_{\leq k_n}|) + \mathcal{O}(1) \\ &\leq P(|C_{\leq k_n}| > m_x) + \mathcal{O}(1). \end{aligned}$$

Eine Schranke der selben Art nach unten wird durch eine analoge Rechnung und Anwendung von Lemma 4.19 bewiesen. Insgesamt ist

$$\begin{aligned} \lim_{n \rightarrow \infty} P_\lambda(|C_{\max}| > m_x) &= \lim_{n \rightarrow \infty} P_\lambda(|C_{\leq k_n}| > m_x) \\ &= \mathfrak{N}(0, \sigma_\lambda^2)((x, \infty)), \end{aligned}$$

was zu zeigen war. □



## Literatur

- [Ald97] David J. Aldous. „Brownian Excursions, Critical Random Graphs and the Multiplicative Coalescent“. In: *Ann. Probab.* 25 (1997), S. 812–854.
- [Als00] Gerold Alsmeyer. *Wahrscheinlichkeitstheorie*. 2000.
- [Als02] Gerold Alsmeyer. *Stochastische Prozesse*. 2002.
- [Ame06] Jens Ameskamp. „Grenzwertsätze in magnetischen Modellen“. Diplomarbeit. 2006.
- [AS00] Noga Alon und Joel Spencer. *The Probabilistic Method*. 2. Aufl. Wiley, 2000.
- [BBV00] D. Barraez, S. Boucheron und W. Fernandez de la Vega. „On the Fluctuations of the Giant Component“. In: *Combinatorics, Probability and Computing* 9.04 (2000), S. 287–304.
- [Bol85] Béla Bollobás. *Random graphs*. London: Academic Press Inc., 1985.
- [Die06] Reinhard Diestel. *Graph Theory (Graduate Texts in Mathematics)*. Springer Verlag, 2006.
- [Dwa69] Meyer Dwass. „The total progeny in a branching process and a related random walk“. In: *Journal of Applied Probability* 6 (1969), S. 682–686.
- [EL03] Peter Eichelsbacher und Matthias Löwe. „Moderate Deviations for I.I.D. Random Variables“. In: *ESAIM, Probability and Statistics* 7 (2003), S. 209–218.

- [Ell85] Richard S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Grundlagen der mathematischen Wissenschaften 271. Springer Verlag, 1985.
- [ER59] Paul Erdős und Alfréd Rényi. „On Random Graphs I“. In: *Publ. Math. Debrecen* 6 (1959), S. 290–297.
- [ER60] Paul Erdős und Alfréd Rényi. „On the evolution of random graphs“. In: *Publ. Math. Inst. Hung. Acad. Sci* 5 (1960), S. 17–61.
- [ER61] Paul Erdős und Alfréd Rényi. „On the evolution of random graphs“. In: *Bull. Inst. Int. Statist. Tokyo* 38 (1961), S. 343–347.
- [Erd47] Paul Erdős. „Some remarks on the theory of graphs“. In: *Bull. Amer. Math. Soc* 53 (1947), S. 292–294.
- [Erd57] Paul Erdős. „Some remarks on a theorem of Ramsey“. In: *Bull. Res. Council Isreal, Section F* 7 (1957), S. 21–24.
- [Gil59] Edgar N. Gilbert. „Random Graphs“. In: *Ann. Math. Statist.* 30.4 (1959), S. 1141–1144.
- [Heu00] Harro Heuser. *Lehrbuch der Analysis, Teil 2*. 11. Aufl. Teubner, 2000.
- [Hof09] Remco van der Hofstad. „Random Graphs and Complex Networks“. 2009.
- [JLR00] Svante Janson, Tomasz Łuczak und Andrzej Ruciński. *Random Graphs*. Wiley, 2000.
- [Łuc90] Tomasz Łuczak. „On the equivalence of two basic models of random graphs“. In: *Random graphs '87 (Poznań, 1987)*. Chichester: Wiley, 1990, S. 151–157.
- [ML98] Anders Martin-Löf. „The Final Size of a Nearly Critical Epidemic, and the First Passage Time of a Wiener Process to a Parabolic Barrier“. In: *Journal of Applied Probability* 35.3 (1998), S. 671–682.

- 
- [O’C98] Neil O’Connell. „Some large deviation results for sparse random graphs“. In: *Probab Theory Relat Fields* 110 (1998), S. 277–285.
- [OV05] Enzo Olivieri und Maria Eulália Vares. *Large deviations and metastability*. Bd. 100. Encyclopedia of Mathematics and its Applications. Cambridge: Cambridge University Press, 2005.
- [Pit90] Boris Pittel. „On tree census and the giant component in sparse random graphs“. In: *Random Structures Algorithms* 1.3 (1990), S. 311–342.
- [Tho00] Hermann Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications. Springer Verlag, 2000.
- [Var85] S.R. Srinivasa Varadhan. *Large Deviations and Applications*. Lecture Notes in Mathematics 1362. Springer Verlag, 1985.

