

**Bachelor-Arbeit**

Motivvariation in komplexen Netzwerken  
und ihre Auswirkung  
auf den Informationsgehalt

*Christian M. Meyer*



**Technische Universität Darmstadt**  
Fachbereich Informatik  
Fachgebiet Algorithmik

Name: Christian M. Meyer  
Matrikelnummer: 1200498  
Betreuer: Dr. habil. Matthias Müller-Hannemann  
Abgabedatum: 10. August 2006



## **Zusammenfassung**

Die Arbeit diskutiert die Auswirkungen von Netzwerkmotiven auf den Informationsgehalt eines Graphen. Es werden hauptsächlich Motive der Größe 3 in gerichteten Graphen betrachtet, welche mit Hilfe zweier vorgestellter Algorithmen variiert werden. Die Auswertung des Informationsgehalts erfolgt mit binären zellulären Automaten, deren Verhalten durch die Shannon- und die Word-Entropie analysiert wird. Die Entropien werden nach jedem Variationsschritt ermittelt und in einem Punktediagramm aufgetragen. Es ergibt sich eine Trajektorie, die die Veränderung des Informationsgehalts angibt.

Ziel ist die Untersuchung eines direkten Zusammenhangs zwischen Motiven und Informationsgehalt. Es ergab sich, dass bestimmte Motivklassen dazu neigen größere Entropien als andere zu bilden, insbesondere wurde hier das Motiv „Feed-forward loop“ näher untersucht. Darüber hinaus ergab sich, dass die Zustandsübergangsfunktion und die Existenz von Zyklen im Graphen einen großen Einfluss auf die Entropien bewirken.

## **Abstract**

This thesis discusses the impact of network motifs on the information processing capacity within graphs. In directed graphs motifs of size 3 are studied and varied using two presented algorithms. Binary cellular automata are applied on the graphs to analyse the dynamics of the network. The information processing capacity is measured by shannon and word entropy, which are determined after every step of variation and visualized in a scatter diagram. A trajectory emerges, which represents the changes in the information processing capacity.

Studying a direct correlation of network motifs and the information processing capacity is the objective of the thesis. Some motif classes show the ability of establishing greater entropies than others. The motif “feed-forward loop” was analyzed particularly in this context. Moreover it turned out, that the used transition function as well as the presence of cycles in the graphs has a big impact on the entropies.



## Erklärung

Ich versichere, dass ich diese Bachelor-Arbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Darmstadt, 10. August 2006  
\_\_\_\_\_  
Ort, Datum

\_\_\_\_\_  
Christian M. Meyer



## Danksagung

Mein besonderer Dank gilt Sandra Ebert für die gute Zusammenarbeit bei der Entwicklung der Software MOTIFCONTROL, deren Auswertungsmöglichkeiten diese Arbeit überhaupt erst ermöglicht haben.

Außerdem danke ich Herrn Prof. Marc-Thorsten Hütt und Herrn Carsten Marr für die zahllosen Ideen und Vorschläge, die mich bei der Analyse der Entropien jedes Mal ein gutes Stück weitergebracht haben.

Ich danke meinem Betreuer Herrn Dr. Matthias Müller-Hannemann für die vielen Gespräche und konstruktiven Diskussionen der Arbeit sowie für die ausgezeichnete Betreuung während der Erstellung.

Meinen Eltern danke ich für die finanzielle Unterstützung während meines Studiums.



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>3</b>
<b>2</b>	<b>Grundlagen</b>	<b>5</b>
2.1	Graphen . . . . .	5
2.2	Zelluläre Automaten . . . . .	9
2.3	Analysemethoden . . . . .	12
<b>3</b>	<b>Netzwerkmotive</b>	<b>15</b>
3.1	Definition . . . . .	15
3.2	Klassifikation von Motivverteilungen . . . . .	18
<b>4</b>	<b>Variation von Netzwerkmotiven</b>	<b>25</b>
4.1	Auswirkungen der Variation . . . . .	25
4.2	Ansatz und Probleme . . . . .	26
4.3	Heuristischer Ansatz . . . . .	29
4.4	Motivtransformation . . . . .	29
4.5	Ergebnisse . . . . .	30
<b>5</b>	<b>Auswirkungen auf die Entropieebene</b>	<b>31</b>
5.1	Testprotokolle . . . . .	31
5.2	Reguläre Kreisgraphen . . . . .	32
5.3	Erdős-Rényi-Graphen . . . . .	36
5.4	Skalenfreie Graphen . . . . .	43
5.5	Beitrag der einzelnen Motivklassen zur Entropie . . . . .	45
<b>6</b>	<b>Fazit</b>	<b>53</b>
	<b>Abbildungsverzeichnis</b>	<b>55</b>
	<b>Tabellenverzeichnis</b>	<b>55</b>
	<b>Literaturverzeichnis</b>	<b>56</b>



# 1 Einführung

Die Arbeit beschäftigt sich mit zusammenhängenden Subgraphen einer bestimmten Größe, so genannten *Netzwerkmotiven*. Diese werden in regulären Kreisgraphen, zufälligen Erdős-Rényi-Graphen sowie in skalenfreien Netzen betrachtet. Die Netzwerke werden variiert, um die Motivverteilung zu verändern und mittels *zellulärer Graphenautomaten* auf ihren Informationsgehalt hin untersucht. Als Maß dienen die *Shannon-* und die *Word-Entropie*.

Zelluläre Automaten sind ein theoretisch wohl fundiertes und gut erforschtes Konzept der Modellierung, welches beispielsweise in der Simulation von

- Diffusionsphänomenen,
- Epidemien,
- Evolutionstheorien oder
- Raumzeitlichen Populationsdynamiken

eingesetzt wird [Hüt01]. Die Untersuchung von Netzwerkmotiven ist dagegen noch eine recht junge und viel versprechende Disziplin.

Ziel dieser Arbeit ist die Verbindung der beiden Konzepte. Insbesondere soll ein möglicher Zusammenhang zwischen Motivverteilung und dynamischem Verhalten des Automaten untersucht werden. Bei bestehendem Zusammenhang könnte der Informationsgehalt eines Netzwerks durch Einstreuen bestimmter Motive oder Motivkonstellationen nachhaltig gesteuert werden.

Außerdem ist man in der Lage auch sehr große Netzwerke einfach anhand der kleinen lokalen Motivstrukturen zu untersuchen und zu klassifizieren. Gerade in der Biologie treten häufig sehr große Netzwerke auf, in denen die Informationsverarbeitung näher untersucht werden soll. Ein Zusammenhang zu der Motivverteilung könnte dabei helfen eine Art Karte des Netzwerks zu erstellen, aus welcher Stellen mit besonders großem Informationsdurchsatz hervortreten.

Die Arbeit gliedert sich wie folgt:

- Kapitel 2: Einführung der grundlegenden Begriffe aus der Graphen- und Automatentheorie, die zum Verständnis der Arbeit notwendig sind. Insbesondere werden *Graph*, *zellulärer Graphenautomat* und *Entropie* eingeführt. Weiterhin finden sich hier die Aufbauschemata bestimmter Graphenklassen, die in der Arbeit verwendet werden.
- Kapitel 3: Definition der Netzwerkmotive und einiger Auswertungsmöglichkeiten. Für bestimmte Graphenklassen werden dann die Motivverteilungen untersucht und klassifiziert. Die verwendeten Motivklassen werden in ihrer graphischen Repräsentation dargestellt.
- Kapitel 4: Das Kapitel beschreibt Algorithmen zur Variation der Motivverteilung eines Graphen. Außerdem werden Erkenntnisse bereits durchgeführter Untersuchungen zur topologischen Variation von Graphen beschrieben.
- Kapitel 5: Der Hauptteil der Arbeit beschäftigt sich mit den Auswirkungen der Motivvariation auf den Informationsgehalt des Netzwerks. Die Experimente werden anhand von Testprotokollen dargestellt und diskutiert, die Ergebnisse werden als Beobachtungen formuliert und am Ende eines jeweiligen Abschnitts zusammengestellt.
- Kapitel 6: Zusammenfassung der Beobachtungen und der gewonnenen Erkenntnisse. Außerdem finden sich hier einige Anregungen für weiterführende Untersuchungen.



## 2 Grundlagen

Zunächst werden Graphen eingeführt und einige charakteristische Aufbauschemata für Graphen betrachtet. Zur Auswertung des Informationsgehalts in einem Graphen sind zelluläre Automaten nötig, die im Anschluss eingeführt werden. Die dazu verwendeten Analysemethoden folgen dann in Abschnitt 2.3.

### 2.1 Graphen

Graphen sind mittlerweile in sehr vielen Forschungsgebieten zu einem wichtigen Hilfsmittel geworden. Angefangen im Operations Research, über Linguistik, bis hin zur Informatik und Mathematik zieht sich das Verbreitungsspektrum. Graphen erlauben es beispielsweise elektrische Netzwerke auf die gleiche Weise wie abstrakte Straßenkarten oder – wie in dieser Arbeit – biologische Netzwerke zu untersuchen.

Die Problemanalyse und -lösung durch Graphenstrukturen geht bis ins 18. Jahrhundert zurück, als LEONHARD EULER 1736 eine notwendige Bedingung zu seinem berühmten Königsberger Brückenproblem [Eul36] formulierte. Die eigentliche *Graphentheorie* gewann allerdings erst gut hundert Jahre später an Bedeutung. Hauptsächlich trugen dazu die Arbeiten von GUSTAV KIRCHHOFF [Kir47] im Jahre 1847 und von ARTHUR CAYLEY [Cay57] 1857 bei.

#### 2.1.1 Notation

Die große Verbreitung der Graphentheorie hat zu einer Vielzahl an Begriffen und Definitionen geführt, so dass ein eigenes umfangreiches Vokabular entstanden ist. Die wenigen Begriffe, die zum Verständnis dieser Arbeit erforderlich sind, werden im Folgenden eingeführt. Darüber hinaus geben beispielsweise [Wil75, Pre97, Wal03] einen guten Überblick der gebräuchlichsten Definitionen.

**Definition 2.1 (Graph)** *Ein (endlicher) Graph  $G$  ist ein Paar  $G = (V, E)$  bestehend aus*

*einer (endlichen) Knotenmenge  $V$  und  
einer (endlichen) Kantenmenge  $E$ .*

*Elemente von  $V$  heißen Knoten, Elemente von  $E$  Kanten. Man unterscheidet gerichtete Graphen mit  $E \subseteq V \times V$  (Knotenpaare) und ungerichtete Graphen, die  $E \subseteq \mathcal{P}_2(V)$  (zweielementige Teilmengen von  $V$ ) definieren.*

**Definition 2.2 (Graphennachbarschaft)** *Sei  $G = (V, E)$  ein Graph. Eine Funktion, die zu jedem Knoten aus  $G$  die über eine Kante verbundenen Nachbarn ermittelt, heißt Graphennachbarschaft  $N$  von  $G$ . Damit gilt:*

*Gerichteter Graph:  $N: V \rightarrow \mathcal{P}(V), v \mapsto \{u \in V \mid (u, v) \in E \vee (v, u) \in E\}$ .  
Ungerichteter Graph:  $N: V \rightarrow \mathcal{P}(V), v \mapsto \{u \in V \mid \{u, v\} \in E\}$ .*

Die Definition der Graphennachbarschaft wird in Definition 2.15 verallgemeinert.

**Definition 2.3 (Knotengrad)** *Sei  $G = (V, E)$  ein Graph und  $v \in V$  ein Knoten. Der Grad  $\deg(v)$  des Knotens  $v$  ist definiert als die Zahl der ein- und ausgehenden Kanten, folglich gilt*

$$\deg(v) = |N(v)|.$$

**Definition 2.4 (Gradverteilung)** *Sei  $G = (V, E)$  ein Graph. Die Funktion*

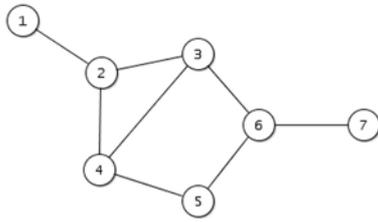
$$d: \mathbb{N} \rightarrow \mathbb{N}, \quad n \mapsto |\{v \mid \deg(v) = n\}|$$

*heißt Gradverteilung. Die Knotengrade werden dabei gezählt und entsprechend ihrer Größe aufgetragen.*

Die Abbildungen 3 und 4 zeigen die Gradverteilung für einen Erdős-Rényi-Graphen und einen skalenfreien Graphen.

Zur Visualisierung eines Graphen werden seine Knoten in der Regel als Kreise gezeichnet, während eine Kante zwei dieser Kreise verbindet.

**Beispiel 2.5** *Abbildung 1 zeigt die Visualisierung eines einfachen Graphen, sowie einige Beispiele obiger Definitionen.*



$$\begin{aligned}
 V &= \{1, 2, 3, 4, 5, 6, 7\} \\
 E &= \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \\
 &\quad \{3, 6\}, \{4, 5\}, \{5, 6\}, \{6, 7\}\} \\
 N(2) &= \{1, 3, 4\} \\
 \text{deg}(2) &= 3 \\
 d &= \{(1 \mapsto 2), (2 \mapsto 1), (3 \mapsto 4), (4 \mapsto 0), \dots\}
 \end{aligned}$$

Abbildung 1: Beispiel eines Graphen

### 2.1.2 Aufbau

Für die Untersuchungen in den folgenden Kapiteln werden Graphen eingesetzt, die einem bestimmten charakteristischen Aufbauschema folgen. Diese Schemata werden in diesem Abschnitt erläutert.

**Bemerkung 2.6** *Da die Aufbauvorschriften sowohl für gerichtete als auch für ungerichtete Graphen auf gleiche Weise funktionieren, wird jeweils nur einer der beiden Fälle betrachtet. Die einzelnen Definitionen lassen sich über folgende Vorschrift übertragen:*

$$(\text{Gerichteter Graph}) \quad (u, v) \in E \wedge (v, u) \in E \quad \rightsquigarrow \quad \{u, v\} \in E \quad (\text{Ungerichteter Graph})$$

**Definition 2.7 (2k-regulärer Kreisgraph)** *Ein 2k-regulärer Kreisgraph  $G_k = (V, E)$ ,  $k \geq 1$  ist ein endlicher Graph mit der Knotenmenge  $V = \{1, 2, \dots, n\}$ ,  $n \in \mathbb{N}$  und der Kantenmenge:*

$$E = \{(u, v) \in V \times V \mid v \equiv u \pm i \pmod{n}, 1 \leq i \leq k\}$$

In der Graphentheorie ist die Definition eines *k-regulären Graphen* bekannt. Dies ist ein Graph, in dem jeder Knoten den Grad  $k$  hat. Jeder 2k-regulärer Kreisgraph bildet einen 2k-regulären Graphen. Eine weitere Einschränkung ist hier jedoch, dass, ausgehend von einem Kreis mit  $n$  Kanten, jeder Knoten mit allen Knoten, die über höchstens  $k$  Kanten erreicht werden können, benachbart ist. Die Gradverteilung hat damit die Gestalt einer Deltafunktion. In [Mar05] wird der 2k-reguläre Kreisgraph einfach als regulärer Graph bezeichnet.

**Beispiel 2.8** *Abbildung 2 zeigt den 4-regulären Kreisgraphen mit 10 Knoten.*

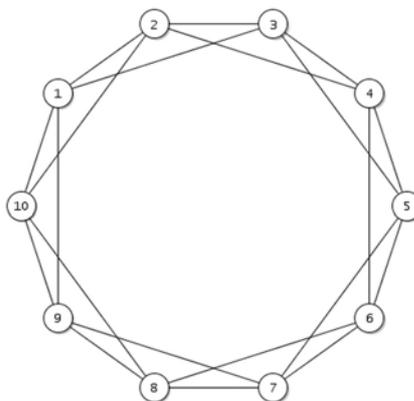


Abbildung 2: 4-regulärer Kreisgraph mit 10 Knoten

**Definition 2.9 (Erdős-Rényi-Graph)** Sei  $G = (V, E)$  ein Graph mit  $n = |V|$ .  $G$  heißt Erdős-Rényi-Graph (ER-Graph) mit Parameter  $p$ , falls jedes Knotenpaar  $(u, v) \in V \times V$  mit der Wahrscheinlichkeit  $p$  durch eine Kante verbunden ist.

Erdős-Rényi-Graphen wurden erstmals 1959 in [Erd59] von PAUL ERDŐS und ALFRÉD RÉNYI eingeführt. Der Knotengrad eines jeden Knotens liegt – bedingt durch die Gleichverteilung im Algorithmus – nahe am mittleren Knotengrad und folgt damit einer Gauß'schen Normalverteilung der Form

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}, \text{ mit Standardabweichung } \sigma \text{ und Erwartungswert } \mu.$$

**Beispiel 2.10** Abbildung 3 zeigt einen ER-Graphen mit 15 Knoten mit seiner (blau) und der typischen Gradverteilung (grau).

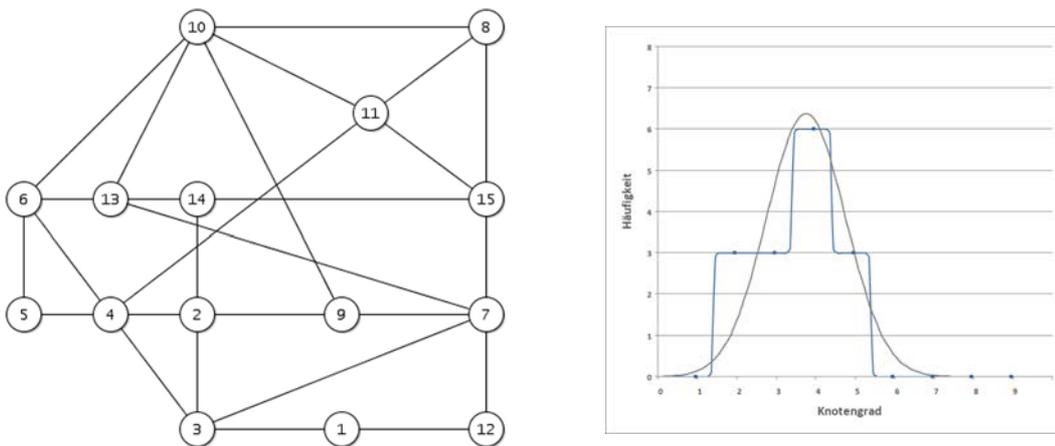


Abbildung 3: Erdős-Rényi-Graph mit 15 Knoten

**Definition 2.11 (Skalenfreier Graph)** Sei  $G = (V, E)$  ein Graph mit einer kleinen Anzahl  $m_0$  von Knoten und einer leeren Kantenmenge  $E$ . Nun wird diesem Graphen in jedem Schritt ein neuer Knoten hinzugefügt. Der neue Knoten ist mit  $m$  Kanten ausgestattet, welche mit existierenden Knoten verbunden werden. Dabei ist die Verbindungswahrscheinlichkeit proportional zum Knotengrad des Zielknotens, insbesondere kann die Verbindungswahrscheinlichkeit  $p$  eines Zielknotens  $v' \in V$  wie folgt gewählt werden:

$$p = \frac{\deg(v')}{\sum_{v \in V} \deg(v)}.$$

Einen Graphen, der mit diesem Algorithmus erzeugt wurde, nennt man skalenfreier Graph oder nach seinen Entdeckern auch Barabási-Albert-Graph (BA-Graph).

Skalenfreie Graphen wurden 1999 von ALBERT-LÁSZLÓ BARABÁSI und REKA ALBERT [Bar99] eingeführt. Die Gradverteilung folgt hier einem Potenzgesetz der Form

$$\lambda e^{-\lambda x} \text{ mit Ausfallrate } \lambda \in \mathbb{R}.$$

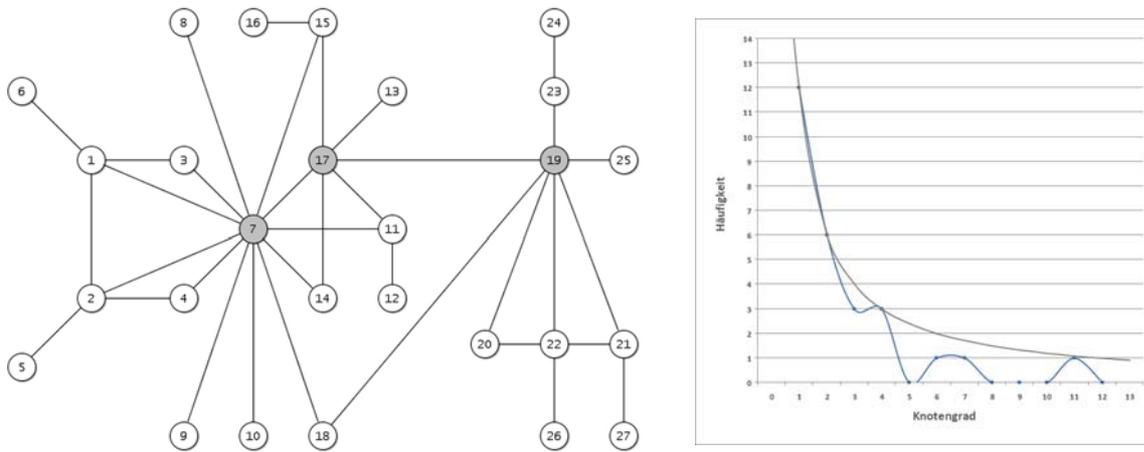


Abbildung 4: Skalenfreier Graph mit 27 Knoten

**Beispiel 2.12** *Abbildung 4 zeigt einen BA-Graphen mit 27 Knoten. Die Gradverteilung ist wieder blau, eine typische Verteilung grau dargestellt.*

Einen guten Überblick über ER- und BA-Graphen bietet beispielsweise [Bar04].

## 2.2 Zelluläre Automaten

### 2.2.1 Notation

Automaten sind in der Informatik ein verbreitetes Mittel zur Analyse und Simulation von Problemen. Es handelt sich um eine gedachte Maschine, die von einem realen oder erfundenen System abstrahiert und sich damit auf die für die Simulation nötigen Aspekte konzentriert. Dadurch wird eine Problemstellung besser verständlich und kann somit leichter analysiert werden. Der hier verwendete Automatenbegriff wird wie folgt definiert:

**Definition 2.13 (Automat)** Ein (endlicher) Automat  $\mathcal{A}$  ist ein 5-Tupel  $\mathcal{A} = (S, I, O, \delta, \lambda)$  bestehend aus

- der (endlichen) Zustandsmenge  $S$ ,
- dem (endlichen) Eingabealphabet  $I$ ,
- dem (endlichen) Ausgabealphabet  $O$ ,
- der Transitionsfunktion  $\delta: S \times I \rightarrow S$  und
- der Ausgabefunktion  $\lambda: S \times I \rightarrow O$ ,

wobei  $S$ ,  $I$  und  $O$  paarweise disjunkte Mengen sind.

Die Transitionsfunktion weist einem Zustand unter Berücksichtigung der aktuellen Eingabe einen neuen Zustand zu. Die Funktion  $\delta$  heißt daher auch (Zustands-)Übergangsfunktion.

Meist werden Automaten zur Erkennung von Sprachen im Bereich der formalen Sprachentheorie genutzt. Zur graphischen Repräsentation eines Automaten werden seine Zustände als Knoten und die Transitionsfunktion als Kanten eines gerichteten Graphen aufgefasst. Ausgaben werden in eckige Klammern gefasst.

**Beispiel 2.14** Der Automat in Abbildung 5 erkennt die reguläre Sprache  $a^*b(bb)^*$ . Bei dem gezeigten Automaten handelt es sich um einen NFA (non-deterministic finite automaton), der obige Definition um einen dezidierten Anfangszustand erweitert, welcher aber hier nicht weiter benötigt wird. Das Beispiel ist [Otto04] entnommen.

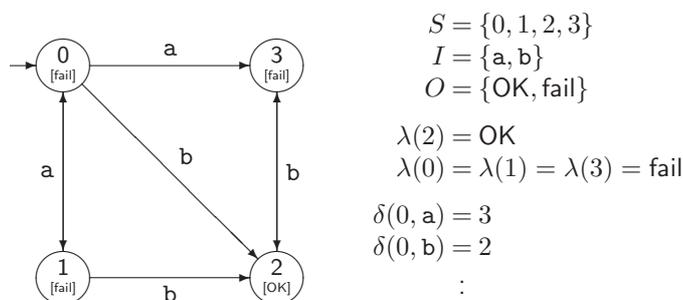


Abbildung 5: Beispiel eines Automaten

Näheres zu Automaten und formalen Sprachen findet sich beispielsweise in [Sch97]. Es folgen zwei kleinere Definitionen, welche zur Definition des zellulären Automaten benötigt werden:

**Definition 2.15 (Nachbarschaft)** Sei  $X$  eine Menge. Eine Nachbarschaft  $N$  auf  $X$  ist eine Funktion

$$N: X \rightarrow \mathcal{P}(X).$$

Die Menge  $N(x)$ , also das Bild von  $x \in X$ , heißt Nachbarmenge von  $x$ . Elemente  $y \in N(x)$  dieser Menge sind Nachbarn zu  $x$ , man sagt  $y$  ist benachbart mit  $x$ .

Insbesondere ist die in Definition 2.2 eingeführte Graphennachbarschaft eine Nachbarschaft im Sinne obiger Definition.

**Definition 2.16 (Zellraum)** Ein Zellraum  $R$  ist ein Tripel  $R = (C, S, \chi_0)$  bestehend aus

der Zellmenge  $C$ ,  
der Zustandsmenge  $S$  und  
einer Funktion  $\chi_0: C \rightarrow S$ ,

die jedem  $c \in C$  einen Zustand  $s = \chi_0(c) \in S$  zuordnet. Elemente der Zellmenge  $C$  heißen Zellen.

Zelluläre Automaten wurden erstmals 1940 von STANISLAW ULAM erwähnt. JOHN VON NEUMANN erweiterte Ulams Idee zu einem universellen Berechnungsmodell. Die in diesem Dokument verwendete Definition lautet wie folgt:

**Definition 2.17 (Zellulärer Automat)** Ein zellulärer Automat  $\mathcal{CA}$  ist ein Tripel  $\mathcal{CA} = (R, N, \delta_\chi)$  bestehend aus

einem Zellraum  $R = (C, S, \chi_0)$ ,  
einer Nachbarschaft  $N$  auf der Zellmenge  $C$  und  
der Transitionsfunktion  $\delta_\chi: C \times \mathcal{P}(C) \rightarrow S$  mit  $\chi: C \rightarrow S$ .

Die Transitionsfunktion ermittelt hier anders als bei dem zuvor betrachteten Automaten zu einer gegebenen Zelle und einiger Nachbarn einen neuen Zustand. Die Zustände der übergebenen Zellen können mit der  $\chi$ -Funktion ermittelt werden.

Die Startkonfiguration  $y_0$  eines zellulären Automaten ist eine Liste von Zuständen aller Zellen zum Zeitpunkt  $t = 0$ :

$$y_0 = (\chi_0(c_0), \chi_0(c_1), \dots)$$

Ausgehend von der Startkonfiguration  $y_0$  ergibt sich der Begriff der (Folge-)Konfiguration durch Anwendung der Transitionsfunktion  $\delta_\chi$ . So kann ein zeitlich dynamischer Prozess simuliert werden. Es gilt:

$$y_t = (\chi_t(c_0), \chi_t(c_1), \dots)$$

wobei

$$\chi_i: C \rightarrow S, \quad c \mapsto \delta_{\chi_{i-1}}(c, N(c)) \quad \text{für } i \geq 1$$

gesetzt wird. Insbesondere gilt damit  $\chi_1(c) = \delta_{\chi_0}(c, N(c))$ , wobei  $\chi_0$  die bereits bekannte Zustandsfunktion des Zellraums ist. Insgesamt kann man sagen, dass die Funktion  $\chi_t(c)$  den Zustand der Zelle  $c$  zum Zeitpunkt  $t$  angibt.

**Bemerkung 2.18** Der zelluläre Automat verknüpft mehrere endliche Automaten, um eine räumlich-zeitliche Simulation zu ermöglichen. Für jede Zelle  $c \in C$  wird ein endlicher Automat  $\mathcal{A}_c$  der Form

$$\mathcal{A}_c = (S, \mathcal{P}(C), S, \delta_\chi, \lambda)$$

definiert. Die Ausgabefunktion emittiert dabei den Zellzustand:

$$\lambda: S \times \mathcal{P}(C) \rightarrow S, \quad (s, N(c)) \mapsto s.$$

**Definition 2.19 (Binärer zellulärer Automat)** Ein binärer zellulärer Automat ist ein Automat mit nur zwei Zuständen. Meist wird die Zustandsmenge  $S = \{0, 1\}$  verwendet.

Ausgangspunkt der späteren Experimente ist ein zellulärer Graphenautomat, also ein zellulärer Automat, dem die Nachbarschaft eines Graphen unterliegt. Dieser Graphenautomat wird wie folgt modelliert:

**Definition 2.20 (Zellulärer Graphenautomat)** Ein zellulärer Graphenautomat  $\mathcal{GA}$  ist ein 4-Tupel  $\mathcal{GA} = (G, S, \chi_0, \delta_\chi)$  bestehend aus:

einem endlichen Graphen  $G = (V, E)$ ,  
einer Zustandsmenge  $S$ ,  
der Zustandsfunktion  $\chi_0: V \rightarrow S$  und  
der Transitionsfunktion  $\delta_\chi: C \times \mathcal{P}(C) \rightarrow S$  mit  $\chi: C \rightarrow S$ .

Dieser zelluläre Graphenautomat bildet einen zellulären Automaten  $\mathcal{CA} = (R, N, \delta_\chi)$  wobei  $R$  den Zellraum  $R = (V, S, \chi_0)$  der Knoten darstellt und  $N$  die Graphennachbarschaft des Graphen  $G$  repräsentiert.

## 2.2.2 Beispiel

Zum besseren Verständnis der zellulären Automatentheorie wird ein populäres Beispiel vorgestellt:

**Beispiel 2.21 (Game of Life)** *Der wohl bekannteste zelluläre Automat wurde von JOHN H. CONWAY in seinem „Spiel des Lebens“ (Game of Life) [Con70] verwendet. Hierbei handelt es sich um eine recht einfache Simulation von komplexen Wachstumsprozessen.*

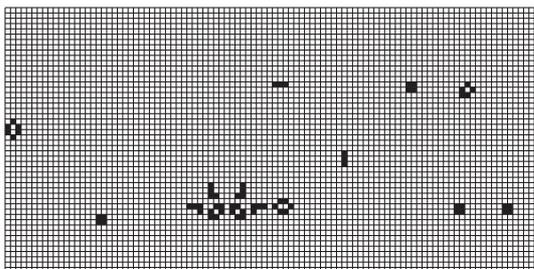


Abbildung 6: Conways „Game of Life“

Das „Game of Life“ definiert zwei Zustände  $S = \{\text{tot}, \text{lebendig}\}$ , die häufig als  $S = \{0, 1\}$  kodiert werden. In Abbildung 6 sind lebendige Zellen schwarz und tote Zellen weiß dargestellt.

Die Nachbarschaft ist als regelmäßiges viereckiges Gitter definiert. Zwei Zellen sind benachbart, wenn sie mindestens eine Ecke des Feldes gemeinsam haben, also alle Zellen, die ein Feld horizontal, vertikal oder diagonal von einer anderen Zelle entfernt sind. Diese Nachbarschaft wird als *Moore-Nachbarschaft* bezeichnet.

Die „Spielregeln“ werden durch die Transitionsfunktion  $\delta_\chi$  angegeben. Diese ist definiert als:

$$\delta_\chi(c, N(c)) = \begin{cases} 1 & \text{falls } \sum_{c' \in N(c)} \chi(c') = 3 \\ \chi(c) & \text{falls } \sum_{c' \in N(c)} \chi(c') = 2 \\ 0 & \text{sonst} \end{cases}$$

Abbildung 7 zeigt die Wirkung dieser Übergangsfunktion anhand von einigen Beispielen.

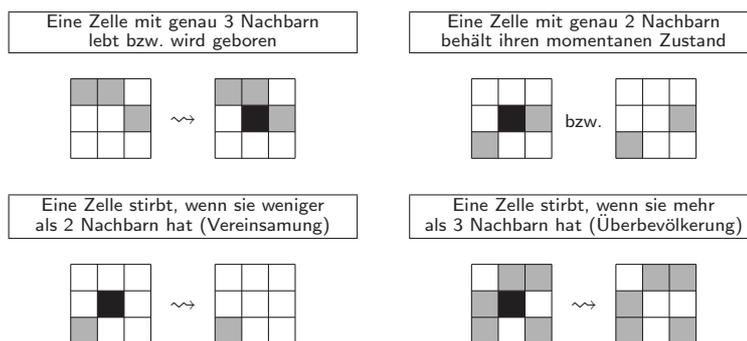


Abbildung 7: Transitionsfunktion im „Game of Life“

Visualisiert man nun die Konfigurationen des verwendeten zellulären Automaten über eine gewisse Anzahl an Zeitschritten, so zeigt das System eine erstaunliche Komplexität in Anbetracht des deterministischen Verhaltens der Transitionsfunktion.

Neben stabilen Endzuständen, können oszillierende Objekte (so genannte „Blinker“) auftreten, die nach einer bestimmten Anzahl von Zeitschritten wieder ihre Ausgangsgestalt annehmen. Außerdem können „Gleiter“ erzeugt werden, die sich innerhalb des Zellraums bewegen können. Auch ein völlig chaotisches Verhalten oder das vollständige Aussterben der Zellpopulation ist möglich.

## 2.3 Analysemethoden

### 2.3.1 Simulation

Die Transitionsfunktion  $\delta_\chi$  lässt die Simulation eines dynamischen Prozesses auf dem zellulären Automaten  $\mathcal{CA} = (R, N, \delta_\chi)$  zu. Die Zustände des Zellraums bilden dabei ein *räumliches Muster*, die Betrachtung über eine gewisse Zeit ein *zeitliches Muster*.

Die beiden folgenden Definitionen bilden die Terminologie einer solchen Simulation:

**Definition 2.22 (Simulation)** Die Simulation der Dynamik eines zellulären Automaten ist ein Paar  $(\mathcal{CA}, T)$  bestehend aus

dem zellulären Automaten  $\mathcal{CA} = (R, N, \delta_\chi)$  und der Simulationszeit  $T \in \mathbb{N}$ .

**Definition 2.23 (Raumzeitliches Muster)** Sei  $(\mathcal{CA}, T)$  eine Simulation von  $\mathcal{CA}$  mit der Zellmenge  $C = \{c_0, c_1, \dots, c_n\}$ . Das aus der Simulation resultierende raumzeitliche Muster  $D$  ist eine Matrix von Zuständen in  $S$ . Die Matrix hat folgende  $(n + 1) \times (T + 1)$  Elemente:

$$D = \begin{pmatrix} \chi_0(c_0) & \chi_0(c_1) & \cdots & \chi_0(c_n) \\ \chi_1(c_0) & \chi_1(c_1) & \cdots & \chi_1(c_n) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_T(c_0) & \chi_T(c_1) & \cdots & \chi_T(c_n) \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_T \end{pmatrix} = ( z_{c_0} \quad z_{c_1} \quad \cdots \quad z_{c_n} )$$

Die Zeilen von  $D$  sind die bereits eingeführten Konfigurationen  $y_t$  des Automaten. Sie bilden die räumlichen Muster der Simulation. Die Spalten geben die zeitlichen Muster  $z_c$  des Automaten wieder.

Das raumzeitliche Muster der Simulation kann visualisiert werden: Jedes Element der Matrix  $D$  wird in einen Bildpunkt umgesetzt. Die Farbe des Bildpunktes gibt dabei den Zustand an. Bei binären zellulären Automaten wird schwarz für den Zustand 1 und weiß für den Zustand 0 verwendet.

**Beispiel 2.24** Abbildung 8 zeigt das raumzeitliche Muster einer Simulation über 200 Zeitschritte. Der verwendete zelluläre Automat verfügt über 200 Zellen.

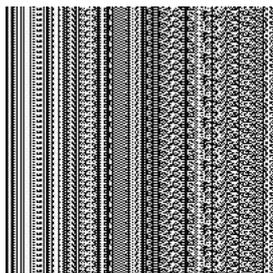


Abbildung 8: Beispiel eines raumzeitlichen Musters

### 2.3.2 Entropie

Ein Ziel bei der Analyse von zellulären Automaten ist es, den Informationstransport im Netzwerk auszuwerten. Zu diesem Zweck definierte CLAUDE ELWOOD SHANNON in seinem berühmten Artikel „A mathematical theory of communication“ [Sha48] die *Entropie*. Der Begriff stammt von dem griechischen Wort *εντροπία* (Wendung, Umwandlung) und ist der Thermodynamik entliehen, wo die Entropie ein Maß für die Unbestimmtheit der mikroskopischen Struktur eines Systems ist [Wik06].

**Definition 2.25 (Entropie)** Sei  $p_i$  eine Wahrscheinlichkeit für den Zustand  $i \in \Sigma$ . Die Entropie  $\mathcal{H}$  ist definiert als

$$\mathcal{H} = - \sum_{i \in \Sigma} p_i \log(p_i).$$

Diese allgemeine Definition wird für unsere Zwecke in zwei Varianten eingeführt. Dabei wird jeweils der zelluläre Automat  $\mathcal{CA} = (R, N, \delta_\chi)$  betrachtet.

**Definition 2.26 (Shannon-Entropie)** Die Shannon-Entropie  $\mathcal{S}$  wertet die Zustände einer Zelle  $c \in C$  im zeitlichen Muster des zellulären Automaten aus. Sie ist definiert als:

$$\mathcal{S}_c = - \sum_{s \in S} p_s \log_2(p_s) \quad \text{und} \quad \mathcal{S} = \sum_{c \in C} \frac{\mathcal{S}_c}{|C|}.$$

Dabei ist  $p_s$  die Wahrscheinlichkeit, dass sich die Zelle  $c$  im Zustand  $s \in S$  befindet.

**Definition 2.27 (Word-Entropie)** Die Word-Entropie  $\mathcal{W}$  betrachtet Blöcke einer bestimmten Länge innerhalb der zeitlichen Muster  $z_c, c \in C$ . Es gilt:

$$\mathcal{W}_c = - \sum_{\ell=1}^{T+1} p_\ell \log_2(p_\ell) \quad \text{und} \quad \mathcal{W} = \sum_{c \in C} \frac{\mathcal{W}_c}{|C|}.$$

Dabei ist  $p_\ell$  die Wahrscheinlichkeit für ein Wort der Länge  $\ell$  in  $z_c$ . Die maximale Wortlänge beträgt dabei sicherlich  $T + 1$ , da das raumzeitliche Muster nur  $T + 1$  Zeilen enthält.

Für obige Entropien wird insbesondere  $\log(0) = 0$  definiert, um die Entropie für alle Wahrscheinlichkeiten auswerten zu können. Die Entropie erreicht ihr Minimum falls ein sicheres Ereignis ( $p_i = 1$ ) beteiligt ist. Das Maximum tritt dagegen bei einer Gleichverteilung der Wahrscheinlichkeiten auf. Als Approximation der Wahrscheinlichkeiten  $p_i$  wird in dieser Arbeit die jeweilige relative Häufigkeit verwendet.

**Beispiel 2.28** Betrachtet wird das raumzeitliche Muster  $D$ , wie in Abbildung 9 angegeben. Der verwendete zelluläre Automat hat die Zellmenge  $C = \{c_0, c_1, c_2, c_3, c_4\}$ , die Simulationszeit beträgt  $T = 11$ .

$D =$	(	0	1	1	1	0						
		0	0	1	1	0						
		1	0	1	1	0						
		1	1	0	1	1						
		0	1	1	1	1						
		0	0	1	1	1						
		1	0	1	1	1						
		1	1	0	1	0						
		0	1	1	1	1						
		0	0	1	1	1						
		1	0	1	1	0						
		1	1	0	1	1						

$c$	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$	
$p_0$	$6/12$	$6/12$	$3/12$	$0/12$	$5/12$	
$p_1$	$6/12$	$6/12$	$9/12$	$12/12$	$7/12$	
$\mathcal{S}_c$	1.000	1.000	0.811	0.000	0.979	$\mathcal{S} = 3.791/5 = 0.758$

$c$	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$	
$p_1$	$0/6$	$2/7$	$3/6$	$0/1$	$3/6$	
$p_2$	$6/6$	$5/7$	$0/6$	$0/1$	$1/6$	
$p_3$	$0/6$	$0/7$	$3/6$	$0/1$	$1/6$	
$p_4$	$0/6$	$0/7$	$0/6$	$0/1$	$1/6$	
$p_{12}$	$0/6$	$0/7$	$0/6$	$1/1$	$0/6$	
$\mathcal{W}_c$	0.000	0.863	1.000	0.000	1.792	$\mathcal{W} = 3.655/5 = 0.731$

Abbildung 9: Beispiel eines raumzeitlichen Musters

Allgemeine Informationen zur Entropie finden sich in [Sha48] und [Hüt01], die hier eingeführten Varianten in [Mar05].

### 2.3.3 Klassifikation

Da die Anzahl der möglichen Transitionsfunktionen je nach Nachbarschaftsgrad beliebig groß werden kann (bei einem binären zellulären Graphenautomaten über einem 4-regulären Kreisgraphen bereits  $2^{32}$  Funktionen), erscheint eine Klassifikation der entstehenden Dynamik sinnvoll. STEPHEN WOLFRAM [Wol84, Wol86] gelang es Ende der sechziger Jahre vier Klassen zu definieren, die eine solche Einteilung der raumzeitlichen Muster zulässt. Diese Klassen sind:

**Wolfram-Klasse I:** In der ersten Klasse befinden sich alle Muster, die sich zu einem unveränderlichen stationären Fixpunkt entwickeln. Dadurch ergeben sich sehr monotone Gebilde, wie sie in Abbildung 10 (a) zu sehen sind.

**Wolfram-Klasse II:** Die zweite Wolfram-Klasse enthält alle Muster, die (lokal) periodische oder oszillierende Strukturen enthalten. Diese Strukturen sind vorhersagbar. In Abbildung 10 (b) beispielsweise endet das zeitliche Muster jeder Zelle in einer Periode der Länge 2, wodurch die nächsten Schritte vorherbestimmt sind.

**Wolfram-Klasse III:** Muster der dritten Wolfram-Klasse sind ungeordnet/chaotisch. Das System ist hier instabil und wechselt kontinuierlich seinen Zustand, wodurch eine Vorhersage der nächsten Konfigurationen unmöglich wird. Abbildung 10 (c) zeigt ein solches chaotisches Muster.

**Wolfram-Klasse IV:** In der vierten Klasse bilden die Muster komplexe Strukturen, die sich durch Zeit und Raum fortbewegen. Diese zeigen zwar eindeutig Organisationsstrukturen, sind jedoch ebenso wenig wie die chaotischen Muster vorhersagbar. Ein Beispiel eines Klasse IV-Musters findet sich in Abbildung 10 (d).

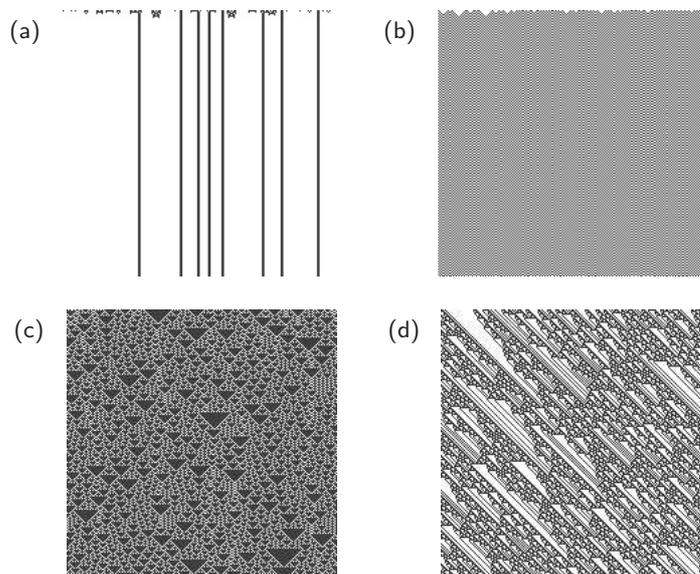


Abbildung 10: Beispiele für die Klassifikation eines raumzeitlichen Musters

### 3 Netzwerkmotive

#### 3.1 Definition

Zur Definition eines Netzwerkmotivs in einem Graphen werden einige weitere Definitionen aus der allgemeinen Graphentheorie benötigt, die nachfolgend beschrieben werden. Die Definitionen beziehen sich größtenteils nur auf gerichtete Graphen, können aber (mit der Vorschrift aus Bemerkung 2.6) auch analog für ungerichtete Graphen definiert werden.

**Definition 3.1 (Subgraph)** Sei  $G = (V, E)$  ein Graph. Ein Graph  $G' = (V', E')$  heißt (induzierter) Subgraph von  $G$ , falls gilt:

$$V' \subseteq V \text{ und } E' = \{(u, v) \in E \mid u \in V' \wedge v \in V'\}.$$

Häufig wird in diesem Fall auch einfach  $G' \subseteq G$  geschrieben.

**Definition 3.2 ( $k$ -Subgraph)** Sei  $G = (V, E)$  wieder ein Graph. Ein  $k$ -Subgraph  $G' = (V', E')$  von  $G$  ist ein Subgraph mit  $k$  Knoten. Es gilt also:

$$G' \subseteq G \text{ und } |V'| = k.$$

**Definition 3.3 (Graphen-Isomorphismus)** Zwei Graphen  $G_1 = (V_1, E_1)$  und  $G_2 = (V_2, E_2)$  heißen isomorph, falls es eine bijektive Abbildung  $\phi: V_1 \rightarrow V_2$  gibt, für die gilt:

$$\{\phi(v) \mid v \in V_1\} = V_2 \text{ und } \{(\phi(u), \phi(v)) \mid (u, v) \in E_1\} = E_2.$$

Eine solche Abbildung  $\phi$  heißt Graphen-Isomorphismus. Man schreibt auch  $G_1 \cong G_2$ , um die Existenz eines Graphen-Isomorphismus zu kennzeichnen.

**Beispiel 3.4** Abbildung 11 zeigt zwei Subgraphen des in Abbildung 1 eingeführten Graphen. Beide Subgraphen sind 3-Subgraphen, da sie jeweils 3 Knoten enthalten. Außerdem gibt es einen Graphen-Isomorphismus  $\phi$ , der wie folgt definiert ist:

$v$	1	2	3
$\phi(v)$	7	6	5

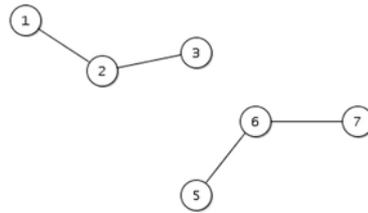


Abbildung 11: Beispiel für Subgraphen und Graphenisomorphie

**Definition 3.5 (Erreichbarkeit)** Sei  $G = (V, E)$  und  $u, v \in V$  zwei Knoten aus  $G$ . Der Knoten  $v$  heißt erreichbar von  $u$ , falls es einen ununterbrochenen Kantenzug von  $u$  nach  $v$  gibt, falls also gilt:

$$\exists w_i \in V, 0 \leq i \leq n : (u, w_1) \in E \wedge (w_1, w_2) \in E \wedge \dots \wedge (w_n, v) \in E.$$

**Definition 3.6 (Zusammenhang)** Ein ungerichteter Graph  $G = (V, E)$  heißt zusammenhängend, falls von jedem Knoten  $u \in V$  alle anderen Knoten  $v \in V, u \neq v$  erreichbar sind. Ein gerichteter Graph  $G' = (V', E')$  heißt (schwach) zusammenhängend, falls der zugehörige ungerichtete Graph

$$\tilde{G}' = (V', \{\{u, v\} \in V' \times V' \mid (u, v) \in E'\})$$

zusammenhängend ist.

**Beispiel 3.7** Fasst man die beiden Komponenten in Abbildung 11 als einen Graphen auf, so ist Knoten 3 von Knoten 1 aus erreichbar, Knoten 6 allerdings nicht. Der Graph ist daher nicht zusammenhängend.

Betrachtet man den linken Subgraphen für sich, so kann von Knoten 1 aus Knoten 2 und 3 erreicht werden (und umgekehrt), außerdem kann Knoten 3 auch von Knoten 2 erreicht werden (und umgekehrt), der Subgraph ist daher zusammenhängend.

Nun kann der Motivbegriff definiert werden: Zunächst betrachtet man hier alle zusammenhängenden Subgraphen einer bestimmten Größe und ordnet diese einer bestimmten (Isomorphie-)Klasse zu.

**Definition 3.8 (Netzwerkmotiv)** Sei  $G = (V, E)$ . Die  $k$ -Motivmenge  $\mathcal{M}_k$  von  $G$  enthält alle zusammenhängenden  $k$ -Subgraphen von  $G$ . Folglich gilt:

$$\mathcal{M}_k = \{G' = (V', E') \mid G' \subseteq G \wedge |V'| = k \wedge G' \text{ zusammenhängend}\}$$

Die Elemente  $M$  von  $\mathcal{M}_k$  heißen  $k$ -Netzwerkmotive oder einfach Motive.

**Definition 3.9 (Motivrelation)** Sei  $\mathcal{M}_k$  die  $k$ -Motivmenge eines Graphen  $G$ . Die Relation  $\Theta$  heißt Motivrelation der Größe  $k$  und ist wie folgt definiert:

$$\Theta \subseteq \mathcal{M}_k \times \mathcal{M}_k, \quad (M_1, M_2) \in \Theta \Leftrightarrow M_1 \cong M_2$$

Aus der Definition von  $\Theta$  mittels des Graphenisomorphismus folgt direkt, dass  $\Theta$  eine Äquivalenzrelation auf der Motivmenge  $\mathcal{M}_k$  bildet. Die Äquivalenzklassen von  $\Theta$  werden als  $[M]_\Theta$ ,  $M \in \mathcal{M}_k$  geschrieben.

**Definition 3.10 (Motivklassen)** Die Äquivalenzklassen

$$\mathcal{K}_k = \mathcal{M}_k / \Theta$$

von  $\Theta$  heißen Motivklassen, ihre Anzahl wird als Motivklassenzahl

$$\varphi_k(G) = |\mathcal{K}_k|$$

bezeichnet.

**Definition 3.11 (Motivgehalt, Motivverteilung)** Die Anzahl der Motive

$$\text{amount}(K) = |K|$$

innerhalb einer Motivklasse  $K \in \mathcal{K}_k$  wird als Motivgehalt bezeichnet. Listet man den Motivgehalt für allen Klassen auf, spricht man von der Motivverteilung

$$\text{amount}(\mathcal{K}_k) = (\text{amount}(K_1), \text{amount}(K_2), \dots, \text{amount}(K_n)),$$

mit  $\mathcal{K}_k = \{K_1, K_2, \dots, K_n\}$ .

**Definition 3.12 (Motivkonzentration, Motivsignatur)** Sei  $n = |\mathcal{M}_k|$  die Gesamtzahl der Motive. Der Wert

$$\text{conctr}(K) = \frac{|K|}{n} = \frac{\text{amount}(K)}{n}$$

für  $K \in \mathcal{K}_k$  spiegelt die Auftrittshäufigkeit der Motivklasse  $K$  im Graphen  $G$  wieder und wird als Motivkonzentration bezeichnet. Die Motivsignatur

$$\text{conctr}(\mathcal{K}_k) = (\text{conctr}(K_1), \text{conctr}(K_2), \dots, \text{conctr}(K_n)),$$

mit  $\mathcal{K}_k = \{K_1, K_2, \dots, K_n\}$  ist eine Liste aller Motivkonzentrationen in  $G$ .

Will man den Motivgehalt zweier Graphen vergleichen, ergibt sich zunächst ein Problem, falls die Graphen über eine unterschiedliche Anzahl von Knoten verfügen. Die Motivsignatur schafft hier Abhilfe, da der Motivgehalt über der Motivanzahl normiert wird. Leider ist auch diese Vergleichsmethode in der Praxis meist ungeeignet, da die Graphen in der Regel eine unterschiedliche Vernetzungsstruktur und Gradverteilung haben. Es wird daher ein Verfahren benötigt, das besser auf die Unterschiede der zu vergleichenden Graphen eingeht.

Die Gruppe um URI ALON schlägt vor, den Graphen mit einer Reihe leicht variiertes Zufallsgraphen mit gleicher Gradverteilung zu vergleichen und dadurch den so genannten *Z-Score* [Aln04] zu bestimmen. Auf diese Weise kann festgestellt werden, welche Motivklassen signifikant erhöht oder vermindert auftreten.

**Definition 3.13 (Z-Score)** Sei  $G$  ein Graph mit Motivklassenmenge  $\mathcal{K}_k$ . Seien außerdem  $\tilde{G}_i$ ,  $1 \leq i \leq m$ , leicht variierte Zufallsgraphen, die (unter Beibehaltung der Gradverteilung) aus  $G$  entstanden sind. Der Z-Score der Motivklasse  $K \in \mathcal{K}_k$  ergibt sich als

$$\mathcal{Z}(K) = \frac{|K| - \langle |K| \rangle}{\text{std}(|K|)},$$

wobei  $\langle |K| \rangle$  den Mittelwert des Motivgehalts aller randomisierten Graphen  $\tilde{G}_i$  und  $\text{std}(|K|)$  die zugehörige Standardabweichung angibt.

Normiert man nun den Z-Score, verfügt man über eine geeignete Methode, um verschiedenste Netzwerke bezüglich ihrer Motive zu vergleichen. Die Auflistung aller normierten Z-Scores nennt man *Signifikanzprofil*:

**Definition 3.14 (Signifikanzprofil)** Sei  $G$  ein Graph mit Motivklassenmenge  $\mathcal{K}_k = \{K_1, K_2, \dots, K_n\}$ . Der normierte Z-Score  $\mathcal{Z}_0$  der Motivklasse  $K \in \mathcal{K}_k$  ergibt sich als:

$$\mathcal{Z}_0(K) = \frac{\mathcal{Z}(K)}{\sum_{K' \in \mathcal{K}_k} \mathcal{Z}(K')}.$$

Das Signifikanzprofil  $\text{SP}_k$  von  $G$  mit Motiven der Größe  $k$  ergibt sich aus den normierten Z-Scores:

$$\text{SP}_k = (\mathcal{Z}_0(K_1), \mathcal{Z}_0(K_2), \dots, \mathcal{Z}_0(K_n)).$$

## 3.2 Klassifikation von Motivverteilungen

### 3.2.1 Betrachtete Motivklassen

Es gibt  $2^{k \cdot k}$  verschiedene  $k$ -Subgraphen. Um die Zahl der möglichen Motivklassen zu bestimmen, müssen die nicht zusammenhängenden Subgraphen abgezogen werden. Außerdem dürfen zueinander isomorphe Subgraphen nur einfach gezählt werden. Die Zahl wächst dennoch sehr schnell, so dass die Auswertung nur für eine relativ kleine Motivgröße möglich wird. Die Motivklassenzahl selbst ist vom betrachteten Graphen unabhängig.

Im Folgenden werden in gerichteten Graphen ausschließlich Motive der Größe 3 (13 Klassen) und in ungerichteten Graphen Motive der Größe 4 (6 Klassen) betrachtet. Die Motivklassen werden zur einfacheren Beschreibung der Reihe nach durchnummeriert. Die Abbildungen 12 und 13 zeigen die genannten Motive mit der zugehörigen Bezeichnung und Motivklassennummer.

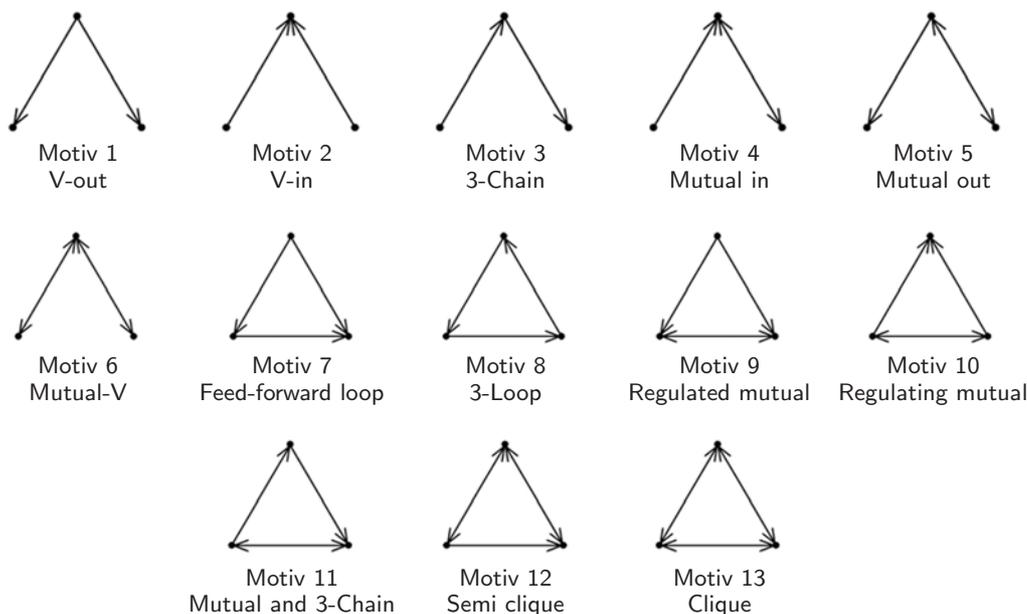


Abbildung 12: Motivklassen der Größe 3 in gerichteten Graphen

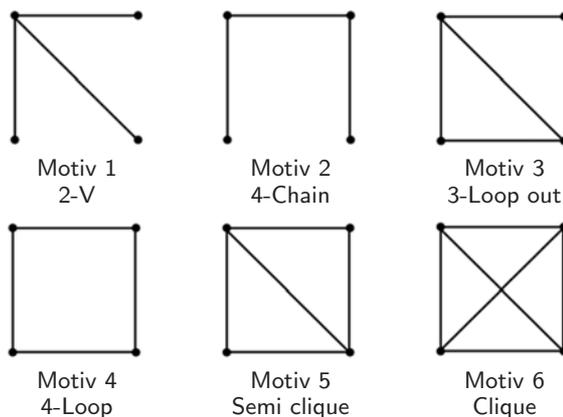


Abbildung 13: Motivklassen der Größe 4 in ungerichteten Graphen

### 3.2.2 Charakteristische Signifikanzprofile in Familien von Graphenklassen

Die Forschungsgruppe um URI ALON hat sich bereits intensiv mit Motiven in den verschiedensten Netzwerktypen auseinander gesetzt.

Eine der ersten Beobachtungen war, dass einzelne Motive in realen Netzwerken eine signifikante Erhöhung im Vergleich zu randomisierten Graphen zeigen [Aln02]. Diese motivspezifischen Erhöhungen sind nicht auf eine bestimmte Klasse von Netzwerken begrenzt: Es wurden charakteristische Signifikanzprofile in Netzwerken aus der Biochemie, der Neurobiologie, der Ökologie und dem Ingenieurwesen nachgewiesen. Ein Ziel der Forschungsgruppe war dabei, bestimmte Bausteine zu identifizieren, aus welchen ein Netzwerk eines bestimmten Typs besteht.

Zwei Jahre später veröffentlichte die Gruppe einen Artikel, in dem reale Netzwerke in verschiedene Klassen (so genannte *Superfamilies*) eingeordnet werden [Aln04]. Neben Netzwerken aus den Naturwissenschaften wurden hier auch soziologische und linguistische Netzwerke betrachtet. Die Untersuchungen lieferten eine verblüffende Ähnlichkeit der Signifikanzprofile innerhalb der einzelnen Familien, beispielsweise einen Korrelationskoeffizienten von über 0.99 in bestimmten biologischen Netzwerken. Die Einordnung von sozialen Netzen und Hyperlink-Netzwerken aus dem Word Wide Web in eine gemeinsame Superfamilie lässt auf den ähnlichen Aufbau der beiden Netzwerktypen schließen. Abbildung 14 zeigt solche Superfamilien in gerichteten Graphen. Betrachtete wurden Motive der Größe 3.

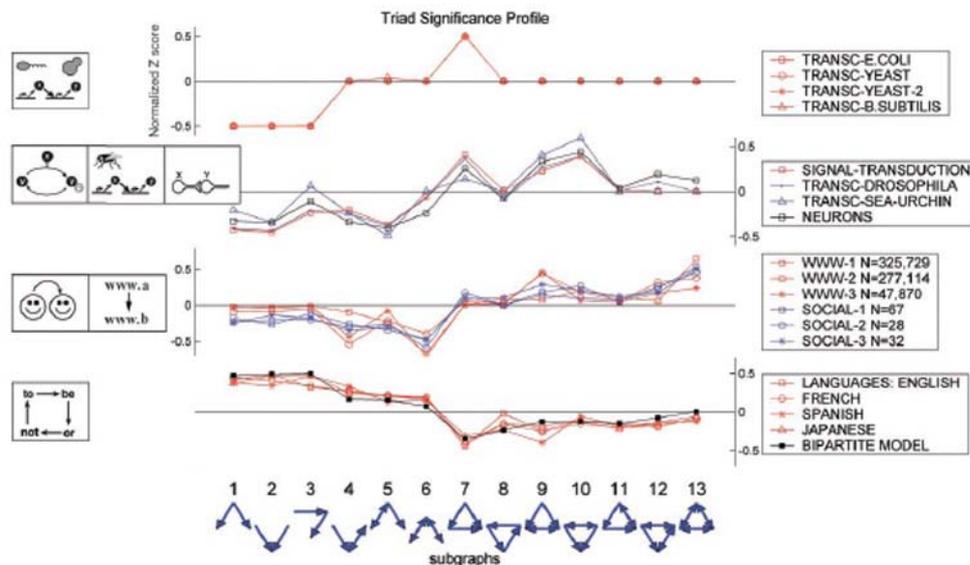


Abbildung 14: Superfamilien in gerichteten Graphen [Aln04]

Neben der großen Ähnlichkeit innerhalb einer Familie zeigte sich ein recht großer Unterschied zwischen zwei verschiedenen Familien. Während soziale Netzwerke beispielsweise eine Überhöhung der Motive 7 (Feed-forward loop) und 13 (Clique) enthalten, sind in linguistischen Netzen eher die Motive 1 (V-out), 2 (V-in) und 3 (3-chain) zu finden.

Die nachfolgenden Abschnitte beschäftigen sich nun mit den charakteristischen Motivverteilungen der in Kapitel 2.1.2 eingeführten Aufbauschemata. Es werden zunächst nur gerichtete Graphen betrachtet.

### 3.2.3 $2k$ -reguläre Kreisgraphen

Der strukturierte Aufbau des  $2k$ -regulären Kreisgraphen bewirkt, dass die Motivverteilung des entstehenden Graphen einer ganz bestimmten Regel folgt. Ausgehend von einem Kreis mit  $k = 1$ , der ausschließlich Motive der Klasse 3 enthält (3-Chain), entstehen durch die Verbindung der  $k$  nächsten Nachbarn für  $k > 1$  Motive der Klasse 7 (Feed-forward loop), welche eine Abkürzung auf dem äußeren Kreis bieten.

Die Motivsignatur ist unabhängig von der Anzahl der Knoten. Interessant ist hier die Beobachtung des Parameters  $k$ . Wird dieser inkrementiert, erhöht sich Motiv 3 um  $(k + 1) \cdot |V|$  und Motiv 7 um  $k \cdot |V|$ . Diese Formel erklärt auch die Tatsache, dass der Motivgehalt von Motiv 7 im  $2k$ -regulären Kreisgraphen genau dem Motivgehalt von Motiv 3 im  $2(k + 1)$ -regulären Kreisgraphen entspricht.

Der normierte Z-Score steigt mit wachsendem  $k$  für das Motiv 7 an und vermindert sich im Motiv 3. Das Signifikanzprofil enthält hier keine negativen Werte.

Tabelle 1 listet die Motivverteilung und das Signifikanzprofil für  $k = 1, \dots, 8$  auf, Abbildung 15 stellt die Werte graphisch dar.

Motivverteilung	1	2	3	4	5	6	7	8	9	10	11	12	13
$k = 1$	0	0	200	0	0	0	0	0	0	0	0	0	0
$k = 2$	0	0	600	0	0	0	200	0	0	0	0	0	0
$k = 3$	0	0	1200	0	0	0	600	0	0	0	0	0	0
$k = 4$	0	0	2000	0	0	0	1200	0	0	0	0	0	0
$k = 5$	0	0	3000	0	0	0	2000	0	0	0	0	0	0
$k = 6$	0	0	4200	0	0	0	3000	0	0	0	0	0	0
$k = 7$	0	0	5600	0	0	0	4200	0	0	0	0	0	0
$k = 8$	0	0	7200	0	0	0	5600	0	0	0	0	0	0
norm. Z-Scores	1	2	3	4	5	6	7	8	9	10	11	12	13
$k = 1$	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$k = 2$	0.000	0.000	0.152	0.000	0.000	0.000	0.848	0.000	0.000	0.000	0.000	0.000	0.000
$k = 3$	0.000	0.000	0.117	0.000	0.000	0.000	0.883	0.000	0.000	0.000	0.000	0.000	0.000
$k = 4$	0.000	0.000	0.105	0.000	0.000	0.000	0.895	0.000	0.000	0.000	0.000	0.000	0.000
$k = 5$	0.000	0.000	0.089	0.000	0.000	0.000	0.911	0.000	0.000	0.000	0.000	0.000	0.000
$k = 6$	0.000	0.000	0.079	0.000	0.000	0.000	0.921	0.000	0.000	0.000	0.000	0.000	0.000
$k = 7$	0.000	0.000	0.077	0.000	0.000	0.000	0.923	0.000	0.000	0.000	0.000	0.000	0.000
$k = 8$	0.000	0.000	0.071	0.000	0.000	0.000	0.929	0.000	0.000	0.000	0.000	0.000	0.000

Tabelle 1: Motivverteilung und normalisierte Z-Scores in gerichteten  $2k$ -regulären Kreisgraphen

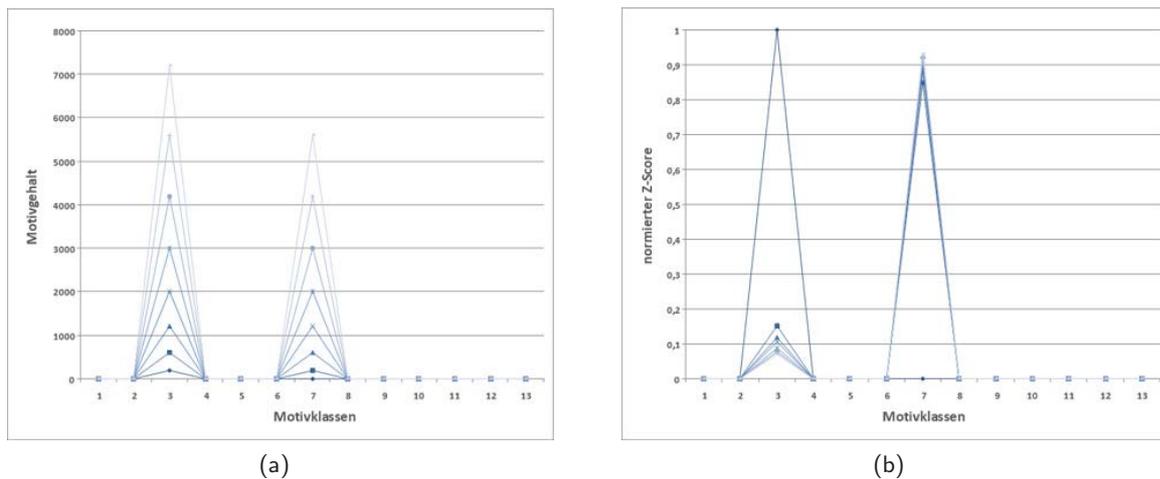


Abbildung 15: Motivverteilung und normierter Z-Score in gerichteten  $2k$ -regulären Kreisgraphen

### 3.2.4 Erdős-Rényi-Graphen

Um eine charakteristische Verteilung in Erdős-Rényi-Graphen zu finden, wurden Beispielgraphen mit verschiedenen Vernetzungswahrscheinlichkeiten  $p$  erzeugt und auf Ihren Motivgehalt untersucht. Die Abbildungen 16 und 17 zeigen die resultierenden Motivverteilungen und den normierten Z-Score, während Tabelle 2 die Durchschnittswerte der einzelnen Vernetzungswahrscheinlichkeiten angibt. Die Durchschnittswerte des normierten Z-Scores sind im Diagramm grün eingezeichnet. Die Graphen verfügen jeweils über 200 Knoten; zu den Vernetzungswahrscheinlichkeiten 0.5, 0.25, 0.2 und 0.18 wurden jeweils 50 Graphen erzeugt.

Prinzipiell können in Erdős-Rényi-Graphen alle 13 Motivklassen enthalten. Dabei sind natürlich Motive mit wenigen Kanten wahrscheinlicher als Motive mit vielen Kanten. Das Motiv 13 (Clique) verfügt mit sechs Kanten über die meisten Kanten, es taucht nur äußerst selten auf.

Interessant ist zunächst einmal, dass Motiv 3 (3-Chain) etwa doppelt so häufig vorkommt wie die beiden anderen Motive mit zwei Kanten, nämlich Motiv 1 (V-out) und Motiv 2 (V-in). Diese treten für sich genommen etwa gleich häufig auf.

Die Motive mit drei Kanten, also Motiv 4 (Mutual-in), 5 (Mutual-out), 7 (Feed-forward loop) und Motiv 8 (3-Loop) treten etwa um den Faktor 50 weniger auf als Motiv 3. Dabei ist Motiv 8 durchschnittlich deutlich weniger aufzufinden als die Übrigen. Die Ähnlichkeit zu Motiv 3 lässt vermuten, dass die Überhöhung von Motiv 3 zumindest teilweise auf Kosten des Motivs 8 entstanden sind. Motive mit mehr als drei Kanten treten entsprechend seltener auf und sind in Ihrer Anzahl im Verhältnis zu den Motiven 1-3 äußerst gering.

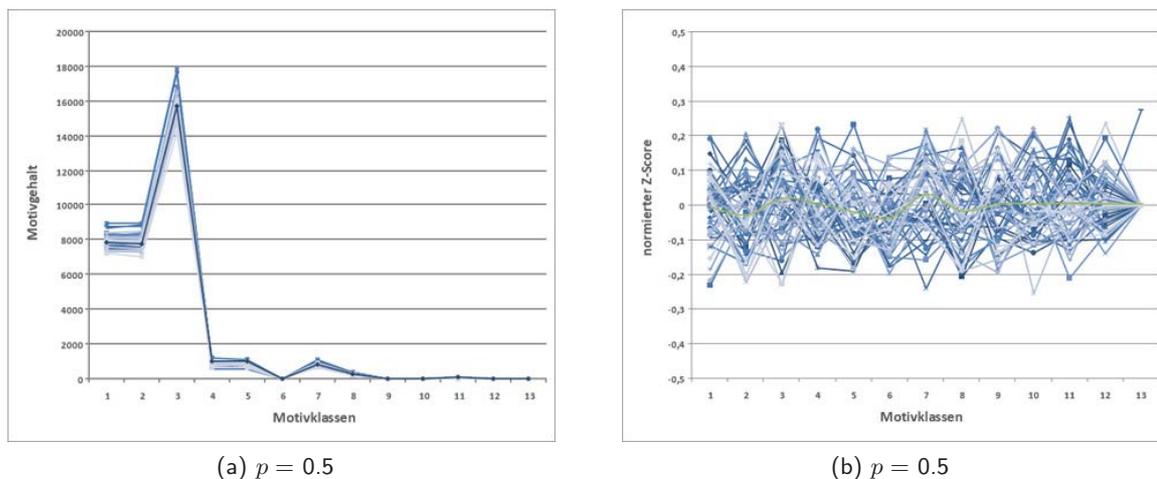


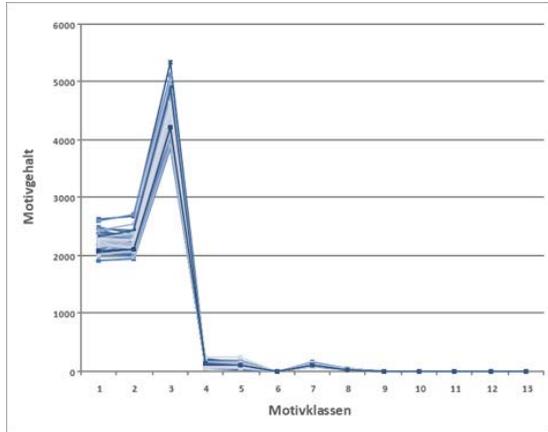
Abbildung 16: Motivverteilung und normierter Z-Score in gerichteten ER-Graphen (1)

Motivverteilung	1	2	3	4	5	6	7	8	9	10	11	12	13
$p = 0.5$	7961	7975	15931	839	841	22	845	274	22	23	44	2	0
$p = 0.25$	2226	2243	4459	121	118	2	115	39	2	1	3	0	0
$p = 0.2$	1446	1450	2905	60	61	1	60	19	1	1	1	0	0
$p = 0.18$	1185	1171	2359	48	48	1	43	15	1	0	1	0	0

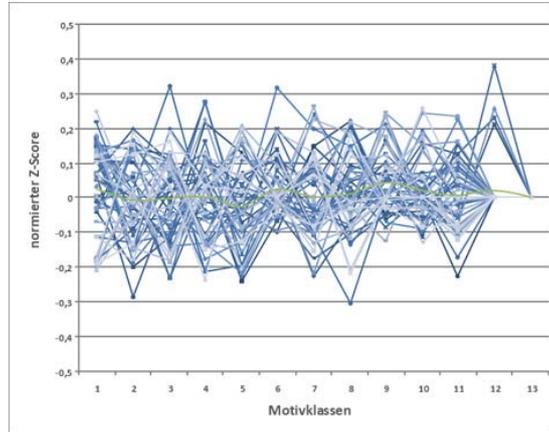
  

norm. Z-Scores	1	2	3	4	5	6	7	8	9	10	11	12	13
$p = 0.5$	-0.004	-0.028	0.021	0.006	-0.018	-0.037	0.031	-0.016	0.004	0.005	0.007	0.005	0.005
$p = 0.25$	0.024	-0.007	-0.003	0.003	-0.025	0.026	0.005	0.013	0.041	0.021	0.008	0.022	0.000
$p = 0.2$	-0.021	-0.006	-0.006	0.012	0.027	0.029	0.011	-0.006	0.061	0.039	0.040	0.000	0.000
$p = 0.18$	0.014	-0.008	-0.034	0.034	0.018	0.040	-0.012	0.019	0.066	0.047	0.037	0.000	0.000

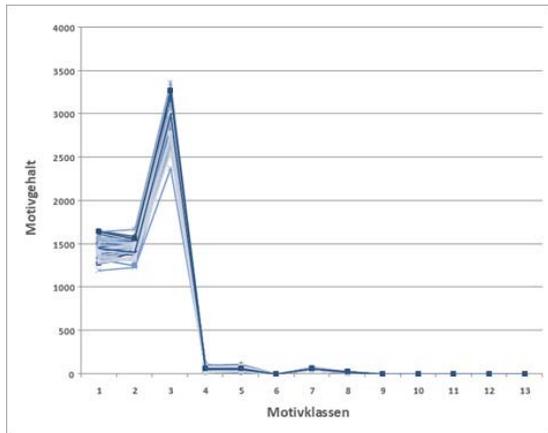
Tabelle 2: Durchschnittliche Motivverteilungen und normalisierte Z-Scores in gerichteten ER-Graphen



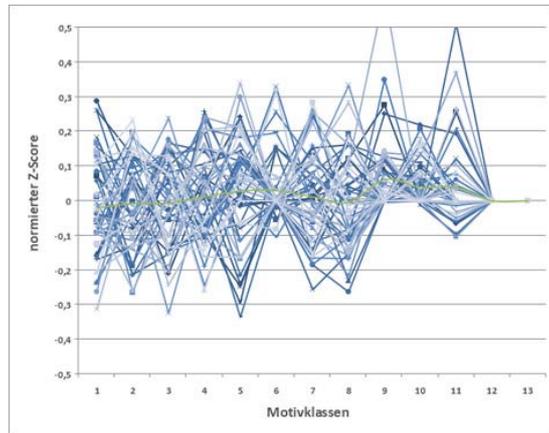
(a)  $p = 0.25$



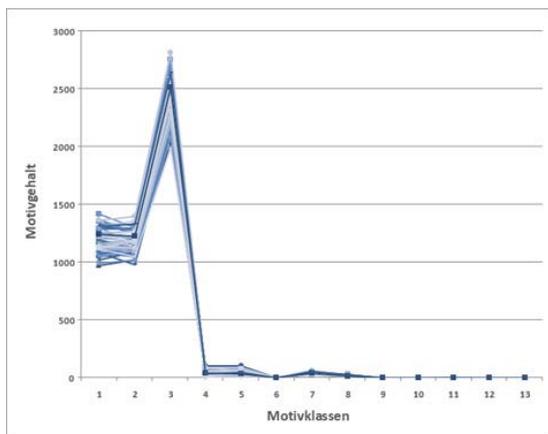
(b)  $p = 0.25$



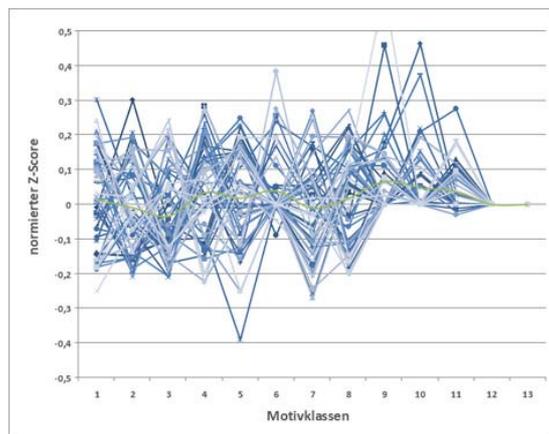
(a)  $p = 0.2$



(b)  $p = 0.2$



(c)  $p = 0.18$



(d)  $p = 0.18$

Abbildung 17: Motivverteilung und normierter Z-Score in gerichteten ER-Graphen (2)

Bei Änderung der Vernetzungswahrscheinlichkeit  $p$ , bleiben Motivverteilung und normierter Z-Score erhalten. Die Motivverteilung wird lediglich in Ihrer absoluten Anzahl erhöht (bedingt durch die größere Kantenzahl), die Charakteristik bleibt jedoch unverändert).

Der normierte Z-Score zeigt in allen Diagrammen zunächst ein chaotisches Verhalten, was scheinbar im Widerspruch zu den Erkenntnissen der ALON-Gruppe steht [Aln04]. Die Graphen folgen allerdings lediglich der gleichen Gradverteilung und stammen nicht aus einer bestimmten Problemdomäne, so dass Erdős-Rényi-Graphen eben keine der dort genannten Superfamilien bildet.

Um auszuschließen, dass es sich um eine zu große Variation beim Z-Score handelt, wurde dieser für einige ausgewählte Graphen mehrfach berechnet. Durch die stochastische Definition des Z-Scores sind diese natürlich unterschiedlich, bleiben aber dennoch sehr nahe zusammen, so dass die im Diagramm dargestellten Werte stabil sind. Für einen Beispielgraphen mit Vernetzungswahrscheinlichkeit  $p = 0.18$  sind die Ergebnisse von fünf Auswertungen des normierten Z-Scores nebst Mittelwert und Streuung in Tabelle 3 gelistet.

Die grünen Durchschnittskurven in Abbildung 16 zeigen durch ihre Nähe zur Nulllinie, dass kein Motiv überaus signifikant überhöht oder vermindert ist, was durchaus dem Aufbauschema von ER-Graphen entspricht.

Lauf	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-0.014	0.306	-0.060	-0.137	0.187	0.000	-0.232	0.065	0.000	0.000	0.000	0.000	0.000
2	0.000	0.313	-0.064	-0.114	0.181	0.000	-0.263	0.065	0.000	0.000	0.000	0.000	0.000
3	-0.011	0.304	-0.067	-0.111	0.205	0.000	-0.237	0.065	0.000	0.000	0.000	0.000	0.000
4	-0.015	0.294	-0.064	-0.114	0.199	0.000	-0.252	0.062	0.000	0.000	0.000	0.000	0.000
5	-0.002	0.305	-0.072	-0.118	0.185	0.000	-0.240	0.077	0.000	0.000	0.000	0.000	0.000
Mittelwert	-0.008	0.304	-0.065	-0.119	0.192	0.000	-0.245	0.067	0.000	0.000	0.000	0.000	0.000
Streuung	0.007	0.007	0.005	0.011	0.010	0.000	0.013	0.006	0.000	0.000	0.000	0.000	0.000

Tabelle 3: Verschiedene normierte Z-Scores des ersten Graphen mit  $p = 0.18$

### 3.2.5 Skalenfreie Graphen

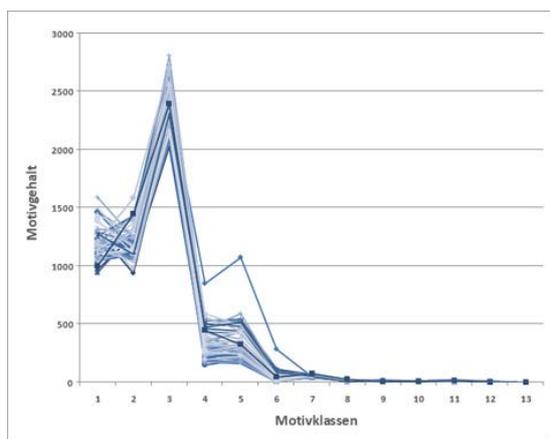
Bei skalenfreien Graphen ist der Motivgehalt von der Wahl der Verbindungswahrscheinlichkeit abhängig. Für den in Abschnitt 2.1.2 beschriebenen Ansatz

$$p = \frac{\deg(v')}{\sum_{v \in V} \deg(v)}$$

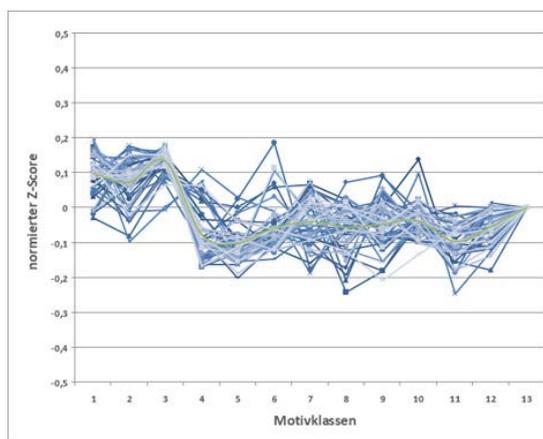
ergibt sich eine große Zahl von Motiven der Klasse 3. Abbildung 18 zeigt wieder die Motivverteilung und den normierten Z-Score von 50 skalenfreien Graphen mit einer unterschiedlichen Kantenzahl. Diese ist durch den Parameter  $m$ , also der Anzahl Kanten, die mit jedem Knoten eingefügt werden, bestimmt. Die Durchschnittswerte sind in Tabelle 4 gelistet und im Diagramm grün dargestellt.

Ebenso wie bei Erdős-Rényi-Graphen treten bei skalenfreien Graphen alle 13 Motivklassen auf. Die Motivklassen 1 (V-out) und 2 (V-in) kommen wieder etwa halb so oft wie Motiv 3 (3-Chain) vor. Die Klassen 4 (Mutual-in) und 5 (Mutual-out) treten dagegen weit häufiger auf als in Erdős-Rényi-Graphen. Bei Erhöhung der Knotenzahl oder des Vernetzungsgrades erhöht sich jeweils die Motivanzahl, bleibt jedoch in Ihrer Charakteristik gleich.

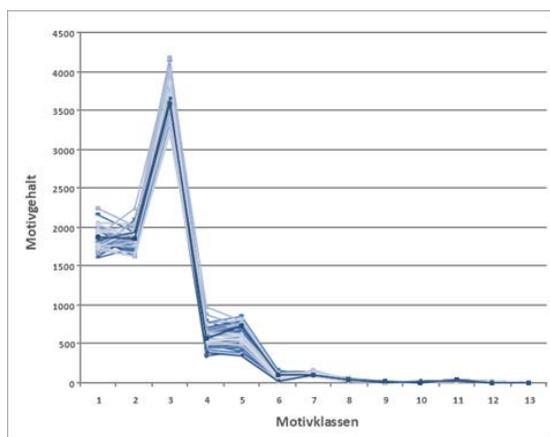
Die normierten Z-Scores zeigen eine Überhöhung der ersten drei Motive. Ab Klasse 4 wird der Z-Score negativ, die Motive treten also vermindert auf (im Gegensatz zu den leicht variierten Zufallsgraphen).



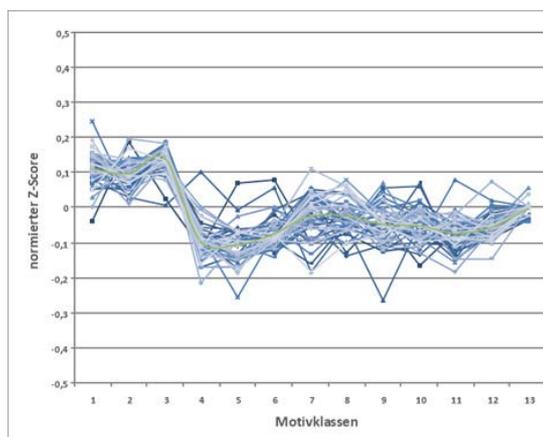
(a)  $m = 3$



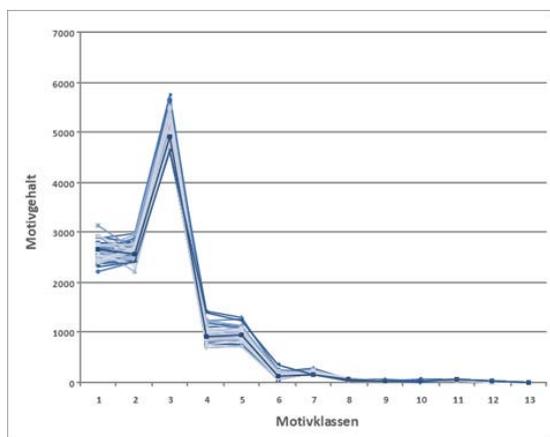
(b)  $m = 3$



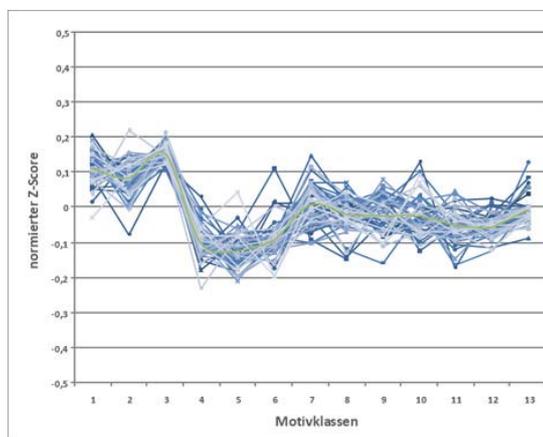
(c)  $m = 4$



(d)  $m = 4$



(e)  $m = 5$



(f)  $m = 5$

Abbildung 18: Motivverteilung und normierter Z-Score in gerichteten skalenfreien Netzwerken

Motivverteilung	1	2	3	4	5	6	7	8	9	10	11	12	13
$m = 3$	1201	1200	2343	338	341	42	60	16	6	7	10	3	0
$m = 4$	1845	1852	3622	590	597	74	119	36	14	14	25	8	0
$m = 5$	2607	2593	5101	966	966	139	209	62	29	30	53	20	2
norm. Z-Scores	1	2	3	4	5	6	7	8	9	10	11	12	13
$m = 3$	0.099	0.074	0.136	-0.077	-0.101	-0.058	-0.039	-0.056	-0.047	-0.038	-0.096	-0.052	0.000
$m = 4$	0.109	0.097	0.139	-0.096	-0.107	-0.073	-0.027	-0.024	-0.049	-0.052	-0.075	-0.057	-0.002
$m = 5$	0.109	0.086	0.154	-0.101	-0.122	-0.093	0.006	-0.021	-0.024	-0.025	-0.052	-0.056	-0.009

Tabelle 4: Durchschnittliche Motivverteilungen und normalisierte Z-Scores in gerichteten BA-Graphen

## 4 Variation von Netzwerkmotiven

### 4.1 Auswirkungen der Variation

Am biologischen Institut der Technischen Universität Darmstadt wird der Zusammenhang von zellulären Automaten und Graphen betrachtet. Der Artikel „*Topology regulates pattern formation capacity of binary cellular automata on graphs*“ [Mar05] von CARSTEN MARR und MARC-THORSTEN HÜTT verwendet zelluläre Graphenautomaten und wertet deren raumzeitliche Muster (Dynamik) aus. Die Graphen werden dann in ihrer Topologie variiert und erneut anhand ihrer Dynamik klassifiziert.

Zunächst wird die *Entropieebene* betrachtet, in der Shannon- und Wordentropie für je ein raumzeitliches Muster eingetragen werden. Hier wird der Zusammenhang zwischen Dynamik, Entropie und Wolfram-Klasse sichtbar: Abbildung 19 zeigt Muster der Wolfram-Klasse I (stationär) in der linken unteren Ecke (grün), Muster der Wolfram-Klasse III (chaotisch) in der rechten oberen Ecke (gelb), Muster der Klasse IV (komplex) halbkreisförmig vor den chaotischen Mustern (rot) und Muster der Wolfram-Klasse II (periodisch) im übrigen Teil der Ebene (blau).

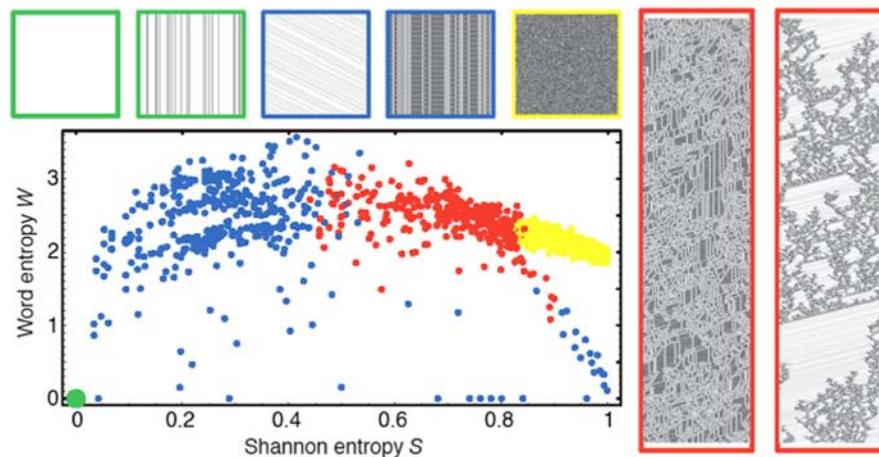


Abbildung 19: Verteilung in der Entropieebene [Mar05]

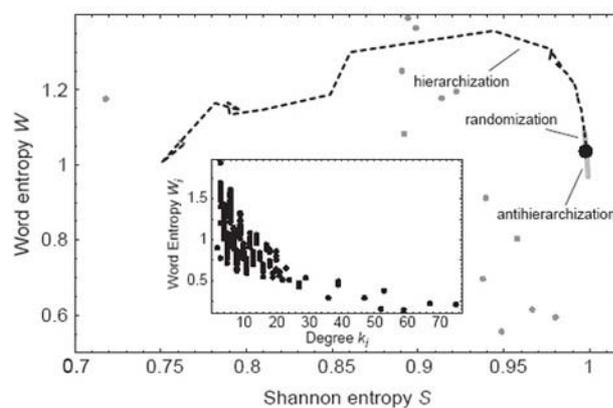


Abbildung 20: Trajektorie in der Entropieebene [Mar05]

Die Untersuchung ergab, dass die Musterbildungsfähigkeit stark von der Topologie abhängig ist. Diese Abhängigkeit erstreckt sich über unterschiedliche Graphentypen und Transitionsfunktionen hinweg. Auch sind die Änderungen nicht auf eine bestimmte topologische Eigenschaft beschränkt. Die Änderung von topologischen Parametern ergab insbesondere eine Trajektorie innerhalb der Entropieebene, was eine

Abhängigkeit von Topologie und Musterbildungsfähigkeit vermuten lässt. Abbildung 20 zeigt eine solche Trajektorie.

## 4.2 Ansatz und Probleme

Um die Auswirkungen von Motiven beobachten zu können, werden Graphen benötigt, die einer bestimmten Motivsignatur folgen. Sicherlich ist es unpraktisch diese per Hand zu erzeugen, so dass geeignete Algorithmen gefunden werden müssen, um diese Aufgabe zu erleichtern.

Man kann solche Algorithmen in zwei Gruppen einteilen:

- **Konstruktive Algorithmen** erzeugen einen neuen Graphen, der einer bestimmten Motivsignatur folgt. Man kann dies als eine Art Baukastensystem sehen, in dem die Motive in einer bestimmten Anzahl vorgegeben werden und der Algorithmus versucht, diese zu kombinieren und damit einen Graphen mit exakt dieser Motivsignatur zu schaffen.
- **Variative Algorithmen** beginnen mit einem Graphen und verändern diesen sukzessive. Dabei wird versucht einzelne Motive in andere zu transformieren, um mehr Motive einer bestimmten Klasse zu erhalten bzw. die Motive einer Klasse zu reduzieren. Der variative Ansatz erlaubt die Analyse zweier Graphen, die zwar eine verschiedene Motivsignatur haben, sich ansonsten aber noch recht ähnlich sind.

Die vorliegende Arbeit beschränkt sich im Folgenden auf variative Algorithmen.

Das Herstellen einer bestimmten Motivsignatur stellt sich in der Praxis leider als schweres Problem heraus. Da die betrachteten Subgraphen der Motive ineinander verwoben sind, wird durch die Erzeugung eines bestimmten Netzwerkmotivs auch dessen Umgebung verändert.

**Beispiel 4.1** Wird das blau hervorgehobene Motiv 1 in Abbildung 21 verändert, schlägt sich die Änderung auch in den Motiven 2-5 nieder. Diese Änderungsfortpflanzung bewirkt, dass sich die Motivsignatur des variierten Graphen nur schwer kontrollieren lässt.

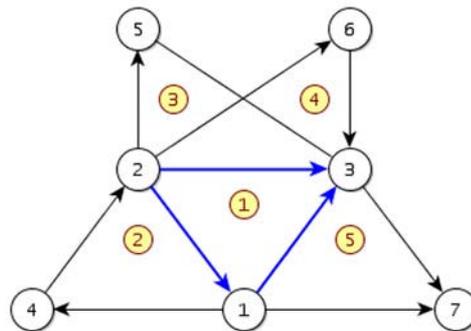


Abbildung 21: Nebenwirkungen bei der Motivvariation

Darüber hinaus werden durch wahlloses Einfügen und Entfernen von Kanten die Topologieeigenschaften des Graphen verändert. Um die entstehende Motivsignatur besser analysieren und vergleichen zu können, sollen aber gerade diese topologischen Größen bei der Variation beachtet werden. Die wichtigsten dieser Kenngrößen sind:

- die Gradsequenz,
- der Clustering-Koeffizient und
- der topologische Overlap.

Zunächst werden diese Kenngrößen definiert:

**Definition 4.2 (Gradsequenz)** Sei  $G = (V, E)$  ein Graph mit  $V = \{v_1, v_2, \dots, v_n\}$ . Eine Folge

$$d_V \in \mathbb{N}^n, \quad d_V = (\deg(v_1), \deg(v_2), \dots, \deg(v_n))$$

heißt Gradsequenz des Graphen  $G$ . Es handelt sich um eine Liste aller Knotengrade in  $G$ .

**Definition 4.3 (Clustering-Koeffizient)** Sei  $G = (V, E)$  ein Graph und  $v \in V$  ein Knoten in  $G$ . Der (lokale) Clustering-Koeffizient misst den Vernetzungsgrad des Knotens  $v$ . Dazu wird die Menge  $L_v$  aller Kanten betrachtet, deren beider Knoten mit  $v$  benachbart sind. Diese wird normiert über der Anzahl möglicher Kanten. Der Clustering-Koeffizient ist also eine Funktion

$$c: V \longrightarrow \mathbb{R}, \quad v \mapsto \frac{2|L_v|}{\deg(v)(\deg(v) - 1)}$$

mit der Menge der untereinander verbundenen Nachbarn

$$L_v = \{(u, w) \in E \mid u \in N(v) \wedge w \in N(v), u \neq v\}.$$

Aufbauend auf den lokalen Clustering-Koeffizienten aller Knoten lässt sich der globale Clustering-Koeffizient zum Graphen  $G$  als Mittelwert aller lokalen Clustering-Koeffizienten angeben.

Die Definition des topologischen Overlaps folgt der in *Hierarchical Organization of Modularity in Metabolic Networks* [Rav02] eingeführten:

**Definition 4.4 (Topologischer Overlap)** Sei  $G = (V, E)$  ein Graph und  $(u, v) \in E$  eine Kante in  $G$ . Der topologische Overlap misst den Vernetzungsgrad dieser Kante, indem die Zahl der gemeinsamen Nachbarn der Kantenendpunkte (plus die betrachtete Kante selbst) durch das Minimum der Knotengrade geteilt wird. Damit ergibt sich die Funktion

$$o: E \longrightarrow \mathbb{R}, \quad (u, v) \mapsto \frac{|N(u) \cap N(v)| + 1}{\min\{\deg(u), \deg(v)\}}.$$

**Beispiel 4.5** Abbildung 22 zeigt einige Beispiele der obigen Definitionen in einem Beispielgraphen.

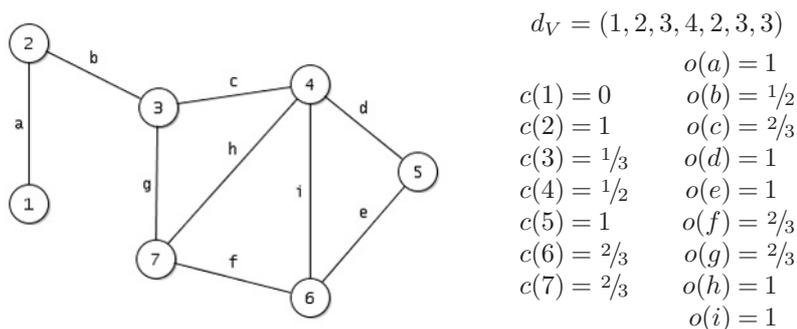


Abbildung 22: Topologische Kenngrößen

Die Gradsequenz ist dabei die wichtigste Kenngröße, um die variierten Graphen miteinander vergleichen zu können. Ein Variationsalgorithmus sollte also in jedem Fall die Gradsequenz beibehalten. Wird eine Kante in einen Graphen eingefügt, werden die beiden Endknoten in ihrem Grad inkrementiert; es muss daher ein Ausgleich vorgenommen werden, welcher beide Knotengrade wieder vermindert. Um dies zu erreichen wird jeweils eine Kante der beiden Knoten entfernt. Auch dies bewirkt wieder eine Änderung in der Gradsequenz, nämlich eine Reduktion in den beiden Endpunkten, welche durch eine Verbindung der beiden ausgeglichen wird.

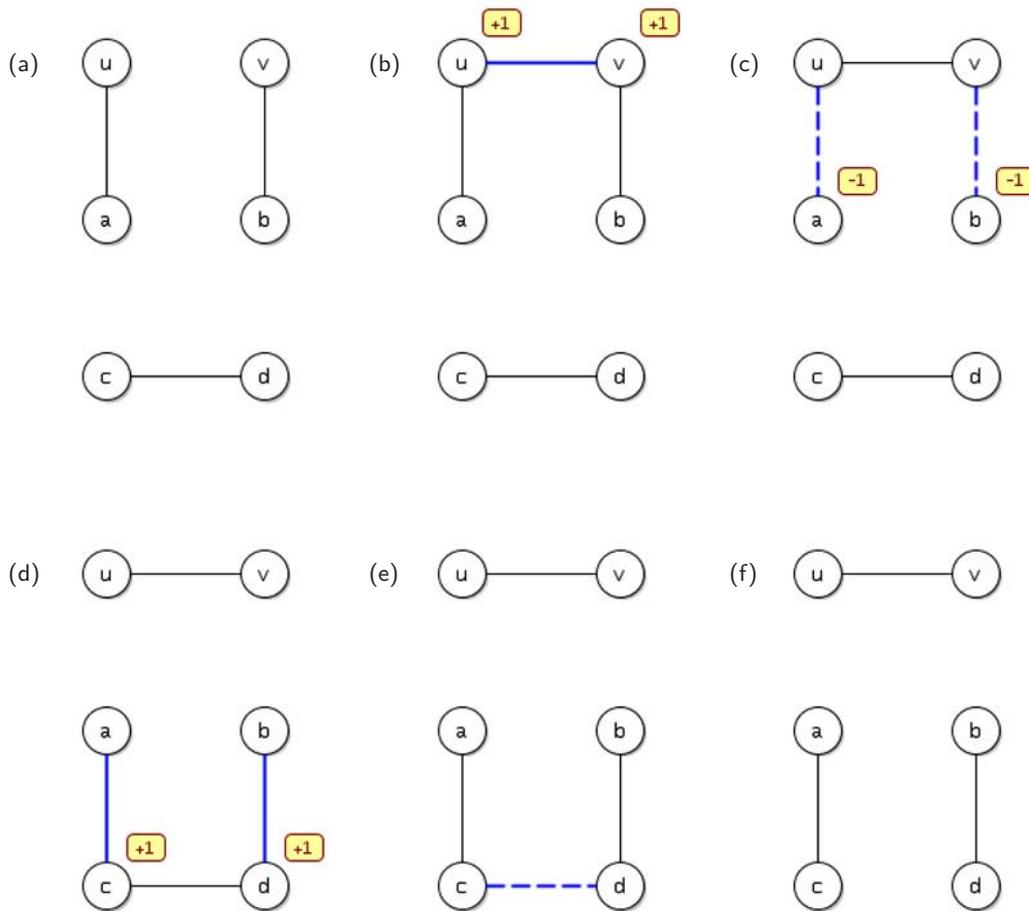


Abbildung 23: Variationsalgorithmus zur Beibehaltung der Gradverteilung

Das Verfahren lässt sich auf eine beliebige Anzahl Knoten ausweiten: In einer Knotenmenge  $v_1, v_2, \dots, v_n$ ,  $n \equiv 0 \pmod 2$  mit  $(v_1, v_2), (v_3, v_4), \dots, (v_{n-1}, v_n) \in E$  und  $(v_2, v_3), (v_4, v_5), \dots, (v_n, v_1) \notin E$  werden die Kanten vertauscht, das heißt es werden die Kanten  $(v_1, v_2), (v_3, v_4), \dots, (v_{n-1}, v_n)$  entnommen und die Kanten  $(v_2, v_3), (v_4, v_5), \dots, (v_n, v_1)$  eingefügt. Die im Rahmen dieser Arbeit verwendeten Algorithmen tauschen die Kanten auf einem solchen zyklischen Kreis mit  $n = 4$  Knoten.

**Beispiel 4.6** *Abbildung 23 zeigt diesen Algorithmus für  $n = 6$  Knoten. Ausgehend von (a) wird die Kante  $(u, v)$  eingefügt (b), was eine Löschung der Kanten  $(u, a)$  und  $(v, b)$  verlangt (c). Anschließend werden zum Ausgleich die Kanten  $(a, c)$  und  $(b, d)$  eingefügt (d). Die Gradverteilung wird ausgeglichen, indem  $(c, d)$  entnommen wird (e). Der entstandene Graph (f) besitzt damit gerade die Kanten, die im Zyklus  $(u, a, c, d, b, v)$  nicht enthalten waren.*

Auch der topologische Overlap und der Clustering-Koeffizient können in den Variationsalgorithmus einfließen. Hier ist etwa eine Variation möglich, die versucht möglichst große Clustering-Koeffizienten oder topologische Overlaps herzustellen. Die beiden Größen werden in dieser Arbeit nicht weiter betrachtet.

Die beiden folgenden Abschnitte beschreiben zwei Variationsalgorithmen, die in der Software MOTIF-CONTROL [EM06] implementiert wurden und für die späteren Untersuchungen verwendet werden.

### 4.3 Heuristischer Ansatz

Die heuristische Variation implementiert einen lokalen Suchalgorithmus, welcher ausgehend vom aktuellen Graphen einen variierten Graphen sucht, dessen Motivgehalt besser den gewünschten Kriterien entspricht.

Der Ablauf der Variation gliedert sich in *Runden* und *Schritte*. Ein Schritt ist dabei eine elementare Variation des Graphen, wie sie in Abschnitt 4.2 erläutert und in Abbildung 23 verdeutlicht wird. Es werden ausschließlich Vertauschungen auf einem Kreis mit vier Knoten durchgeführt ( $n = 4$ ).

Nun wird in jeder Runde eine gewisse Anzahl dieser Schritte durchgeführt und der Motivgehalt des Graphen neu berechnet. Ausgehend von den Motivklassen  $\mathcal{K}_k$  ergeben sich nach der Variation die Klassen  $\mathcal{K}'_k$ . Eine Ratingfunktion  $r$  bewertet nun die Variation der einzelnen Motivklassen. Die Funktion ist wie folgt definiert:

$$r: \mathcal{K}_k \times \mathcal{K}'_k \longrightarrow \mathbb{R}, \quad (K, K') \mapsto c_K (\text{amount}(K) - \text{amount}(K')) \quad \text{mit} \quad c_K \in \mathbb{R}.$$

Das Gesamtrating  $R$  der Variation ergibt sich als:

$$R = \sum_{\substack{K \in \mathcal{K}_k \\ K' \in \mathcal{K}'_k \\ K \cong K'}} r(K, K')$$

Die Koeffizienten  $c_K$  ermöglichen die motivspezifische Bewertung. Soll beispielsweise Motiv 3 erhöht werden, dabei aber die anderen Motive möglichst unberührt bleiben und Motiv 2 nur im Äußersten erhöht werden, empfiehlt sich die Koeffizientenwahl etwa wie im Beispiel in Tabelle 5 zu treffen.

$K$	1	2	3	4	5	6	
amount( $K$ )	826	747	834	300	0	198	
amount( $K'$ )	800	748	986	204	0	198	
amount( $K$ ) - amount( $K'$ )	-26	1	152	-96	0	0	$R = 1014$
$c_K$	-1	-20	6	-1	-1	-1	
$r(K, K')$	26	-20	912	96	0	0	

Tabelle 5: Beispiel einer Ratingberechnung

Die Akzeptanzschwelle des resultierenden Ratings kann ebenfalls eingestellt werden. Alle Ratings oberhalb dieser Schwelle werden akzeptiert. Wird ein Rating am Ende einer Variationsrunde akzeptiert, so wird der Ausgangsgraph durch den variierten ersetzt. Ist das Rating aber kleiner als die Akzeptanzschwelle, so werden keinerlei Änderungen am Graphen in die nächste Runde übernommen.

### 4.4 Motivtransformation

Anders als bei dem vorgestellten heuristischen Variationsalgorithmus werden Änderungen in dem folgenden Algorithmus nicht rückgängig gemacht. Grundidee ist hier das zufällige Auswählen eines Motives und dessen Transformation in ein bestimmtes Zielmotiv, das erhöht werden soll.

Betrachtet wird also das Motiv  $M \in \mathcal{M}_k$ , welches in das Zielmotiv  $\hat{M}$  transformiert werden soll. Sind  $M$  und  $\hat{M}$  äquivalent bezüglich  $\Theta$ , stellen sie also bereits das gleiche Motiv dar, wird ein anderes  $M$  gewählt. Zunächst muss dann eine isomorphe Form des Zielmotivs  $\hat{M}$  gefunden werden, die möglichst nahe am aktuell betrachteten Motiv  $M$  liegt. Als Metrik wird die Anzahl der Unterschiede in der Adjazenzmatrix der beiden Motive verwendet.

Anschließend wird eine Liste aller Änderungen erstellt, die zur Transformation notwendig sind. Eine Änderung kann das Einfügen oder das Löschen einer Kante sein. Auch im Transformationsalgorithmus soll die Gradsequenz erhalten bleiben, für jede Operation in der Änderungsliste muss also ein zyklischer Kantentausch, wie in Abschnitt 4.2 beschrieben, durchgeführt werden. Zu den beiden Knoten, die durch eine Kante verbunden werden sollen (oder zwischen denen eine Kante entfernt werden soll), werden folglich zwei weitere Knoten gesucht, die einen solchen alternierenden Zykel bilden.

Um diese beiden Knoten zu finden, wird zunächst die Adjazenzmatrix der beiden Knoten im Ausgangsmotiv  $M$  erzeugt. Nun wird diese Adjazenzmatrix um zwei zufällig gewählte Knoten  $v_1$  und  $v_2$  erweitert. Die entstehende Matrix muss eine von vier bestimmten Strukturen besitzen, damit ein Kantentausch möglich wird. Für das Einfügen einer Kante sind das

$$\begin{pmatrix} 0 & 0 & 1 & * \\ * & 0 & * & * \\ * & * & 0 & * \\ * & 1 & 0 & 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 & 0 & * & 1 \\ * & 0 & * & * \\ * & 1 & 0 & 0 \\ * & * & * & 0 \end{pmatrix}.$$

Vertauscht man die blau hervorgehobenen Einsen und Nullen, ergeben sich die beiden Strukturen für das Löschen einer Kante:

$$\begin{pmatrix} 0 & 1 & 0 & * \\ * & 0 & * & * \\ * & * & 0 & * \\ * & 0 & 1 & 0 \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} 0 & 1 & * & 0 \\ * & 0 & * & * \\ * & 0 & 0 & 1 \\ * & * & * & 0 \end{pmatrix}.$$

Ein  $*$  steht hier für einen beliebigen Wert in der Adjazenzmatrix. Der Test, ob eine Matrix über einen alternierenden Zykel verfügt – und damit den Kantentausch erlaubt – kann effizient durchgeführt werden, da lediglich die Positionen  $(1, 3), (4, 2), (4, 3)$  bzw.  $(1, 4), (3, 2), (3, 4)$  verglichen werden müssen. Die Position  $(1, 2)$  ergibt sich bereits durch die Operation, die durchgeführt werden soll (Kante einfügen respektive löschen), während die Nullen auf der Hauptdiagonalen durch die Forderung, keine Schleifen im Graphen zuzulassen, vorhanden sind.

Abbildung 24 zeigt ein Beispiel für eine Adjazenzmatrix, in der ein Kantentausch möglich ist. Die Matrix entspricht der linken Struktur der Operation „Kante einfügen“.

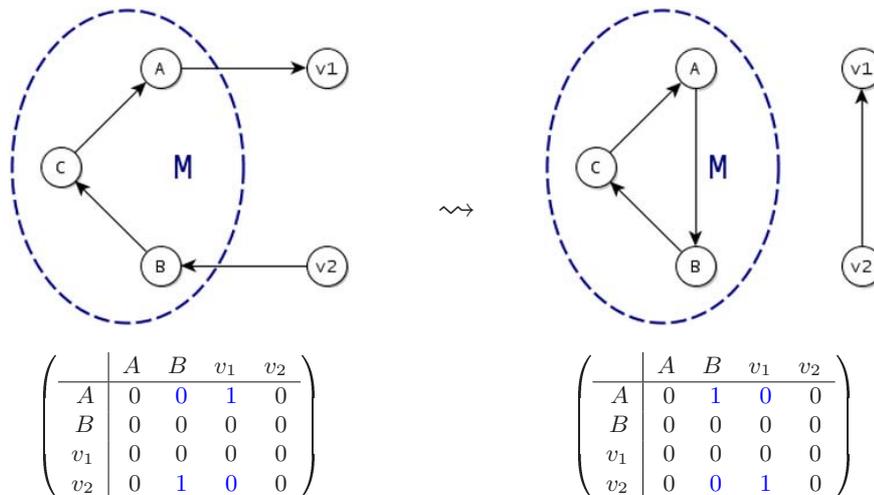


Abbildung 24: Variationsalgorithmus zur Beibehaltung der Gradverteilung

## 4.5 Ergebnisse

Prinzipiell zeigen beide Ansätze ein sehr ähnliches Verhalten: Motive mit vielen Kanten und Motive, die vorher schon in hoher Anzahl vorhanden sind, lassen sich sehr schlecht herstellen, während bei allen anderen Motiven signifikante Änderungen zu erwarten sind. Der heuristische Ansatz liefert dabei immer wesentlich schneller ein Ergebnis und ist für die Gesamtberechnung oft besser geeignet als die Motivtransformation.

Ein detaillierter Vergleich der beiden Variationsalgorithmen findet sich in [EM06].

## 5 Auswirkungen auf die Entropieebene

Nachdem die Motivverteilung der einzelnen Graphen analysiert ist und diese mit den beiden vorgestellten Variationsalgorithmen auch verändert werden kann, soll nun der Informationsgehalt der Netzwerke untersucht werden.

Dazu wird, wie in Abschnitt 2.3 beschrieben, die Simulation  $(\mathcal{GA}, T)$  des zellulären Graphenautomaten  $\mathcal{GA}$  durchgeführt. Die Shannon- und die Word-Entropie werden zur Bestimmung des Informationsgehalts verwendet.

Als Transitionsfunktion  $\delta_\chi$  bieten sich die in [Mar05] eingeführten  $\Omega^{(1)}$  und  $\Omega^{(2)}$  an, mit

$$\Omega_\chi^{(k)}(N(c), c) = \begin{cases} \chi(c) & \text{falls } \rho_k(N(c), c) \leq \kappa \\ 1 - \chi(c) & \text{falls } \rho_k(N(c), c) > \kappa \end{cases}, \quad k \in \{1, 2\}$$

wobei  $\rho_1$  über die Zustände der Nachbarn normiert und  $\rho_2$  darüber hinaus auch den eigenen Zustand der Zelle beachtet. Die beiden Hilfsfunktionen sind wie folgt definiert:

$$\rho_1(N(c), c) = \frac{1}{|N(c)|} \sum_{c' \in N(c)} \chi(c') \quad \text{bzw.} \quad \rho_2(N(c), c) = \frac{1}{|N(c)| + 1} (\chi(c) + \sum_{c' \in N(c)} \chi(c')).$$

Die Dynamik lässt sich über den Parameter  $\kappa$  regeln. Der Artikel von CARSTEN MARR beschreibt die Auswirkungen dieses Parameters und gibt günstige Werte für die einzelnen Graphen vor. Meist ist die Wahl zwischen 0.3 und 0.4 sinnvoll.

### 5.1 Testprotokolle

Für viele der nachfolgenden Untersuchungen werden Testprotokolle angegeben, welche die wichtigsten Eckdaten einer Simulation enthalten. Die Testprotokolle wurden von der Software MOTIFCONTROL erzeugt und für die Verwendung in dieser Arbeit lediglich noch geringfügig umformatiert. Der Aufbau eines Testprotokolls gliedert sich in drei Bereiche: Kopf, Diagramm und Datentabelle. Ein Beispiel findet sich in Abbildung 28.

Der Kopf beginnt mit dem Namen des Graphen bzw. der Graphenklasse sowie der Angabe der Knotenzahl. In dieser Arbeit werden zusätzlich die zur Erzeugung verwendeten Parameter, also etwa den Vernetzungsgrad  $p$  in Erdős-Rényi-Graphen oder die Anzahl der Kanten  $m$  die in skalenfreien Graphen in jedem Schritt eingefügt werden (siehe Abschnitt 2.1.2). Es folgt dann die Beschreibung des Variationsalgorithmus, der zwischen den einzelnen Simulationen auf den Graphen angewendet wird.

Die Parameterreihenfolge für den heuristischen Variationsalgorithmus (Abschnitt 4.3) lautet: Anzahl der Runden  $\times$  Anzahl der Schritte, die Bewertungsfunktion (beschrieben durch die Faktoren  $c_K$ ) und die Akzeptanzschwelle für das resultierende Rating.

Für die Motivtransformation (Abschnitt 4.4) wird zunächst das Zielmotiv angegeben, das erzeugt werden soll. Es folgt der Prozentsatz der Subgraphen, für die eine Transformation durchgeführt werden soll, die Operationsbeschränkung pro Transformation und die Suchschranke, welche die Anzahl der Versuche angibt, eine gültige Vertauschung wie in Abbildung 24 dargestellt, zu finden. Der letzte Parameter gibt an, wie oft der Algorithmus ausgeführt werden soll.

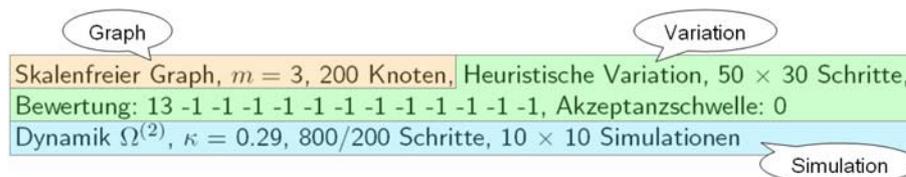


Abbildung 25: Beispiel eines Kopfes in einem Testprotokoll

Nach dem verwendeten Variationsalgorithmus werden die Parameter der Dynamik und der Simulation angegeben. Zunächst wird die Transitionsfunktion nebst zugehörigem Parameter  $\kappa$  ausgegeben, danach folgt die Simulationszeit, welche sich in den Transienten und die tatsächliche Simulationszeit aufteilt. Der Transient (erster Wert) wird nicht zur Auswertung der Dynamik herangezogen. Dadurch wird versucht eine größere Unabhängigkeit vom Anfangszustand zu erreichen und damit mögliche Störungen im System zu harmonisieren. Der zweite Wert ist die Anzahl der Zeitschritte, die für die Berechnung der Entropien relevant sind. Die letzten beiden Werte geben die Anzahl der Simulationen (erster Wert) und Simulationsrunden (zweiter Wert) an. Simulationsrunden sind dabei eigenständige Simulationen, deren Entropiewerte gespeichert und anschließend gemittelt werden, um auch hier Störungen zu minimieren. Abbildung 25 zeigt die einzelnen Bereiche des Kopfes noch einmal graphisch auf.

Der zweite Bereich, das Diagramm, enthält für jeden der berechneten (gemittelten) Entropiewerte einen Punkt. Die Punkte werden gemäß ihrer Berechnungsreihenfolge verbunden, um eine Trajektorie sichtbar zu machen.

Die Datentabelle enthält schließlich die Motivverteilung, den Z-Score und die Entropiewerte. Die erste Spalte (No.) gibt die Simulationsnummer an, 0 steht dabei für die erste Simulation, die noch auf dem Ausgangsgraphen agiert, während Simulation 1 die Motivverteilung und die Entropien nach der ersten Anwendung des Variationsalgorithmus angibt. Die mit  $\mathcal{Z}$  eingeleiteten Zeilen enthalten die Z-Scores der einzelnen Motivklassen, diese sind für den Ausgangsgraphen und für den Graphen, der nach Anwendung der letzten Variation entstanden ist, angegeben. Zur besseren Übersicht enthält die letzte Zeile noch einmal die Motivklassen in ihrer graphischen Repräsentation.

## 5.2 Reguläre Kreisgraphen

Zunächst wurden  $2k$ -reguläre Kreisgraphen getestet. Die Motivverteilungen für  $k = 1, \dots, 8$  wurden bereits in Abschnitt 3.2.3 betrachtet. Es sind ausschließlich die Motive 3 (3-Chain) und 7 (Feed-forward loop) vorhanden. Obwohl die Motivstruktur für gegebene  $k$  hier einem ganz bestimmten Schema folgt, sind die Werte für die Word-Entropie stark unterschiedlich. Sie reichen von 0.97 bis 1.93, während die Shannon-Entropie stets bei 0.99 bleibt. Einige typische Werte zeigt Tabelle 6.

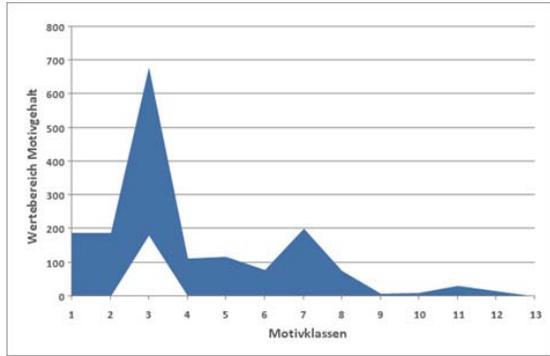
$k$	Shannon-Entropie		Word-Entropie	
	Mittelwert	Streuung	Mittelwert	Streuung
1	0.996333264	0.000430199	1.932964096	0.008430199
2	0.999101488	0.000088819	1.055249923	0.010646618
3	0.996389701	0.000460256	1.932725731	0.012989819
4	0.998742699	0.000157964	1.270710804	0.013121472
5	0.996380740	0.000454510	1.932160980	0.017220707
6	0.998537878	0.000158293	1.375511513	0.015009835
7	0.999380778	0.000074631	0.971022841	0.016293419
8	0.998381929	0.000191649	1.445584185	0.018492464

Tabelle 6: Entropiewerte von 200  $2k$ -regulären Kreisgraphen mit Transition  $\Omega^{(1)}$  und  $\kappa = 0.4$

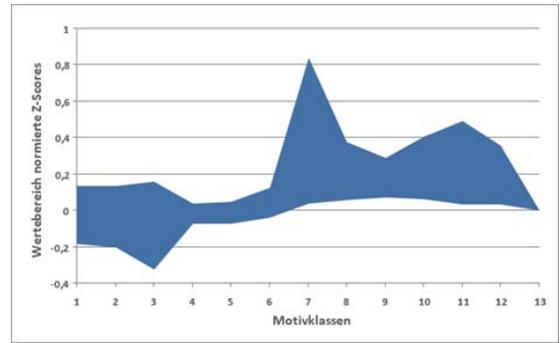
Beide vorgestellten Variationsalgorithmen liefern ein breites Spektrum von Motivverteilungen und normierten Z-Scores. Abbildung 26 zeigt die Bereiche, die von Motivgehalt und normiertem Z-Score angenommen werden im  $2 \cdot 2$ -regulären Kreisgraphen mit 200 Knoten.

Dabei zeigt sich, dass Motiv 3 noch höher geregelt werden kann (von 600 auf 681), während die Motivklasse 7 konsequent zerstört wird. Von den übrigen Motivklassen können bis zu 190 erzeugt werden, wobei sich Motive mit wenigen Kanten besser herstellen lassen, als Motive mit vielen Kanten.

Betrachtet man die normierten Z-Scores, durchläuft die Klasse 7 das größte Intervall. Dies ist mit der starken Abnahme des Motivgehalts (von anfänglich 200 bis zur völligen Vernichtung des Motivs) zu erklären.



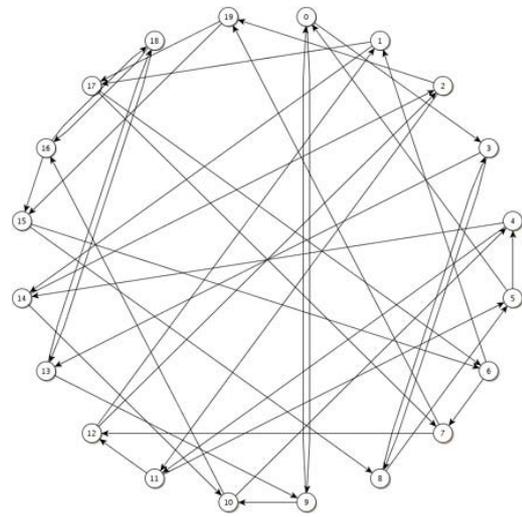
(a)



(b)

Abbildung 26: Auswirkungen der Variation im  $2 \cdot 2$ -regulären Kreisgraphen mit 200 Knoten

(a)



(b)

Abbildung 27:  $2 \cdot 2$ -regulärer Kreisgraphen mit 20 Knoten vor und nach der Variation

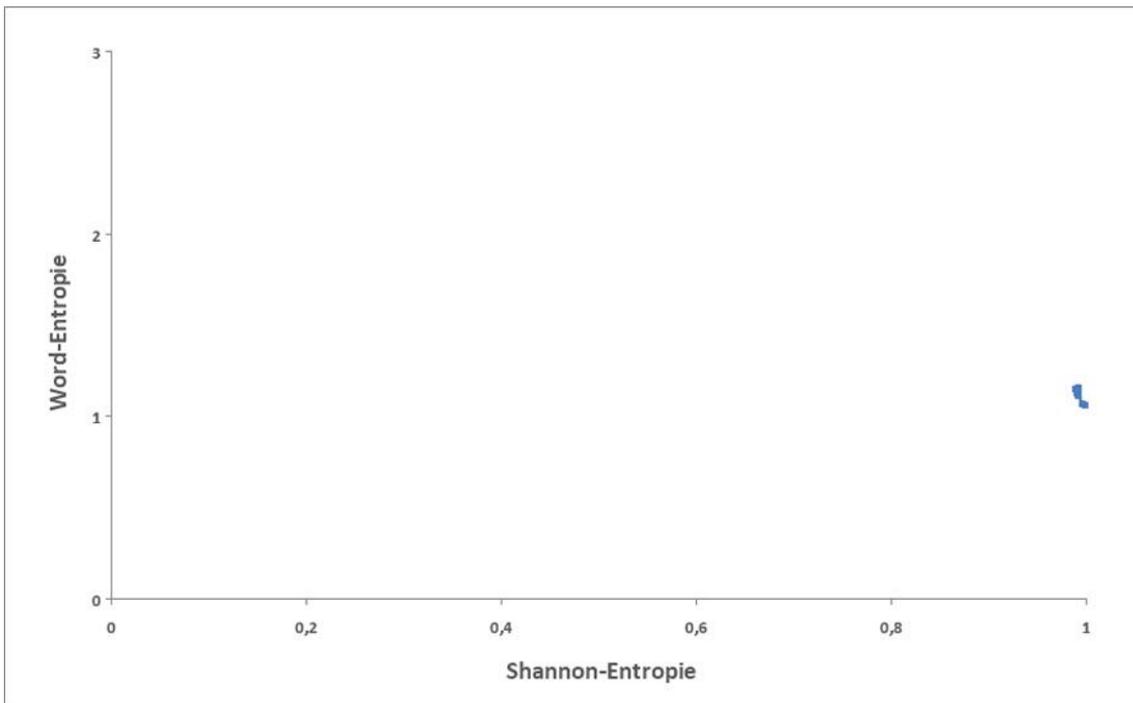
Ogleich sich die Motivverteilung der  $2k$ -regulären Kreisgraphen recht stark beeinflussen lässt, ist zunächst nur wenig Änderung in der Entropieebene zu erkennen. Bereits in Tabelle 6 war ersichtlich, dass die Shannon-Entropie im Ausgangsgraphen 0.99 beträgt. Die Zellzustände 0 und 1 müssen also in jedem Knoten nahezu gleichverteilt sein, was durch die homogene Struktur jedes Knoten erklärt werden kann. Während der Variation bleibt die Shannon-Entropie in dieser Größenordnung erhalten. Auch die Word-Entropie schwankt nur wenig im Bereich von 0.97 bis 1.15 ( $k = 2$ ) bzw. von 1.20 bis 1.49 ( $k = 4$ ).

Im regulären Kreisgraphen mit nur 20 Knoten, verhält sich die Dynamik nahezu gleich wie im Test mit 200 Knoten. Durch die geringe Knoten- und Kantenzahl kann hier jedoch der resultierende Graph visualisiert und untersucht werden. Abbildung 27 zeigt den  $2 \cdot 2$ -regulären Kreisgraphen mit 20 Knoten vor und nach der Variation. Es ist einsehbar, dass die reguläre Struktur des Graphen zwar zerstört ist, der Kreis-Charakter jedoch nach wie vor präsent ist, insbesondere, wenn man die Einbettung des Graphen ändert.

Die Abbildungen 28 und 29 zeigen die Entropieebene eines Tests von  $2k$ -regulären Kreisgraphen mit 200 Knoten und  $k \in \{2, 4\}$ .

## Testprotokoll

2 · 2-regulärer Kreisgraph, 200 Knoten, gezielte Motivtransformation, Motiv 6,  
 10.0 % aller Subgraphen, max. 3 Operationen, max. 300 Kantensuchen, 1 Schritt  
 Dynamik  $\Omega^{(1)}$ ,  $\kappa = 0.4$ , 800/200 Schritte,  $10 \times 10$  Simulationen

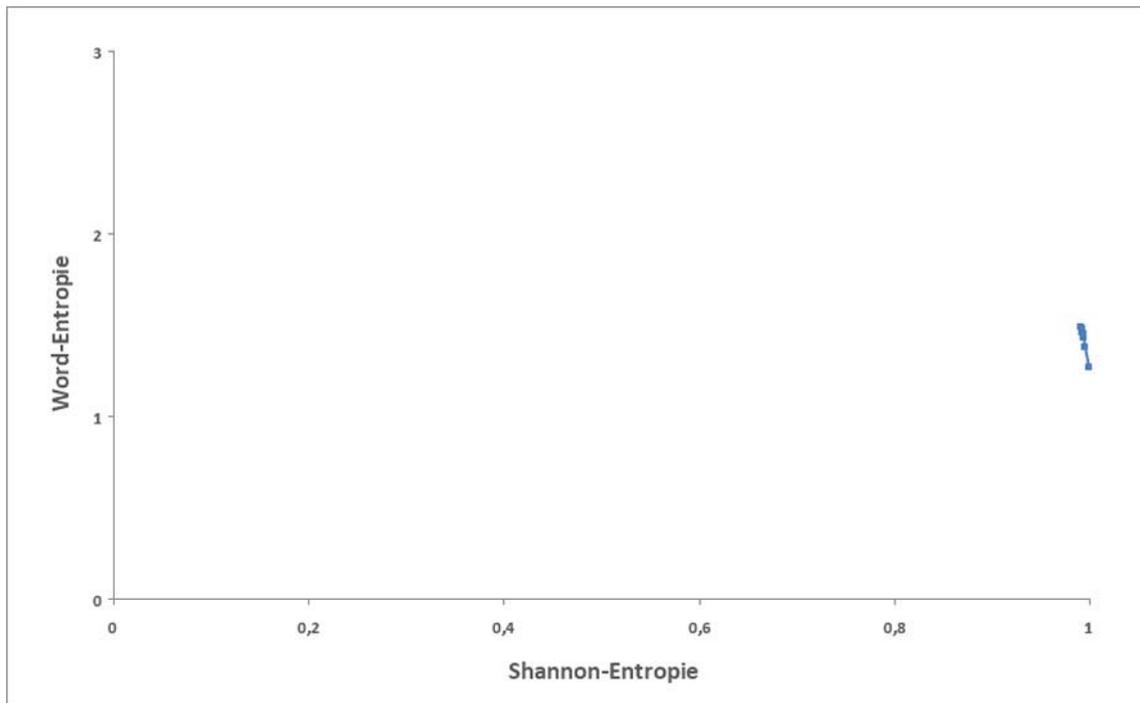


No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Shannon	Word
$\bar{z}$	0.0	0.0	37.4	0.0	0.0	0.0	207	0.0	0.0	0.0	0.0	0.0	0.0		
0	0	0	600	0	0	0	200	0	0	0	0	0	0	0.999	1.059
1	76	76	476	55	55	13	51	12	0	0	5	0	0	0.996	1.066
2	61	56	358	95	85	22	16	10	0	5	4	1	0	0.992	1.111
3	57	54	324	103	97	28	4	9	1	4	1	2	0	0.990	1.127
4	42	41	275	108	106	40	4	5	0	1	2	2	0	0.991	1.137
5	34	34	233	101	101	57	3	5	0	0	1	2	0	0.989	1.151
6	35	34	239	104	102	56	2	4	0	1	1	1	0	0.990	1.144
7	32	31	215	96	94	69	1	4	0	1	2	0	0	0.991	1.150
8	23	22	190	98	96	72	0	2	1	2	4	0	0	0.991	1.158
9	23	23	185	93	93	81	2	2	0	0	1	0	0	0.991	1.152
10	25	25	180	95	95	79	0	5	0	0	1	0	0	0.991	1.144
$\bar{z}$	1.3	1.3	-0.7	-0.9	-0.9	1.6	0.0	4.7	0.0	0.0	1.1	0.0	0.0		
	△	△	△	△	△	△	△	△	△	△	△	△	△		

Abbildung 28: Testprotokoll 2 · 2-regulärer Kreisgraphen mit 200 Knoten

## Testprotokoll

2 · 4-regulärer Kreisgraph, 200 Knoten, gezielte Motivtransformation, Motiv 6,  
 10.0 % aller Subgraphen, max. 3 Operationen, max. 300 Kantensuchen, 1 Schritt  
 Dynamik  $\Omega^{(1)}$ ,  $\kappa = 0.4$ , 800/200 Schritte,  $10 \times 10$  Simulationen



No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Shannon	Word
$\bar{z}$	0.0	0.0	37.6	0.0	0.0	0.0	321	0.0	0.0	0.0	0.0	0.0	0.0		
0	0	0	2000	0	0	0	1200	0	0	0	0	0	0	0.999	1.269
1	329	326	1194	418	412	129	121	34	34	37	63	19	0	0.995	1.386
2	288	294	968	522	534	204	31	19	16	10	46	26	3	0.993	1.434
3	254	251	852	551	545	274	14	17	9	12	29	24	1	0.993	1.457
4	237	240	803	542	548	325	9	12	6	3	20	14	7	0.990	1.490
5	203	205	729	568	572	365	5	5	4	2	17	9	4	0.991	1.485
6	236	237	823	591	593	328	5	6	3	2	10	9	1	0.993	1.456
7	214	217	773	594	600	350	7	7	5	2	7	5	1	0.992	1.464
8	237	237	821	598	598	340	4	5	0	0	6	6	1	0.992	1.465
9	239	238	832	611	609	329	8	8	1	2	4	3	0	0.993	1.447
10	237	238	843	609	611	326	5	2	3	2	8	3	0	0.993	1.449
$\bar{z}$	0.8	0.9	1.4	-1.4	-1.3	1.4	0.6	-1.1	1.1	0.3	0.9	-0.5	0.0		
	△	△	△	△	△	△	△	△	△	△	△	△	△		

Abbildung 29: Testprotokoll 2 · 4-regulärer Kreisgraphen mit 200 Knoten

Getestet wurden außerdem die gerichteten  $2k$ -regulären Kreisgraphen mit 200 Knoten und  $k \in \{1, \dots, 8\}$ . Die Variation wurde auf die Maximierung einer der 13 Motivklassen ausgerichtet; es wurde die heuristische und die gezielte Variation eingesetzt.

Keiner der Tests wies eine größere Änderung in den Entropien auf, so dass man folgende Beobachtungen festhalten kann:

**Beobachtung 1** *Nicht jede Änderung in der Motivverteilung wirkt sich direkt auf die Entropieebene aus. Ein Graph, der aus einem anderen durch Variation entsteht, kann das gleiche dynamische Verhalten wie der Ausgangsgraph haben.*

**Beobachtung 2** *Graphen, deren Gradverteilung einer Deltafunktion folgt, zeigten in den Tests ein ähnliches dynamisches Verhalten. Eine Abhängigkeit von der Motivverteilung konnte nicht nachgewiesen werden.*

### 5.3 Erdős-Rényi-Graphen

Bei einer Testreihe mit Erdős-Rényi-Graphen zeigte sich zunächst ebenfalls keinerlei Auswirkung auf die Entropieebene. Alle getesteten Graphen wiesen durchweg eine hohe, nahezu maximale Shannon-Entropie auf, während die Word-Entropie im Bereich zwischen 1.397 und 1.522 zu finden war. Eine Variation der Motive zeigte in keinem Testfall eine Änderung in den Entropien. Abbildung 30 zeigt ein solches Testprotokoll, bei dem keine Änderung zu beobachten war.

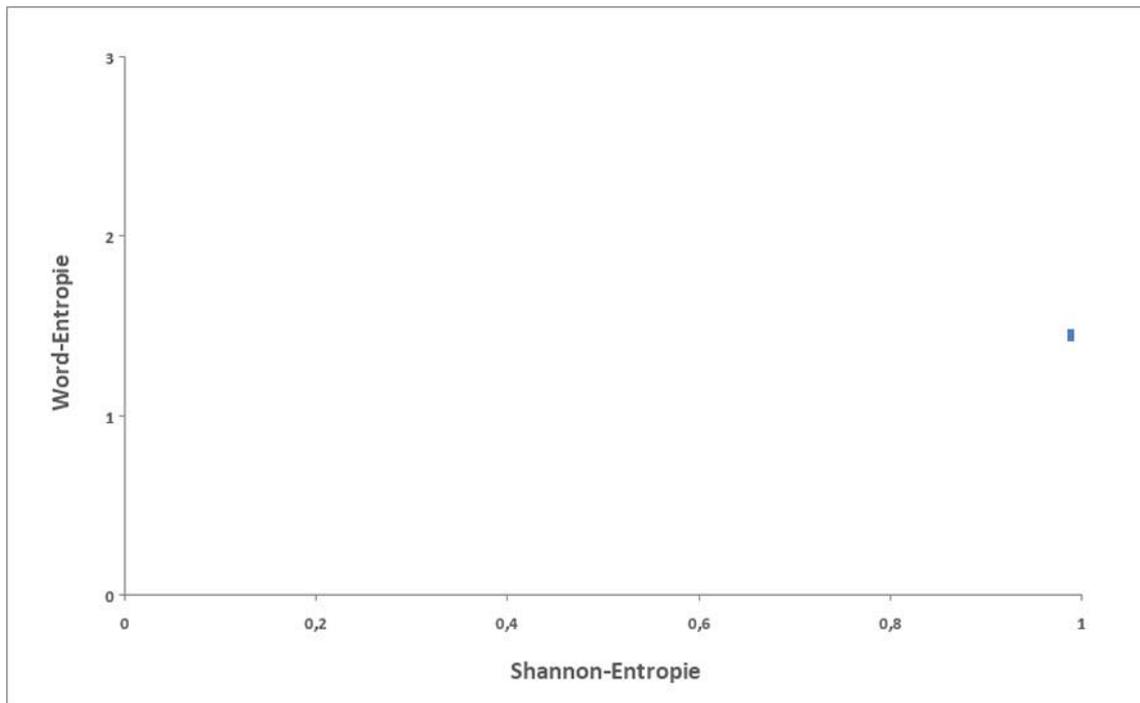
Daraufhin wurden zur Erzeugung des raumzeitlichen Musters genutzte Transitionsfunktion näher betrachtet. Die beiden Funktionen  $\Omega^{(1)}$  und  $\Omega^{(2)}$  zeigen für verschiedene Parameter  $\kappa$  unterschiedliche Ergebnisse der beiden Entropiemaße. Zu beobachten ist, dass eine komplexere Musterbildung nur in bestimmten Grenzen des Parameters  $\kappa$  auftreten können. Der Bereich liegt je nach Netzwerk etwa zwischen 0.2 und 0.4, sowie zwischen 0.6 und 0.8. Der zweite Bereich ergibt invertierte Muster, bei denen Nullen und Einsen vertauscht sind, die Entropie wird dadurch allerdings nicht verändert, so dass es genügt den ersten Bereich zu betrachten. Außerhalb der genannten Bereiche ist die Word-Entropie nahezu Null. Für  $\kappa$  unterhalb der Untergrenze, ist die Shannon-Entropie ebenso nahezu Null, während ein  $\kappa$  über der Obergrenze von ca. 0.4 einen nahezu maximalen Wert annimmt. Diese minimalen bzw. maximalen Werte führen dazu, dass die Muster ausschließlich die Wolfram-Klassen I und II annehmen.

Wendet man die Transitionsfunktion  $\Omega^{(1)}$  auf Erdős-Rényi-Graphen an, zeigen sich zwar verschiedenste Werte der Word-Entropie, die Shannon-Entropie bleibt jedoch auf dem Ausgangswert von ca. 0.973. Nicht so bei der Transitionsfunktion  $\Omega^{(2)}$ . Hier löst sich die Shannon-Entropie für mittlere  $\kappa$  von ihrer Maximalposition. In diesem Fall kann dann auch eine Trajektorie in der Entropieebene beobachtet werden, welche die einzelnen Entropien der Motivvariation durchläuft. Die Grenzen der komplexen Muster bilden sich bei  $\kappa \approx 0.24$  und  $\kappa \approx 0.35$ . Dazwischen ergibt sich ein Trend nach links oben für steigende  $\kappa$ , während die Motivvariation gerade in die entgegengesetzte Richtung strebt, nämlich nach rechts unten. Abbildung 31 zeigt die Entropiewerte der einzelnen Variationsschritte und die Trajektorien, die sich dadurch ergeben.

Besonders interessant ist hier der Wert  $\kappa = 0.29$  für  $\Omega^{(2)}$ , da sich hier die deutlichste Trajektorie bildet. Abbildung 32 zeigt das Testprotokoll des Versuchs. Die ersten beiden Entropiepaare weichen deutlich von der weiteren Variation ab. Die Motivverteilung zeigt, dass Motive mit drei Kanten, also die Klassen 4, 5 und 7 abgebaut werden. Die Motivklasse 8 besitzt auch nur drei Kanten, wird allerdings sogar ein wenig erhöht. Stark erhöht werden die Motive mit zwei Kanten. Vor allem im ersten Variationsschritt ist eine große Veränderung der Motivverteilung festzustellen, was den ersten Ausschlag erklären könnte. Der zweite Variationsschritt zeigt leider keine solch deutliche Veränderung in den Motiven, die den enormen Sprung in der Shannon-Entropie rechtfertigen würde.

## Testprotokoll

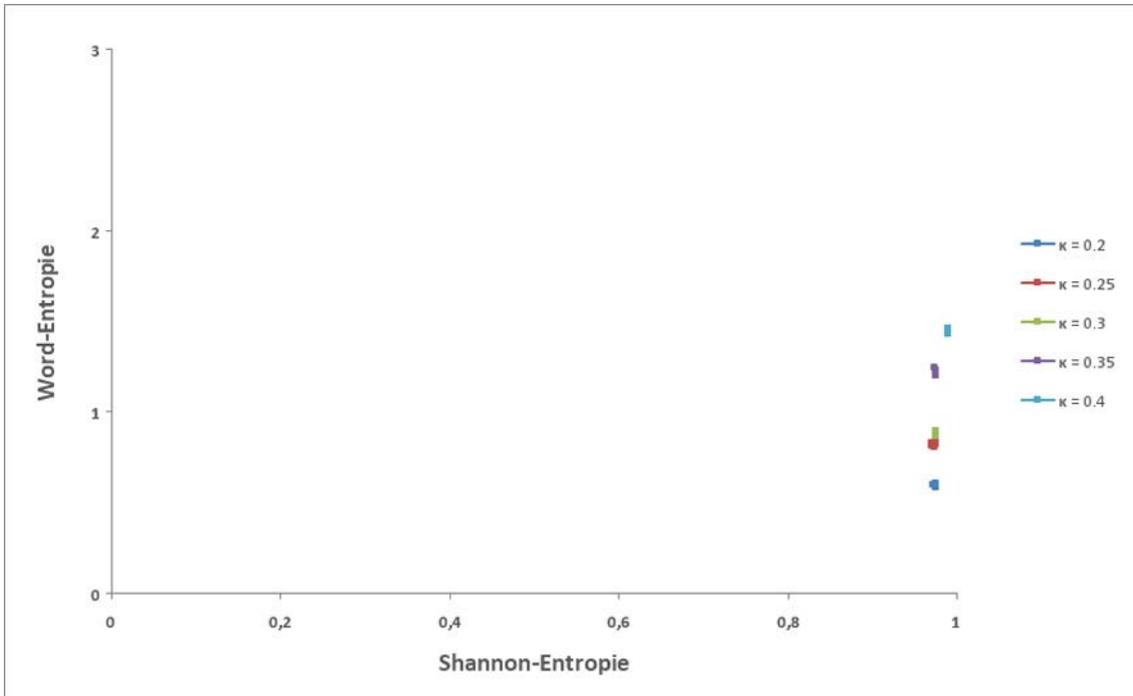
Erdős-Rényi-Graph,  $p = 0.02$ , 200 Knoten, Heuristische Variation,  $50 \times 30$  Schritte,  
 Bewertung: 13 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1, Akzeptanzschwelle: 0  
 Dynamik  $\Omega^{(1)}$ ,  $\kappa = 0.4$ , 800/200 Schritte,  $10 \times 10$  Simulationen



No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Shannon	Word
$\bar{z}$	-0.3	-0.5	0.4	0.1	-0.1	0.0	0.2	-0.6	1.5	1.3	1.1	0.0	0.0		
0	1653	1593	3175	70	72	0	71	20	2	2	3	0	0	0.988	1.463
1	1735	1668	3299	17	11	0	57	31	0	1	0	0	0	0.988	1.460
2	1753	1691	3341	0	0	0	52	30	0	0	0	0	0	0.988	1.456
3	1753	1691	3341	0	0	0	52	30	0	0	0	0	0	0.988	1.450
4	1763	1701	3351	0	0	0	42	30	0	0	0	0	0	0.988	1.462
5	1766	1704	3351	0	0	0	39	31	0	0	0	0	0	0.988	1.457
6	1768	1706	3356	0	0	0	37	30	0	0	0	0	0	0.988	1.458
7	1768	1706	3356	0	0	0	37	30	0	0	0	0	0	0.988	1.461
8	1768	1706	3356	0	0	0	37	30	0	0	0	0	0	0.988	1.456
9	1768	1706	3356	0	0	0	37	30	0	0	0	0	0	0.988	1.435
10	1770	1708	3364	0	0	0	35	28	0	0	0	0	0	0.988	1.463
$\bar{z}$	4.4	4.4	-0.7	0.0	0.0	0.0	-4.8	0.8	0.0	0.0	0.0	0.0	0.0		
	△	△	△	△	△	△	△	△	△	△	△	△	△		

Abbildung 30: Testprotokoll Erdős-Rényi-Graph mit  $p = 0.02$  und 200 Knoten

### Transition $\Omega^{(1)}$



### Transition $\Omega^{(2)}$

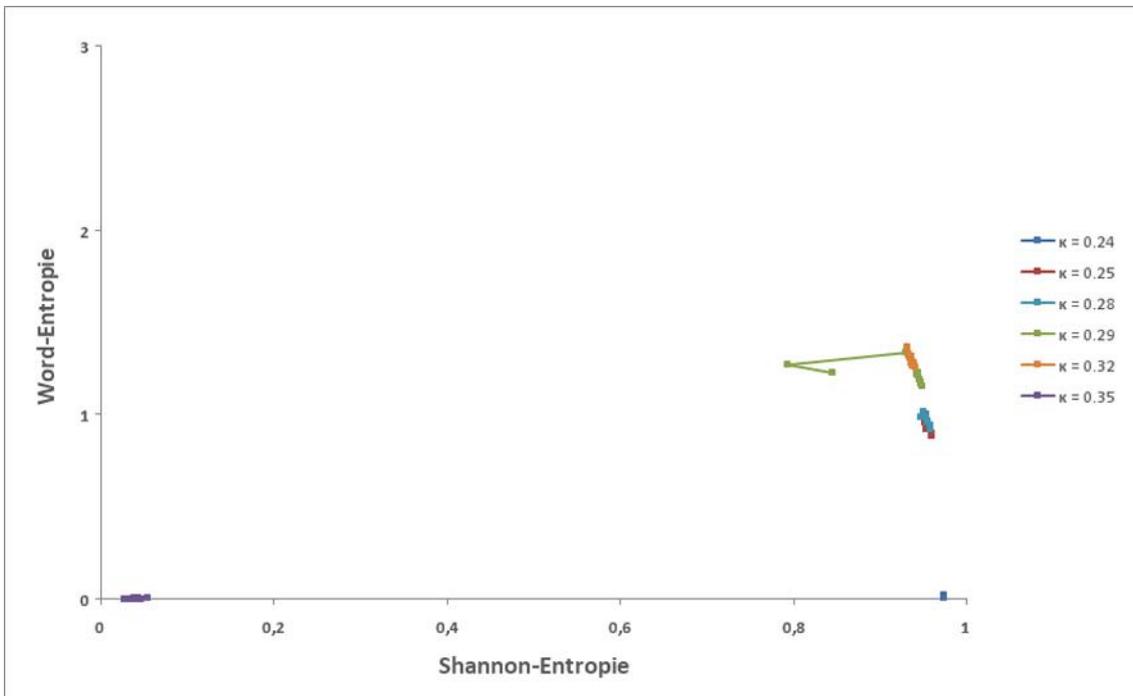
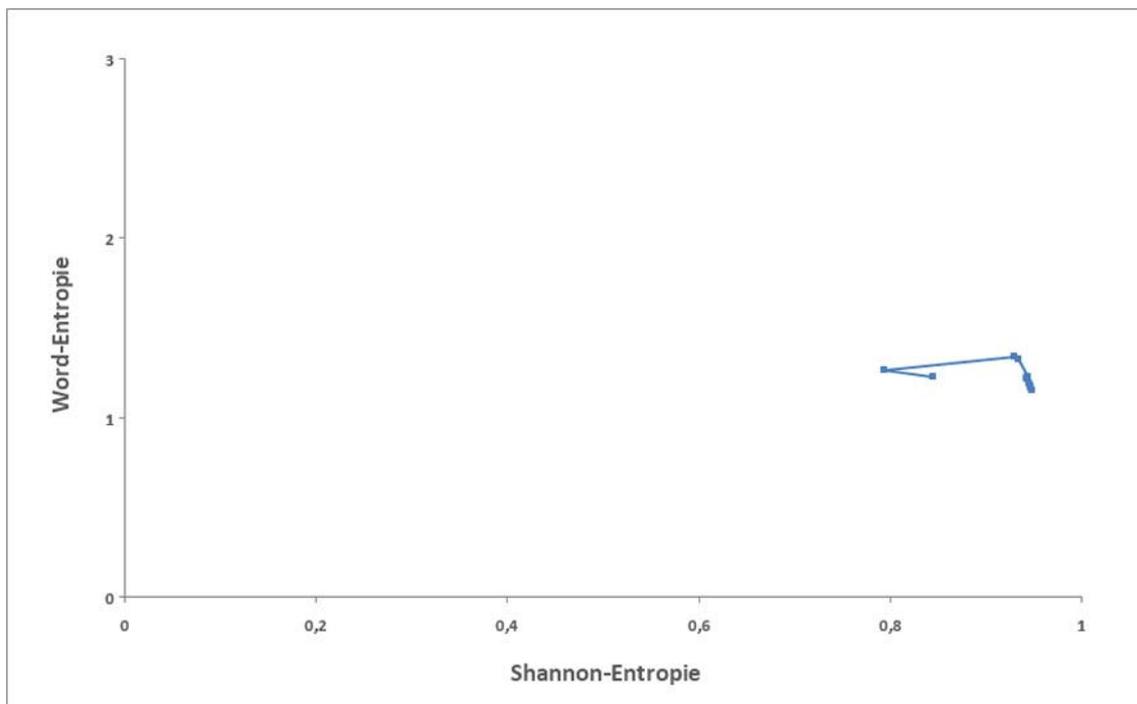


Abbildung 31: Entropiewerte der beiden Transitionsfunktionen  $\Omega$  mit verschiedenen Parametern  $\kappa$

## Testprotokoll

Erdős-Rényi-Graph,  $p = 0.02$ , 200 Knoten, Heuristische Variation,  $50 \times 30$  Schritte,  
 Bewertung: 13 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1, Akzeptanzschwelle: 0  
 Dynamik  $\Omega^{(2)}$ ,  $\kappa = 0.29$ , 800/200 Schritte,  $10 \times 10$  Simulationen



No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Shannon	Word
$\bar{z}$	-0.3	-0.5	0.4	0.1	-0.1	0.0	0.2	-0.6	1.5	1.3	1.1	0.0	0.0		
0	1280	1340	2681	99	111	1	53	23	0	1	1	0	0	0.844	1.228
1	1396	1436	2902	13	6	0	46	25	0	0	0	0	0	0.793	1.271
2	1408	1453	2921	5	3	0	37	26	0	0	0	0	0	0.929	1.340
3	1414	1461	2937	0	0	0	34	25	0	0	0	0	0	0.933	1.326
4	1417	1464	2943	0	0	0	31	24	0	0	0	0	0	0.943	1.232
5	1419	1466	2942	0	0	0	29	25	0	0	0	0	0	0.943	1.216
6	1419	1466	2942	0	0	0	29	25	0	0	0	0	0	0.942	1.222
7	1422	1469	2951	0	0	0	26	23	0	0	0	0	0	0.946	1.179
8	1426	1471	2934	4	2	0	20	28	0	0	0	0	0	0.948	1.153
9	1426	1471	2934	4	2	0	20	28	0	0	0	0	0	0.945	1.191
10	1426	1471	2925	4	2	0	20	31	0	0	0	0	0	0.946	1.172
$\bar{z}$	5.8	5.4	-1.9	-1.4	-2.2	0.0	-5.1	2.4	0.0	0.0	0.0	0.0	0.0		
	∧	∧	∧	∧	∧	∧	△	△	△	△	△	△	△		

Abbildung 32: Testprotokoll Erdős-Rényi-Graph mit Transition  $\Omega^{(2)}$

Die zusammengefassten Beobachtungen bezüglich der Transitionsfunktion sind also:

**Beobachtung 3** Die Transitionsfunktionen  $\Omega^{(1)}$  und  $\Omega^{(2)}$  bilden Grenzen aus, zwischen denen eine komplexe Musterbildung möglich ist. Unterhalb und oberhalb der Grenzwerte ist die Word-Entropie nahezu Null, wodurch keine Änderungen mehr sichtbar werden.

**Beobachtung 4** Die Wahl der Transitionsfunktion und des Parameters  $\kappa$  ist entscheidend für die Beobachtung von Trajektorien in der Entropieebene.

**Beobachtung 5** Bei extremen Entropiewerten, also einer Word-Entropie von nahezu Null oder einer Shannon-Entropie von nahezu Null bzw. Eins, konnte keine Trajektorie beobachtet werden. Die Entropiewerte bleiben in einem solchen Fall weitgehend konstant.

Neben den Erdős-Rényi-Graphen, wie sie in Abschnitt 2.1.2 eingeführt wurden, wurden auch azyklische Erdős-Rényi-Graphen für die Experimente verwendet. Azyklische Graphen sind wie folgt definiert:

**Definition 5.1 (azyklischer Graph)** Sei  $G = (V, E)$  ein gerichteter Graph.  $G$  heißt azyklisch, falls es keine Zyklen in  $G$  gibt, also keinen Kantenzug der Form

$$(v, u_1) \in E, (u_1, u_2) \in E, \dots, (u_n, v) \in E \quad \text{für alle } v \in V$$

existiert. Meist spricht man auch von einem zyklensfreien Graphen.

Für den algorithmischen Aufbau von azyklischen Graphen ist ein griffigeres Kriterium für Zyklensfreiheit sinnvoll, damit die Auswertung im Programm effizient durchgeführt werden kann:

**Bemerkung 5.2** Sei  $G = (V, E)$  ein gerichteter Graph mit  $V = \{v_1, v_2, \dots, v_n\}$  und einer Ordnung  $v_1 < v_2 < \dots < v_n$ .  $G$  ist azyklisch, falls  $E$  nur Kanten enthält, deren Startknoten kleiner als der Zielknoten ist, falls also gilt:

$$u < v \quad \text{für alle } (u, v) \in E.$$

Die hier betrachteten azyklischen Erdős-Rényi-Graphen werden mit dem in Definition 2.9 eingeführten Algorithmus erzeugt. Soll ein Knotenpaar  $(u, v)$  verbunden werden, wird die Kante so eingefügt, dass der Startknoten kleiner dem Zielknoten ist. Das Verfahren entspricht dem Erzeugen eines „normalen“ ER-Graphen und das anschließende Sortieren der einzelnen Start- und Zielknoten.

Azyklische Erdős-Rényi-Graphen zeigen eine andere Motivverteilung und andere Entropiewerte. Insbesondere tritt die Motivklasse 3 (3-Chain) nicht mehr am häufigsten auf, vielmehr entspricht die Anzahl etwa der der Klassen 1 und 2. Außerdem können natürlich nur noch Motive auftreten, die keinen Zyklus

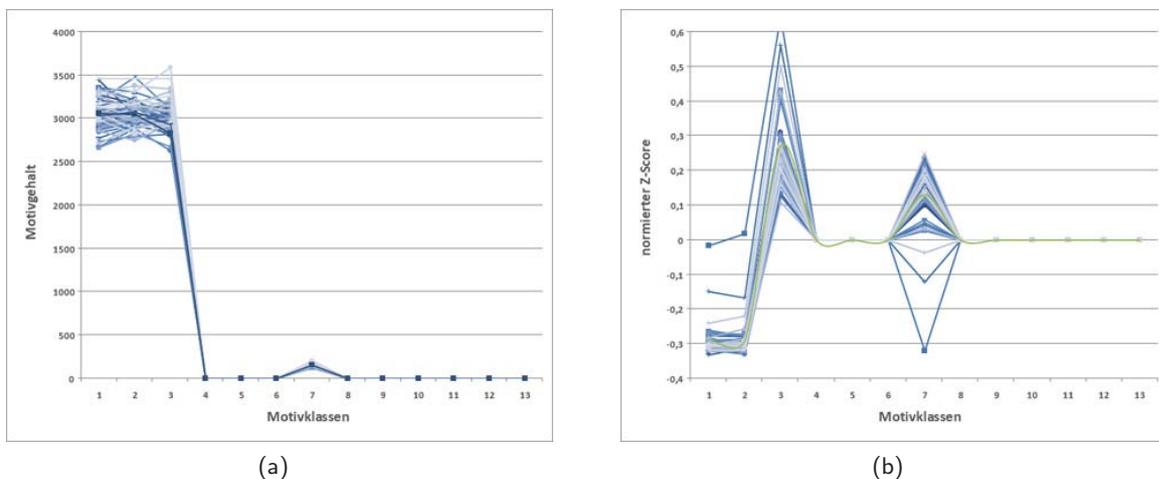


Abbildung 33: Motivverteilung und normierter Z-Score in gerichteten azyklischen ER-Graphen

enthalten, diese sind: 1, 2, 3 und 7. Die Z-Scores zeigen entgegen den nicht zyklischen ER-Graphen nun eine Regelmäßigkeit: Motiv 3 ist signifikant überhöht, während die ersten beiden Motive vermindert auftreten. Abbildung 33 zeigt die Motivverteilung und den normalisierten Z-Score von azyklischen Erdős-Rényi-Graphen.

Die Entropien zeigen einen ebenso deutlichen Unterschied, wie die Motivverteilung. Beide Werte sind deutlich niedriger als in nicht zyklischen ER-Graphen. Tabelle 7 zeigt einen Vergleich der beiden Entropien.

Graph	Shannon-Entropie	Word-Entropie
zyklisch	0.974	1.497
azyklisch	0.640	0.461

Tabelle 7: Typische Entropiewerte in zyklischen und nicht zyklischen Erdős-Rényi-Graphen

Das Interessante an der Betrachtung von azyklischen Graphen ist, dass nahezu mit jeder Konfiguration eine Trajektorie in der Entropieebene sichtbar wurde. Der Trend dieser Trajektorien war immer der gleiche: Die Entropien laufen in Richtung der rechten oberen Ecke. Abbildung 34 zeigt einige dieser Trajektorien, ein vollständiges Testprotokoll, das auch die Motivverteilung vor und nach der Variation enthält findet sich dann in Abbildung 35

Der Vergleich von zyklischen und nicht zyklischen Erdős-Rényi-Graphen bringt nun folgende Beobachtungen mit sich:

**Beobachtung 6** Die Betrachtung von zyklischen Erdős-Rényi-Graphen wies deutlich häufiger eine Trajektorie in der Entropieebene auf, als nicht zyklische. Die Trajektorien sind außerdem wesentlich länger. Zyklen scheinen also einen großen Einfluss auf die Entropiewerte zu haben.

**Beobachtung 7** Die Trajektorie in der Entropieebene von azyklischen Erdős-Rényi-Graphen hat für verschiedene Graphen eine ähnliche Gestalt und läuft nach rechts oben (Erhöhung der Shannon- und der Word-Entropie).

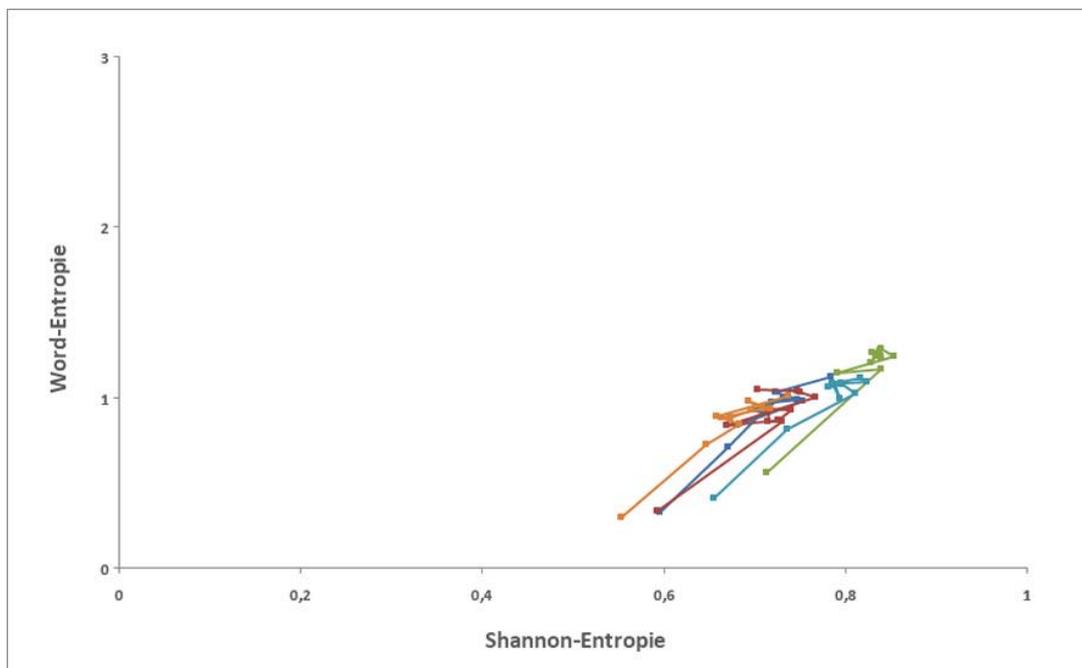


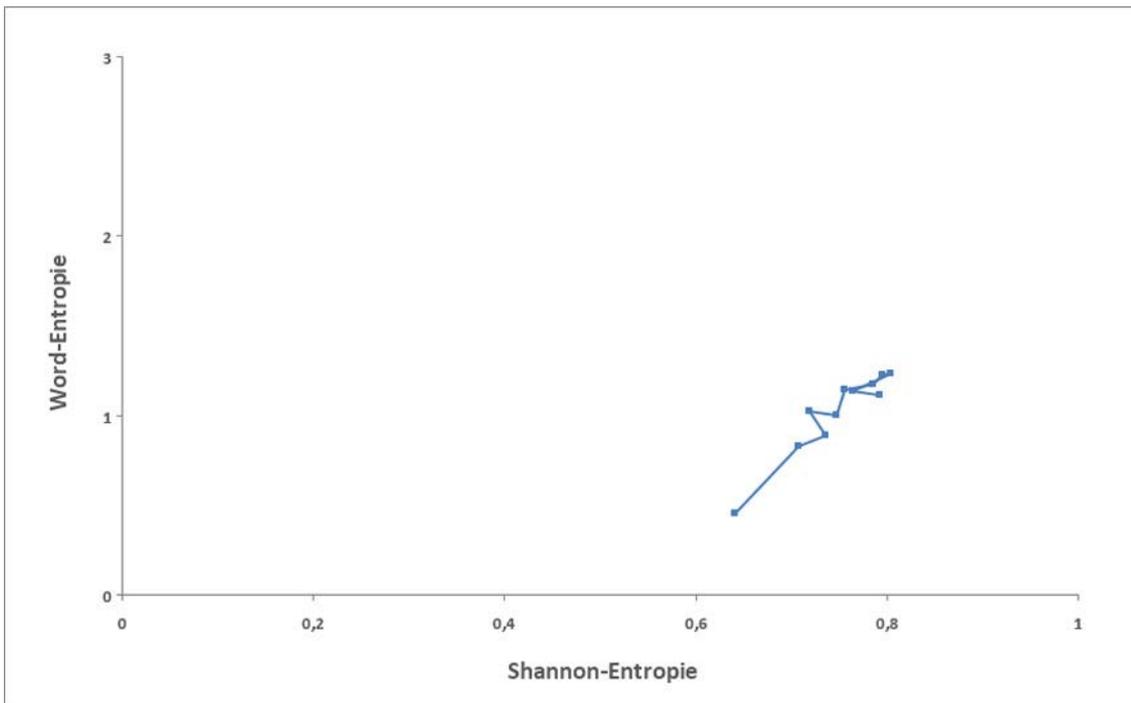
Abbildung 34: Trajektorien in azyklischen Erdős-Rényi-Graphen

## Testprotokoll

Erdős-Rényi-Graph,  $p = 0.025$ , 200 Knoten, Heuristische Variation,  $50 \times 30$  Schritte,

Bewertung: 13 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1, Akzeptanzschwelle: 0

Dynamik  $\Omega^{(1)}$ ,  $\kappa = 0.4$ , 800/200 Schritte,  $10 \times 10$  Simulationen



No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Shannon	Word
$\bar{z}$	-3.0	-2.9	2.6	0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0		
0	1598	1653	1693	0	0	0	60	0	0	0	0	0	0	0.640	0.461
1	1615	1670	1707	0	0	0	43	1	0	0	0	0	0	0.707	0.829
2	1633	1688	1722	0	0	0	25	2	0	0	0	0	0	0.735	0.890
3	1634	1689	1714	0	0	0	24	5	0	0	0	0	0	0.718	1.030
4	1634	1689	1714	0	0	0	24	5	0	0	0	0	0	0.747	1.008
5	1634	1689	1714	0	0	0	24	5	0	0	0	0	0	0.755	1.150
6	1635	1690	1712	0	0	0	23	6	0	0	0	0	0	0.785	1.175
7	1638	1693	1721	0	0	0	20	4	0	0	0	0	0	0.795	1.227
8	1638	1693	1718	0	0	0	20	5	0	0	0	0	0	0.803	1.234
9	1641	1696	1721	0	0	0	17	5	0	0	0	0	0	0.763	1.140
10	1641	1696	1721	0	0	0	17	5	0	0	0	0	0	0.792	1.117
$\bar{z}$	4.7	4.6	2.3	0.0	0.0	0.0	-5.5	0.0	0.0	0.0	0.0	0.0	0.0		
	△	△	△	△	△	△	△	△	△	△	△	△	△		

Abbildung 35: Testprotokoll azyklischer Erdős-Rényi-Graph mit 200 Knoten

## 5.4 Skalenfreie Graphen

Auch bei skalenfreien Graphen ist die verwendete Transitionsfunktion ausschlaggebend für die entstehenden Trajektorien innerhalb der Entropieebene. Ausgehend von Beobachtung 4 werden also die Funktionen  $\Omega^{(1)}$  und  $\Omega^{(2)}$  mit verschiedenen  $\kappa$ -Werten getestet. Abbildung 37 zeigt die Ergebnisse.

Das Verhalten der Transitionsfunktion  $\Omega^{(1)}$  ähnelt sehr stark den Erdős-Rényi-Graphen: Die Shannon-Entropie ist nahezu maximal, die Word-Entropie nimmt je nach  $\kappa$  einen mittleren Wert an. Bereits in Beobachtung 5 konnte festgestellt werden, dass sich bei maximalen Entropiewerten keine großen Trajektorien bilden, dies bewahrheitet sich nun auch für skalenfreie Graphen.

Die Funktion  $\Omega^{(2)}$  zeigte bereits für ER-Graphen eine Ausbildung von Trajektorien in der Entropieebene. Bei skalenfreien Graphen werden diese Trajektorien noch wesentlich deutlicher. Die Grenzen für eine komplexere Dynamik liegen hier bei etwa 0.2 und 0.35. Zu beobachten ist, dass die Werte innerhalb der Grenzen stark variieren, so bewegen sich die Word-Entropien nach den einzelnen Variationsschritten für  $\kappa = 0.3$  zwischen 0.005 und 0.629, während die Shannon-Entropien die Werte 0.081 bis 0.443 durchlaufen.

Außerdem wurden mehrere Testläufe mit einer gleichen Parametrisierung durchgeführt, um die Abweichungen der einzelnen Experimente besser einschätzen zu können. Abbildung 36 zeigt fünf Trajektorien, die durch die Transition  $\Omega^{(2)}$  und Parameter  $\kappa = 0.3$  entstanden sind. Die Entropiewerte sammeln sich hier zunächst nahe bei Null. Im Laufe der Variation zeigen die Werte einen einheitlichen Trend nach rechts oben, wobei es oftmals zu einem Einbruch (und zu etwaigen Neuanstiegen) der Entropie kommt.

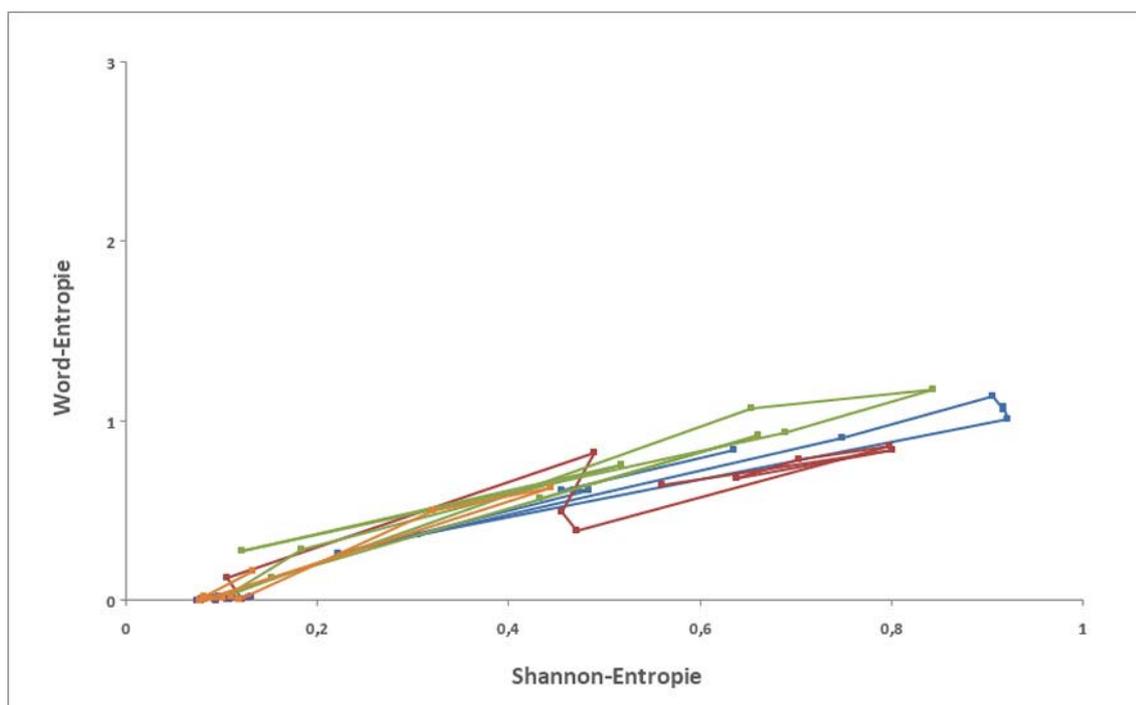
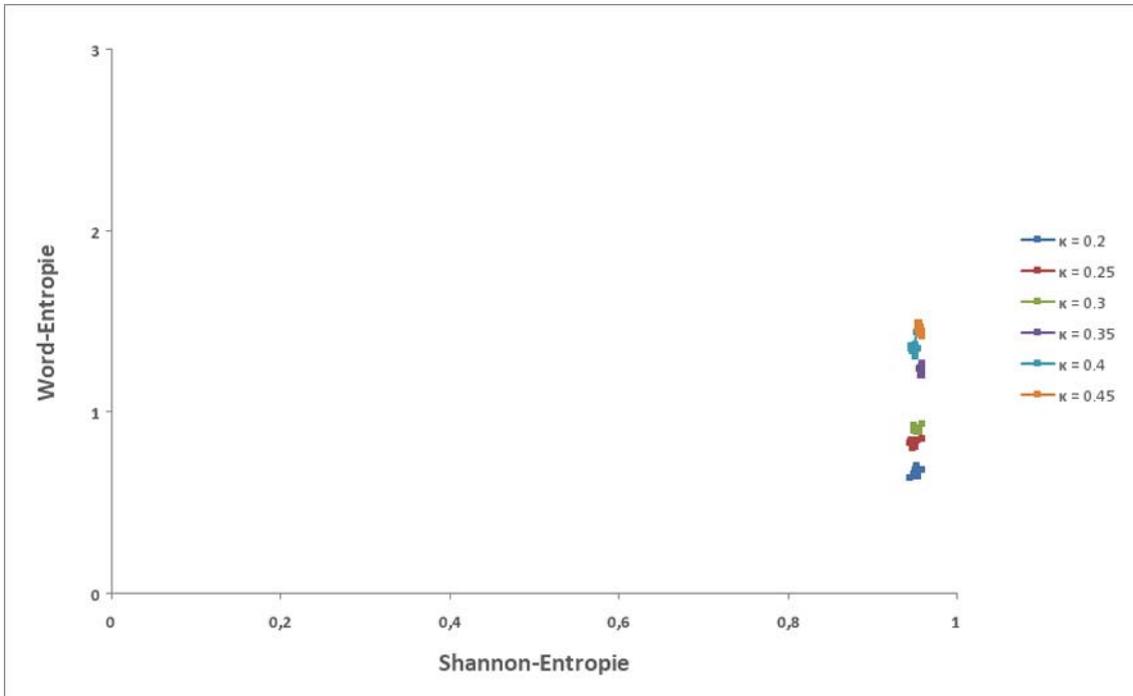


Abbildung 36: Verschiedene Trajektorien für  $\Omega^{(2)}$  mit Parameter  $\kappa = 0.3$

### Transition $\Omega^{(1)}$



### Transition $\Omega^{(2)}$

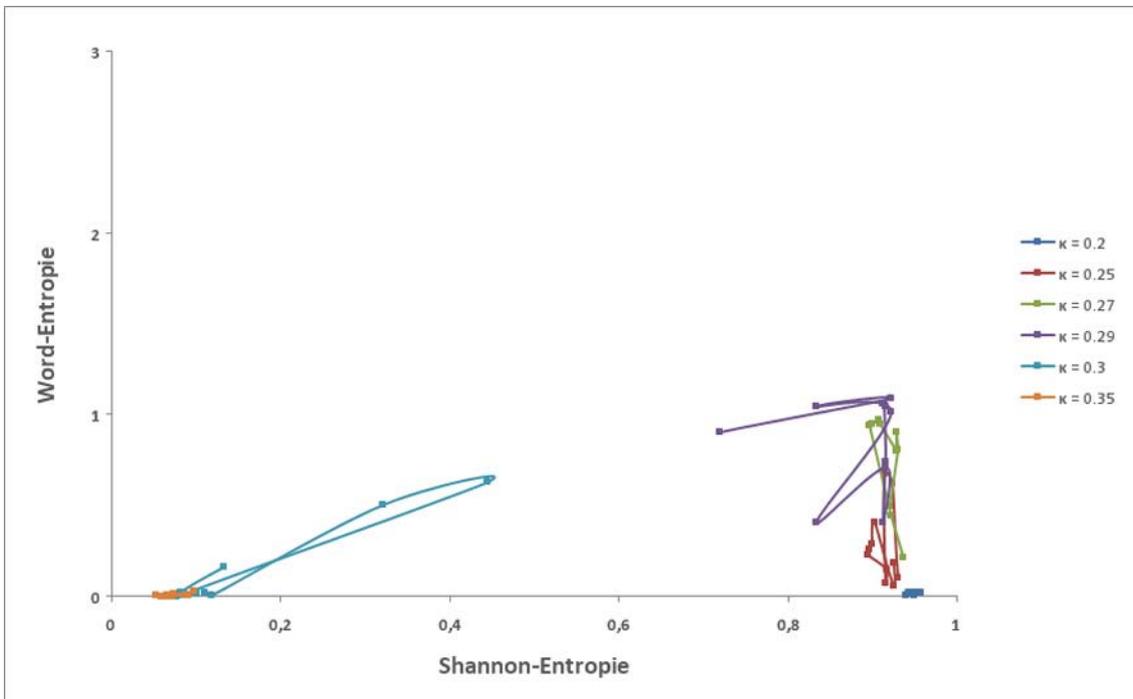


Abbildung 37: Entropiewerte der beiden Transitionsfunktionen  $\Omega$  mit verschiedenen Parametern  $\kappa$

## 5.5 Beitrag der einzelnen Motivklassen zur Entropie

Die starken Änderungen und Sprünge in den Entropie von skalenfreien Graphen sollen nun näher untersucht werden, insbesondere soll festgestellt werden, ob ein Zusammenhang zur Motivverteilung besteht. Grundidee der Analyse ist, die Entropiewerte zu zerlegen und den Anteil einer jeden Motivklasse am Entropiewert anzugeben. Um einen solchen Wert zu erhalten wird die Shannon-Entropie nicht mehr knotenweise bestimmt, sondern für jede Motivklasse gebildet. Die so entstandenen Werte werden zuletzt noch normiert, so dass die Werte besser verglichen werden können.

Diese so genannte Motiv-Entropie wird wie folgt definiert:

**Definition 5.3 (Motiv-Entropie)** Sei  $\mathcal{GA} = (G, S, \chi_0, \delta_\chi)$  ein zellulärer Graphenautomat und  $\mathcal{K}_k$  die Motivklassenmenge des Graphen  $G$ . Die Motiv-Entropie  $\mathcal{S}_K$  der Motivklasse  $K \in \mathcal{K}_k$  ist definiert als

$$\mathcal{S}_K = - \sum_{s \in S} p_s(K) \log_2(p_s(K)),$$

wobei  $p_s(K)$  die Wahrscheinlichkeit bezeichnet, mit der sich eine Zelle  $c$  eines Motivs  $M \in K$  der Klasse  $K$  im Zustand  $s \in S$  befindet. Sie werden wieder durch die relativen Häufigkeiten approximiert.

Die Motiv-Entropie kann dann über der Summe aller Motiv-Entropien normiert werden, um für Vergleiche verwendet werden zu können. Es ergibt sich:

$$\mathcal{S}_K^0 = \frac{\mathcal{S}_K}{\sum_{K' \in \mathcal{K}_k} \mathcal{S}_{K'}}.$$

**Beispiel 5.4** Angenommen wir betrachteten Motive der Größe  $k = 3$  in einem zellulären Graphenautomaten über  $T = 200$  Zeitschritte hinweg. Während der Simulation wird für jede Motivklasse  $K \in \mathcal{K}_k$  gezählt, wie oft die enthaltenen Zellen den Zustand  $1 \in S$  annehmen. Dieser Wert wird mit  $c_1$  bezeichnet. Ermittelt wurden:

$K$	1	2	3	4	5	6	7	8	9	10	11	12	13
$ K $	2390	2475	4933	0	0	0	187	80	0	0	0	0	0
$c_1$	163700	182750	349700	0	0	0	14150	5000	0	0	0	0	0

Aus diesen Werten werden nun die Wahrscheinlichkeiten  $p_1(K)$  und  $p_0(K)$  berechnet, woraus sich die Motiv-Entropie  $\mathcal{S}_K$  ergibt. Die Wahrscheinlichkeit  $p_1(K)$  ergibt sich dabei aus der Anzahl der 1-Zustände normiert über der Gesamtzahl der möglichen Zustände im Motiv. Letztere ergibt sich aus der Anzahl der Zellen in der Motivklasse  $|K| \cdot k$  multipliziert mit der Simulationszeit  $T$ . Für die oben angegebenen Werte ergibt sich also:

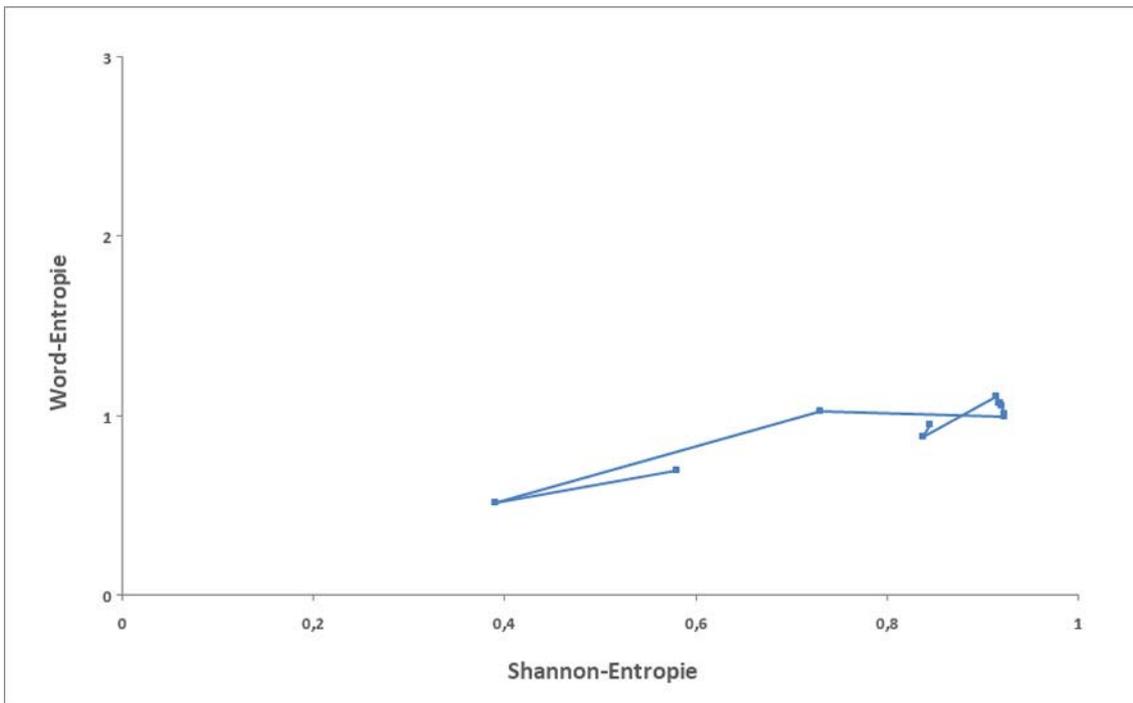
$K$	1	2	3	4	5	6	7	8	9	10	11	12	13
$c =  K  \cdot k \cdot T$	1434000	1485000	2959800	0	0	0	112200	48000	0	0	0	0	0
$p_1(K) = c_1/c$	0.1142	0.1231	0.1181	0	0	0	0.1261	0.1042	0	0	0	0	0
$\log_2(p_1(K))$	-3.1309	-3.0225	-3.0813	0	0	0	-2.9872	-3.2630	0	0	0	0	0
$p_0(K) = 1 - p_1(K)$	0.8858	0.8769	0.8819	1	1	1	0.8739	0.8958	1	1	1	1	1
$\log_2(p_0(K))$	-0.1749	-0.1895	-0.1814	0	0	0	-0.1945	-0.1587	0	0	0	0	0
$\mathcal{S}_K$	0.5123	0.5381	0.5240	0	0	0	0.5467	0.4821	0	0	0	0	0
$\mathcal{S}_K^0$	19.68%	20.67%	20.13%	0	0	0	21.00%	18.52%	0	0	0	0	0

Das Beispiel zeigt, dass die Motiv-Entropie der Klasse 7 höher ist, als die der Klasse 1. Der Schluss liegt nahe, dass eine größere Menge von Motiven der Klasse 7 insgesamt zu einer höheren Entropie führen.

Abbildung 38 zeigt das Testprotokoll eines gerichteten skalenfreien Graphen mit 200 Knoten. Die Entropien wurden mit der Transitionsfunktion  $\Omega^{(2)}$  und dem Parameter  $\kappa = 0.29$  ausgewertet. Die normierten Motiv-Entropien für die einzelnen Variationsschritte sind in Tabelle 8 gelistet.

## Testprotokoll

Skalenfreier Graph,  $m = 3$ , 200 Knoten, Heuristische Variation,  $50 \times 30$  Schritte,  
 Bewertung: 13 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1, Akzeptanzschwelle: 0  
 Dynamik  $\Omega^{(2)}$ ,  $\kappa = 0.29$ , 800/200 Schritte,  $10 \times 10$  Simulationen



No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Shannon	Word
$\bar{z}$	0.6	0.1	1.3	0.2	-0.7	-1.0	-0.8	-0.9	0.1	1.2	-1.5	0.0	0.0		
0	1298	941	2244	196	281	16	53	16	4	10	5	0	0	0.579	0.699
1	1567	1112	2670	36	30	0	76	32	1	0	3	0	0	0.390	0.515
2	1597	1144	2751	7	2	0	78	29	0	0	0	0	0	0.729	1.030
3	1609	1156	2754	7	2	0	66	32	0	0	0	0	0	0.922	0.994
4	1609	1156	2754	7	2	0	66	32	0	0	0	0	0	0.922	1.010
5	1623	1170	2774	7	2	0	52	30	0	0	0	0	0	0.918	1.072
6	1623	1170	2774	7	2	0	52	30	0	0	0	0	0	0.917	1.072
7	1623	1170	2774	7	2	0	52	30	0	0	0	0	0	0.919	1.055
8	1624	1176	2784	0	0	0	53	30	0	0	0	0	0	0.914	1.108
9	1635	1187	2774	0	0	0	42	37	0	0	0	0	0	0.837	0.883
10	1639	1191	2781	0	0	0	38	36	0	0	0	0	0	0.844	0.956
$\bar{z}$	4.7	5.0	-0.7	0.0	0.0	0.0	-4.7	0.7	0.0	0.0	0.0	0.0	0.0		
	△	△	△	△	△	△	△	△	△	△	△	△	△		

Abbildung 38: Testprotokoll skalenfreier Graph mit 200 Knoten

No.	$S_1^0$	$S_2^0$	$S_3^0$	$S_4^0$	$S_5^0$	$S_6^0$	$S_7^0$	$S_8^0$	$S_9^0$	$S_{10}^0$	$S_{11}^0$	$S_{12}^0$	$S_{13}^0$
0	9.6%	9.7%	9.7%	9.0%	9.2%	8.8%	9.3%	8.6%	9.9%	8.3%	8.0%	0.0%	0.0%
1	12.2%	12.3%	12.3%	10.2%	9.8%	0.0%	13.3%	12.4%	8.1%	0.0%	9.4%	0.0%	0.0%
2	14.5%	14.4%	14.4%	13.2%	13.2%	0.0%	14.9%	15.3%	0.0%	0.0%	0.0%	0.0%	0.0%
3	14.3%	14.2%	14.3%	14.2%	14.2%	0.0%	14.4%	14.4%	0.0%	0.0%	0.0%	0.0%	0.0%
4	14.3%	14.2%	14.3%	14.2%	14.3%	0.0%	14.4%	14.4%	0.0%	0.0%	0.0%	0.0%	0.0%
5	14.4%	14.3%	14.3%	14.0%	14.2%	0.0%	14.4%	14.4%	0.0%	0.0%	0.0%	0.0%	0.0%
6	14.4%	14.3%	14.3%	14.0%	14.1%	0.0%	14.4%	14.4%	0.0%	0.0%	0.0%	0.0%	0.0%
7	14.4%	14.3%	14.3%	14.0%	14.1%	0.0%	14.4%	14.4%	0.0%	0.0%	0.0%	0.0%	0.0%
8	20.0%	19.9%	19.9%	0.0%	0.0%	0.0%	20.1%	20.1%	0.0%	0.0%	0.0%	0.0%	0.0%
9	20.0%	19.9%	20.0%	0.0%	0.0%	0.0%	20.1%	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10	20.0%	19.9%	20.0%	0.0%	0.0%	0.0%	20.0%	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Tabelle 8: Motiv-Entropien eines skalenfreien Graphen

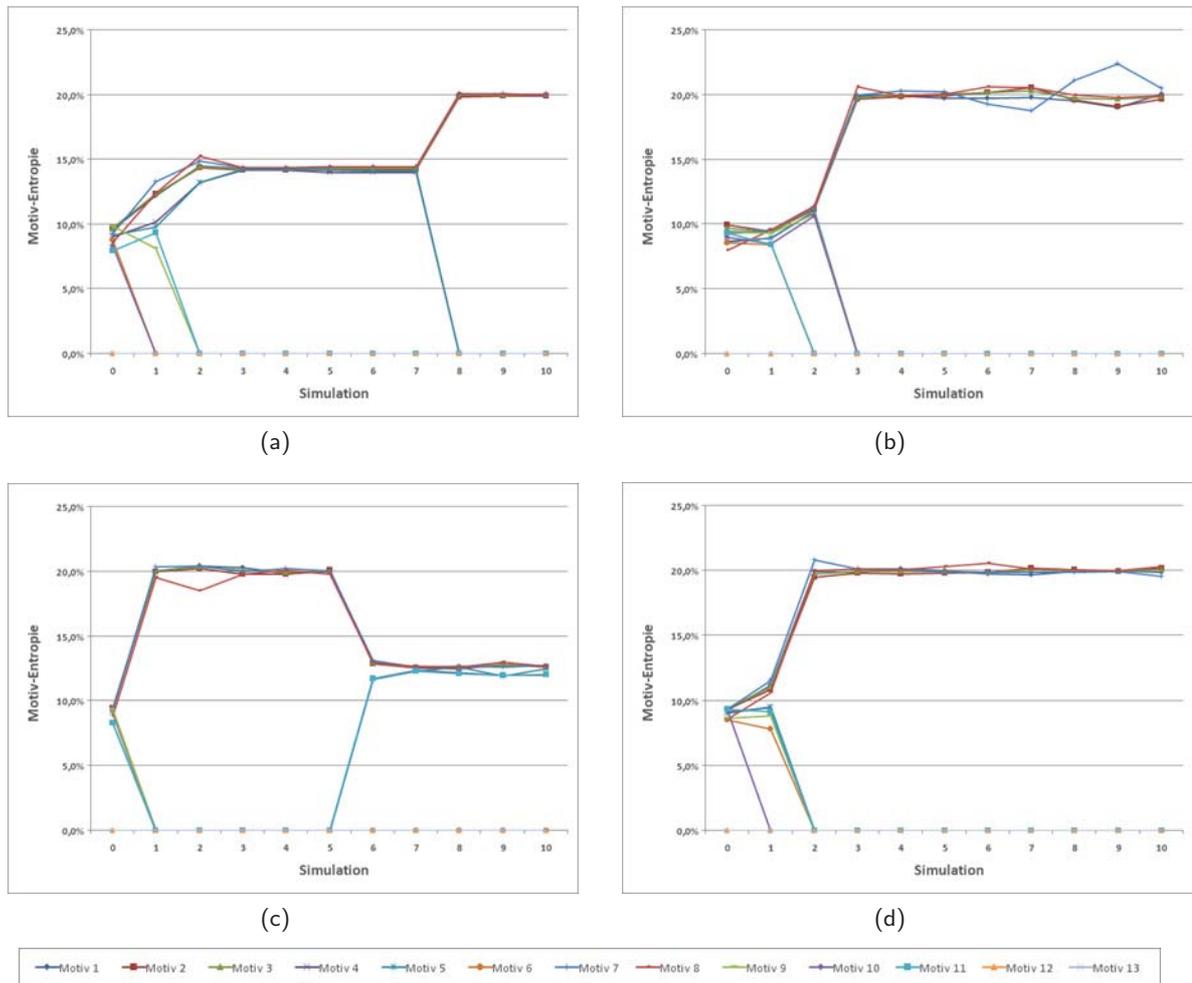


Abbildung 39: Motiv-Entropien in verschiedenen skalenfreien Netzwerken (1)

Zunächst ist festzustellen, dass eine Motivklasse, die im Graphen nicht vorkommt, auch keine Motiv-Entropie hat, was nach Definition sofort ersichtlich ist. Die Werte, deren Motiv-Entropie größer Null ist, sind darüber hinaus innerhalb einer Simulation meist gleich groß, was bedeutet, dass jede Motivklasse in etwa die gleiche Shannon-Entropie bildet. Interessant sind also Motiv-Entropien, die von den übrigen Werten abweichen. In einem solchen Fall bildet nämlich eine Motivklasse eine signifikant höhere oder niedrigere Entropie als die übrigen Klassen. Um diese Abweichungen besser identifizieren zu können, zeigt Abbildung 39 die Motiv-Entropien einiger Experimente, unter anderem auch die Werte aus Tabelle 8 in Bild (a).

In Bild (a) ist eine Abweichung nach den Variationsschritten 1 und 2 zu erkennen. Die Entropiewerte reichen hier von 8.1 (Motiv 9) bis 13.3 (Motiv 7). Letzteres bildet also eine höhere Shannon-Entropie aus, als Motiv 9. In Bild (b) zeigt sich eine erhöhte Motiv-Entropie ebenfalls in der Motivklasse 7, während Bild (c) einen verminderten Motiv-Entropiewert in Klasse 8 nach der zweiten Variation zeigt. Bild (d) zeigt lediglich zu Beginn eine geringfügige Streuung, nach der zweiten Variation leistet aber kein Motiv einen besonderen Beitrag zur Shannon-Entropie.

Die Diagramme in Abbildung 40 enthalten beide größere Abweichungen der einzelnen Motiv-Entropien, die einzelnen Klassen bilden damit völlig verschiedene Shannon-Entropien. Motiv 7 zeigt auch hier einen hohen Motiv-Entropiewert, ein eindeutiges Minimum ist allerdings nicht erkennbar: In (e) bildet Motiv 2 und teilweise Motiv 8 eine niedrige Motiv-Entropie, während in (f) eher Motiv 4 verminderte Entropiewerte bildet. Das Motiv 8 hat hier deutlich höhere Motiv-Entropiewerte als in (e).

Abbildung 41 enthält die Entropieebenen aus Shannon- und Word-Entropie der sechs Experimente. Die großen Abweichungen aus den Bildern (e) und (f) führen zu niedrigen Entropiewerten und zu keiner erkennbaren Trajektorie, während die anderen vier Experimente eine deutliche Trajektorie zeigen.

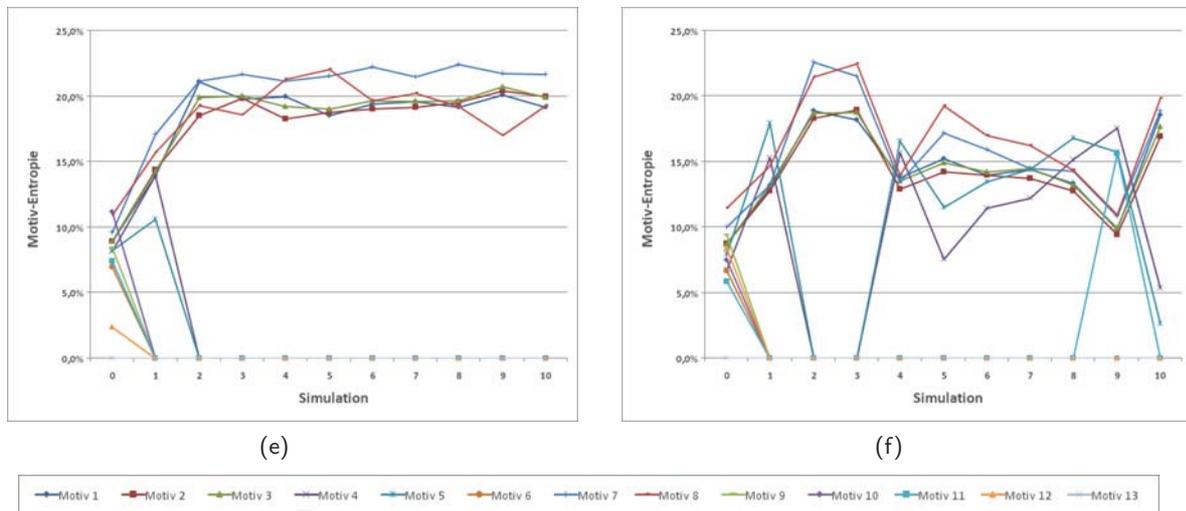


Abbildung 40: Motiv-Entropien in verschiedenen skaleneutralen Netzwerken (2)

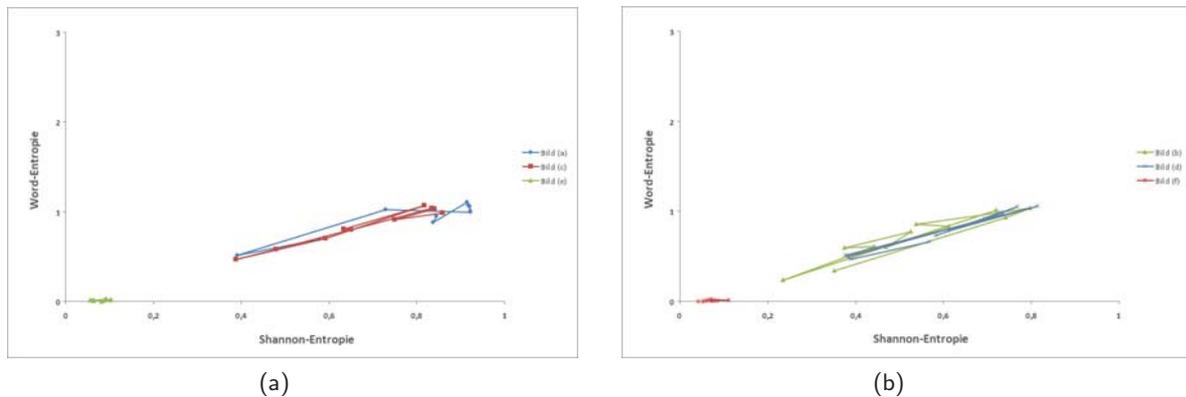


Abbildung 41: Entropieebene der sechs Experimente aus den Abbildungen 39 und 40

Die Experimente zeigen häufig eine hohe Motiv-Entropie für die Motivklasse 7. In der neunten Simulation von (b) zeigt sich sogar ein deutlicher Ausschlag. Der Wert der Motiv-Entropie dieses Experiments wird nun der zahlenmäßigen Shannon- und Word-Entropie des gesamten Netzwerks aus (b) gegenübergestellt. Tabelle 9 listet diese auf.

Simulation	$\mathcal{S}$	$\mathcal{W}$	$\mathcal{S}_7^0$	$\min \mathcal{S}_K^0$	$\max \mathcal{S}_K^0$	$\text{avg } \mathcal{S}_K^0$	$\mathcal{S}_7^0 - \text{avg } \mathcal{S}_K^0$
0	0.352	0.345	9.3%	8.0%	10.0%	9.1%	+0.2%
1	0.741	0.933	9.6%	8.4%	9.6%	9.1%	+0.5%
2	0.797	1.038	11.4%	10.6%	11.5%	11.1%	+0.2%
3	0.538	0.862	20.0%	19.7%	20.6%	20.0%	+0.0%
4	0.611	0.841	20.3%	19.9%	20.3%	20.0%	+0.3%
5	0.468	0.603	20.2%	19.8%	20.2%	20.0%	+0.2%
6	0.526	0.780	19.3%	19.3%	20.6%	20.0%	-0.7%
7	0.374	0.601	18.8%	18.8%	20.6%	20.0%	-1.2%
8	0.442	0.614	21.1%	19.5%	21.1%	20.0%	+1.1%
9	0.234	0.240	22.4%	19.1%	22.4%	20.0%	+2.4%
10	0.719	1.015	20.5%	19.7%	20.5%	20.0%	+0.5%

Tabelle 9: Motiv-Entropie der Klasse 7 im Vergleich zu anderen Entropie-Kennzahlen

Dabei steht  $\min \mathcal{S}_K^0$  für das Minimum der Motiv-Entropie über alle Klassen  $K \in \mathcal{K}_k$  deren Motivgehalt größer Null ist,  $\max \mathcal{S}_K^0$  für das entsprechende Maximum der Motiv-Entropie und  $\text{avg } \mathcal{S}_K^0$  für das arithmetische Mittel der Motiv-Entropien größer Null. Die letzte Spalte der Tabelle zeigt den Abstand von Motiv 7 zum Mittelwert der Motiv-Entropien.

Die größte Abweichung in Simulation 9, bei der  $\mathcal{S}_7^0$  den Maximalwert von 22.4 annimmt, führt zu einer geringen Shannon- und Word-Entropie. In Simulation 7 nimmt  $\mathcal{S}_7^0$  dann das Minimum  $\min \mathcal{S}_K^0$  an, was interessanterweise ebenfalls zu einer recht geringen Entropie führt. Die Umkehrung, dass die Motivklasse 7 eine größere Gesamtentropie induziert, falls  $\mathcal{S}_7^0$  nahe dem Mittelwert liegt, lässt sich nicht zeigen.

Um diese Beobachtung besser verstehen zu können, wird die Struktur der Motivklasse 7 näher betrachtet. Abbildung 42 zeigt das Motiv in seiner graphischen Repräsentation. Das besondere an dieser Motivklasse ist, dass es eine „Abkürzung“ von dem oberen Knoten zum Knoten rechts bietet. Eine solche Konstellation von Kanten wird in der Literatur als transitives Triplett bezeichnet, da aus  $(a, b)$  und  $(b, c)$  der Weg  $(a, c)$  folgt.

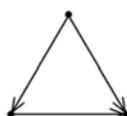


Abbildung 42: Motivklasse 7 (Feed-forward loop)

In der Biologie ist dieses Motiv bereits bekannt dafür, die Signalverarbeitung in metabolischen Netzwerken zu beschleunigen [Aln04, Aln06]. Die Beobachtung hoher Motiv-Entropiewerte der Klasse 7 ist also durchaus sinnvoll, da eine schnelle Signalverarbeitung zu einer hohen Entropie führt. Tatsächlich bewegt sich die nicht normierte Motiv-Entropie, also die Shannon-Entropie aller Knoten im Motiv bei geeignetem  $\kappa$ , im Bereich von 0.430 bis 0.983 mit Mittelwert 0.768 durchaus im oberen Bereich.

Die Vermutung liegt nahe, dass eine große Anzahl von Motiven der Klasse 7 zu einer hohen Entropie führt. Diese Behauptung ist so jedoch nicht haltbar. In einem Experiment wurden die Motive der Klasse 7 um über 200% erhöht, trotzdem ergab sich ausgehend von den Entropien  $(\mathcal{S}, \mathcal{W}) = (0.456, 0.589)$  nach der Variation nur eine Entropie von  $(\mathcal{S}, \mathcal{W}) = (0.300, 0.342)$ . Unter Beachtung der in Abschnitt 4.2 angesprochenen Probleme bei der Variation, ist einsehbar, dass die Herstellung der Ziel motive viele andere Motive transformiert hat, was die Entropien natürlich nachhaltig beeinflusst. Außerdem ist anzunehmen, dass die Entropie von dem am häufigsten auftretenden Motiv dominiert wird, bei skalenfreien Graphen ist

das das zweikantige Motiv 3. Dennoch ist ein gewisser Einfluss des Motivs 7 unter obiger Argumentation sinnvoll, wenn auch vermutlich noch an diverse Nebenbedingungen gebunden.

Abschließend werden die Beobachtungen bezüglich der Motiv-Entropie und des Feed-forward loops noch einmal zusammengefasst:

**Beobachtung 8** Große Unterschiede in den Motiv-Entropien der einzelnen Motivklassen führten zu einer niedrigen Gesamtentropie.

**Beobachtung 9** In vielen Experimenten konnte eine hohe, oftmals maximale, Motiv-Entropie für das Motiv 7 (Feed-forward loop) beobachtet werden. Diese Klasse bildete meist eine hohe Shannon-Entropie, im Mittel bei ca. 0.768.

**Beobachtung 10** Ein direkter (erhöhender) Einfluss der Motivklasse 7 auf die Shannon- bzw. Word-Entropie erscheint sinnvoll, konnte jedoch in den Untersuchungen so nicht nachgewiesen werden.

Die obigen Beobachtungen konnten in gleicher Weise in Erdős-Rényi-Graphen gemacht werden. Auch hier zeigten große Unterschiede in der Motiv-Entropie eine niedrige Gesamtentropie. Abbildung 43 gibt ein Beispiel für eine Motiv-Entropie in einem ER-Graphen mit  $p = 0.18$ . Die Abbildung zeigt oben links wieder die Motiv-Entropien der einzelnen Klassen pro Simulation. Da bereits nach dem ersten Variationsschritt alle Motive der Klassen 4-6 und 9-13 zerstört sind, beschränkt sich das Diagramm auf

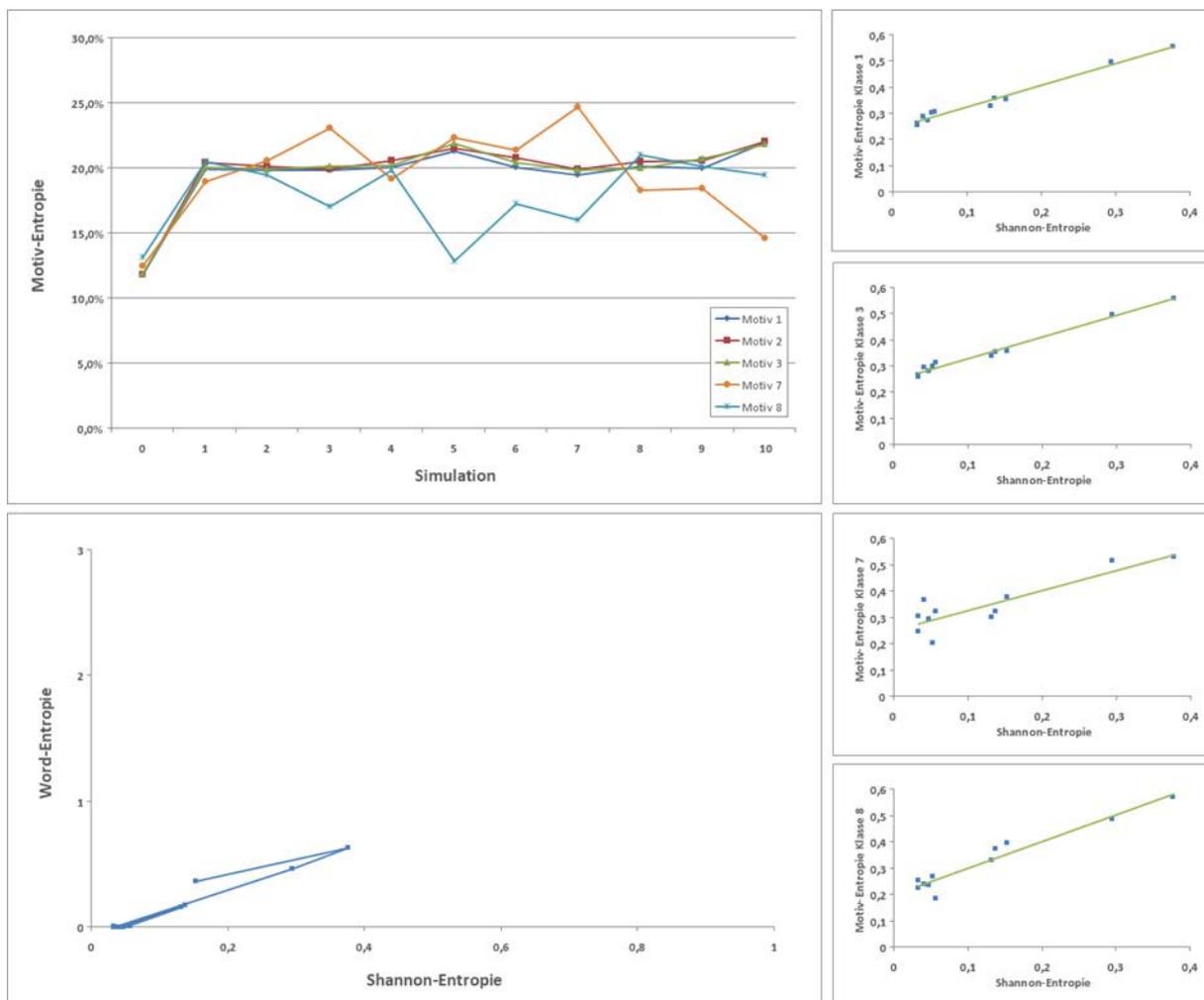


Abbildung 43: Motiv-Entropie in Erdős-Rényi-Graphen

die verbleibenden Motivklassen. Man sieht, dass die Klassen 1-3 einer einheitlichen Linie folgen und stets etwa gleich groß sind. Die Motiv-Entropie der Klasse 8 ist meist niedriger als die der übrigen, während Klasse 7 mit Ausnahme der letzten beiden Schritte höher ist.

Große Abweichungen sind vor allem in den Schritten 4, 6-8 und 11 zu beobachten. Tabelle 10 gibt die Werte der Gesamtentropie an, welche wie erwartet bei diesen Simulationen einen niedrigen Wert enthält. In Simulation 5 würde man nach obigen Beobachtungen nicht unbedingt eine niedrige Entropie erwarten, da die Motiv-Entropien etwa gleich groß sind. Dennoch sind die Entropien hier ähnlich niedrig wie in den anderen genannten Simulationen. Ein Maximum der Gesamtentropie ergibt sich in Simulation 1, die Motiv-Entropien sind hier etwa gleich groß. Das Diagramm unten links zeigt die Entropieebene und die darin enthaltene Trajektorie.

Simulation	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_7$	$\mathcal{S}_8$	$\mathcal{S}$	$\mathcal{W}$
0	0.356	0.357	0.359	0.376	0.396	0.352	0.345
1	0.555	0.570	0.560	0.530	0.572	0.741	0.933
2	0.498	0.505	0.499	0.517	0.488	0.797	1.038
3	0.262	0.264	0.267	0.305	0.225	0.538	0.862
4	0.258	0.264	0.259	0.246	0.254	0.611	0.841
5	0.308	0.311	0.316	0.323	0.186	0.468	0.603
6	0.276	0.286	0.281	0.294	0.237	0.526	0.780
7	0.290	0.297	0.296	0.369	0.239	0.374	0.601
8	0.358	0.364	0.356	0.324	0.373	0.442	0.614
9	0.328	0.338	0.340	0.303	0.331	0.234	0.240
10	0.303	0.304	0.301	0.202	0.269	0.719	1.015
$r_{\mathcal{S},\mathcal{S}_i}$	0.9897	0.991	0.991	0.879	0.968		

Tabelle 10: Kennzahlen des Experiments aus Abbildung 43

Die Diagramme auf der rechten Seite zeigen die Abhängigkeit von nicht normierter Motiv-Entropie (also die Shannon-Entropie über allen Knoten, die in einer Motivklasse beteiligt sind) zur Gesamt-Shannon-Entropie. Es zeigt sich eine lineare Abhängigkeit, die für die Klassen 1-3 besonders gut ist. Auch für Motiv 8 beträgt der Korrelationskoeffizient noch 0.96. Für Motivklasse 7 ist lediglich 0.879 zu beobachten, was erneut für die Sonderstellung dieser Motivklasse spricht. Die Korrelationskoeffizienten  $r_{\mathcal{S},\mathcal{S}_i}$  sind ebenfalls in Tabelle 10 enthalten.

**Beobachtung 11** *Das Verhalten der Motiv-Entropie von skalenfreien Graphen konnte auch in Erdős-Rényi-Graphen beobachtet werden.*

**Beobachtung 12** *Zwischen Gesamt-Shannon-Entropie und nicht-normierter Motiv-Entropie besteht ein linearer Zusammenhang. Gemessen wurden Korrelationskoeffizienten von 0.968 bis 0.991, wobei die Motivklasse 7 hier wieder eine Sonderstellung einnimmt mit nur 0.879.*



## 6 Fazit

Es wurden gerichtete  $2k$ -reguläre Kreisgraphen, Erdős-Rényi-Graphen und skalenfreie Netzwerke auf ihren Motivgehalt untersucht. Anschließend wurden die Motive in ihrer Anzahl variiert, um eine vorgegebene Motivsignatur zu erzeugen. Nach jedem Variationsschritt wurde anhand der Shannon- und Word-Entropie der Informationsgehalt des Netzwerkes ausgewertet.

Es hat sich gezeigt, dass sich nicht jede Änderung der Motivverteilung unmittelbar in den Entropien niederschlägt. Vielmehr ergab sich das die zur Auswertung verwendete Transitionsfunktion und das Aufbauschema des Graphen eine große Rolle spielen. So konnten lange Trajektorien bezüglich der Transition  $\Omega^{(2)}$  beobachtet werden, während die Transition  $\Omega^{(1)}$  keine Bewegung zeigte.

Einen besonderen Einfluss schien auch die Zyklenzahl des Graphen zu haben. Die Trajektorien azyklischer Erdős-Rényi-Graphen zeigte beispielsweise in eine einheitliche Richtung, während nicht-zyklenfreie Erdős-Rényi-Graphen mit gleichen Simulationsbedingungen meist gar keine eindeutige Trajektorie ausbildeten.

Für skalenfreie Graphen wurde die Shannon-Entropie lokal für die einzelnen Motivklassen als so genannte Motiv-Entropie definiert. Hierbei zeigte sich, dass sehr unterschiedliche Motiv-Entropien zu einer niedrigen Gesamtentropie führen. Außerdem wurde die Motivklasse 7 eingehend untersucht, da diese meist eine recht hohe Motiv-Entropie bildet. Ein direkter Einfluss auf die Gesamtentropie erschien sinnvoll, konnte aber nicht nachgewiesen werden.

Ausgehend von den Untersuchungen und Beobachtungen in Kapitel 5 sind viele weitere Forschungen denkbar. Interessant wäre hier sicherlich das Verhalten azyklischer Graphen, welchen in jedem Variationsschritt sukzessive neue Zyklen zugefügt werden, näher zu betrachten. Auch ist eine Untersuchung denkbar, die sich mit Motiven *verschiedener* Größen beschäftigt und diese beispielsweise nach der Anzahl der enthaltenen Zyklen auswertet.

Die Betrachtung der Motiv-Entropie kann sicherlich auch noch weitergeführt werden, hier wäre eine Unterscheidung nach Shannon-Motiv-Entropie und Word-Motiv-Entropie möglich. Ein Variationsalgorithmus könnte an die beiden lokalen Entropien gebunden werden, welcher versucht Motive mit geringem (oder hohem) Entropiewert zu zerstören.

Die Analyse der Korrelation von Gesamtentropie und Motiv-Entropie lässt darüber hinaus noch Raum für weitere Beobachtungen. Es kann hier auch die Korrelation von Word-Entropie zu Word-Motiv-Entropie gemessen und untersucht werden.

Interessant ist natürlich auch die Untersuchung anderer Graphenklassen, insbesondere scheint eine Untersuchung der von der ALON-Gruppe gefundenen Superfamilien [Aln04] sinnvoll.



## Abbildungsverzeichnis

1	Beispiel eines Graphen . . . . .	6
2	4-regulärer Kreisgraph mit 10 Knoten . . . . .	6
3	Erdős-Rényi-Graph mit 15 Knoten . . . . .	7
4	Skalenfreier Graph mit 27 Knoten . . . . .	8
5	Beispiel eines Automaten . . . . .	9
6	Conways „Game of Life“ . . . . .	11
7	Transitionsfunktion im „Game of Life“ . . . . .	11
8	Beispiel eines raumzeitlichen Musters . . . . .	12
9	Beispiel eines raumzeitlichen Musters . . . . .	13
10	Beispiele für die Klassifikation eines raumzeitlichen Musters . . . . .	14
11	Beispiel für Subgraphen und Graphenisomorphie . . . . .	15
12	Motivklassen der Größe 3 in gerichteten Graphen . . . . .	18
13	Motivklassen der Größe 4 in ungerichteten Graphen . . . . .	18
14	Superfamilien in gerichteten Graphen [Aln04] . . . . .	19
15	Motivverteilung und normierter Z-Score in gerichteten $2k$ -regulären Kreisgraphen . . . . .	20
16	Motivverteilung und normierter Z-Score in gerichteten ER-Graphen (1) . . . . .	21
17	Motivverteilung und normierter Z-Score in gerichteten ER-Graphen (2) . . . . .	22
18	Motivverteilung und normierter Z-Score in gerichteten skalenfreien Netzwerken . . . . .	24
19	Verteilung in der Entropieebene [Mar05] . . . . .	25
20	Trajektorie in der Entropieebene [Mar05] . . . . .	25
21	Nebenwirkungen bei der Motivvariation . . . . .	26
22	Topologische Kenngrößen . . . . .	27
23	Variationsalgorithmus zur Beibehaltung der Gradverteilung . . . . .	28
24	Variationsalgorithmus zur Beibehaltung der Gradverteilung . . . . .	30
25	Beispiel eines Kopfes in einem Testprotokoll . . . . .	31
26	Auswirkungen der Variation im $2 \cdot 2$ -regulären Kreisgraphen mit 200 Knoten . . . . .	33
27	$2 \cdot 2$ -regulärer Kreisgraphen mit 20 Knoten vor und nach der Variation . . . . .	33
28	Testprotokoll $2 \cdot 2$ -regulärer Kreisgraphen mit 200 Knoten . . . . .	34
29	Testprotokoll $2 \cdot 4$ -regulärer Kreisgraphen mit 200 Knoten . . . . .	35
30	Testprotokoll Erdős-Rényi-Graph mit $p = 0.02$ und 200 Knoten . . . . .	37
31	Entropiewerte der beiden Transitionsfunktionen $\Omega$ mit verschiedenen Parametern $\kappa$ . . . . .	38
32	Testprotokoll Erdős-Rényi-Graph mit Transition $\Omega^{(2)}$ . . . . .	39
33	Motivverteilung und normierter Z-Score in gerichteten azyklischen ER-Graphen . . . . .	40
34	Trajektorien in azyklischen Erdős-Rényi-Graphen . . . . .	41
35	Testprotokoll azyklischer Erdős-Rényi-Graph mit 200 Knoten . . . . .	42
36	Verschiedene Trajektorien für $\Omega^{(2)}$ mit Parameter $\kappa = 0.3$ . . . . .	43
37	Entropiewerte der beiden Transitionsfunktionen $\Omega$ mit verschiedenen Parametern $\kappa$ . . . . .	44
38	Testprotokoll skalenfreier Graph mit 200 Knoten . . . . .	46
39	Motiv-Entropien in verschiedenen skalenfreien Netzwerken (1) . . . . .	47
40	Motiv-Entropien in verschiedenen skalenfreien Netzwerken (2) . . . . .	48
41	Entropieebene der sechs Experimente aus den Abbildungen 39 und 40 . . . . .	48
42	Motivklasse 7 (Feed-forward loop) . . . . .	49
43	Motiv-Entropie in Erdős-Rényi-Graphen . . . . .	50

## Tabellenverzeichnis

1	Motivverteilung und normalisierte Z-Scores in gerichteten $2k$ -regulären Kreisgraphen . . . . .	20
2	Durchschnittliche Motivverteilungen und normalisierte Z-Scores in gerichteten ER-Graphen . . . . .	21
3	Verschiedene normierte Z-Scores des ersten Graphen mit $p = 0.18$ . . . . .	23
4	Durchschnittliche Motivverteilungen und normalisierte Z-Scores in gerichteten BA-Graphen . . . . .	24
5	Beispiel einer Ratingberechnung . . . . .	29
6	Entropiewerte von 200 $2k$ -regulären Kreisgraphen mit Transition $\Omega^{(1)}$ und $\kappa = 0.4$ . . . . .	32
7	Typische Entropiewerte in zyklischen und nicht zyklischen Erdős-Rényi-Graphen . . . . .	41
8	Motiv-Entropien eines skalenfreien Graphen . . . . .	47
9	Motiv-Entropie der Klasse 7 im Vergleich zu anderen Entropie-Kennzahlen . . . . .	49
10	Kennzahlen des Experiments aus Abbildung 43 . . . . .	51

## Literaturverzeichnis

- [Aln02] Alon, Uri; Milo, Ron; Itzkovitz, Shalev et. al.:  
*Network Motifs: Simple Building Blocks of Complex Networks*,  
Science Magazine, Ausgabe 298, 2002, Seite 1538-1542.
- [Aln04] Alon, Uri; Milo, Ron; Itzkovitz, Shalev et. al.:  
*Superfamilies of designed and evolved networks*,  
Science Magazine, Ausgabe 303, 2004, Seite 824-827.
- [Aln06] Alon, Uri; Mangan, Shmoolik; Itzkovitz, Shalev et. al.: *The Incoherent Feed-forward Loop Accelerates the Response-time of the gal System of Escherichia coli*,  
JMB, Ausgabe 356, 2006, Seite 1073-1081
- [Bar99] Barabási, Albert-László und Albert, Réka: *Emergence of scaling in random networks*,  
Science Magazine, Ausgabe 286, 1999, Seite 509-512.
- [Bar04] Barabasi, Albert-Laszlo und Oltvai, Zoltan N.:  
*Network biology: Understanding the cell's functional organization*,  
Nature Reviews Genetics, Ausgabe 2/5, Februar 2004, Seite 101-113.
- [Cay57] Cayley, Arthur: *On the theory of the analytical forms called trees*,  
Philos. Mag., Ausgabe 13/1857, Seite 172-176.
- [Con70] Conway, John Horton und Gardner, Martin:  
*The Fantastic Combinations of John Conway's New Solitaire Game of Life*,  
Scientific American, Ausgabe 223, 1970, Seite 120-123.
- [EM06] Ebert, Sandra und Meyer, Christian M.:  
MOTIFCONTROL – *Ein Werkzeug zur Analyse und Variation von Motiven in Graphen*,  
Technische Universität Darmstadt, Fachbereich Informatik, Fachgebiet Algorithmik, 2006.
- [Erd59] Erdős, Paul und Rényi, Alfréd: *On Random Graphs I*,  
Publicationes Mathematicae, Ausgabe 6, 1959, Seite 290-297.
- [Eul36] Euler, Leonhard: *Solutio Problematis ad Geometriam situs Pertinentis*,  
Comment. Acad. Sci. I. Petropolitanae, Ausgabe 8/1736, Seite 128-140.
- [Hüt01] Hütt, Marc-Thorsten: *Datenanalyse in der Biologie*,  
Berlin/Heidelberg/New York, Springer, 2001, ISBN: 3-540-42311-7.
- [Kir47] Kirchhoff, Gustav B.: *Über die Auflösung der Gleichungen auf welche man bei den Untersuchungen der linearen Verteilung galvanischer Ströme geführt wird*,  
Poggend. Ann. Physik, Ausgabe 72/1847, Seite 497-508.
- [Mar05] Marr, Carsten und Hütt, Marc-Thorsten:  
*Topology regulates pattern formation capacity of binary cellular automata on graphs*,  
Physica A: Statistical Mechanics and its Applications, Ausgabe 354,  
15. August 2005, Seite 641-662.
- [Otto04] Otto, Martin: *Automaten und Formale Sprachen*,  
Technische Universität Darmstadt, Fachbereich Mathematik, AG1: Algebra und Logik, 2004.
- [Pre97] Preuß, Wolfgang und Wenisch, Günter [Hrsg.]:  
*Lehr- und Übungsbuch Mathematik für Informatiker*,  
Fachbuchverlag Leipzig im Carl Hanser Verlag, München/Wien, 1997.
- [Rav02] Ravasz, Erzsébet; Somera, A. L.; Mongru, D. A. et al.:  
*Hierarchical Organization of Modularity in Metabolic Networks*,  
Science Magazine, Ausgabe 297, 2002, Seite 1551-1555.

- [Sch97] Schöning, Uwe: *Theoretische Informatik – kurzgefaßt*,  
3. Auflage, Spektrum Akademischer Verlag, Heidelberg/Berlin, 1999, ISBN: 3-8274-0250-6.
- [Sha48] Shannon, Claude Elwood: *A mathematical theory of communication.*,  
Bell Systems Technical Journal, Ausgabe 27, 1948, Seite 379-423.
- [Wal03] Waldschmidt, Helmut und Guntermann, Klaus:  
*Nichtnumerische Algorithmen und Datenstrukturen*,  
Technische Universität Darmstadt, Fachbereich Informatik, Fachgebiet Systemprogrammierung, 3. verbesserte Auflage, 2003.
- [Wik06] Wikipedia, die freie Enzyklopädie: *Entropie*,  
<http://de.wikipedia.org/wiki/Entropie>, Stand: 02.04.2006.
- [Wil75] Wilson, Robin J.: *Introduction to graph theory*,  
Longman Group Ltd., London, 1975.
- [Wol84] Wolfram, Stephen: *Universality and complexity in cellular automata*,  
Physica D: Nonlinear Phenomena, Band 10, Ausgabe 1-2,  
Januar 1984, Seite 1-35.
- [Wol86] Wolfram, Stephen: *Theory and applications of cellular automata*,  
Singapore, World Scientific Publishing, 1986, ISBN: 9971-50-123-6.