

Vorlesung SS 2009

Prof. Dr. Ulrich Schwarz

Physik Biologischer Netzwerke

Abteilung für Theoretische Biophysik

Universität Karlsruhe

Homepage: <http://www.zi.uni-karlsruhe.de/tbp>

Inhaltsverzeichnis

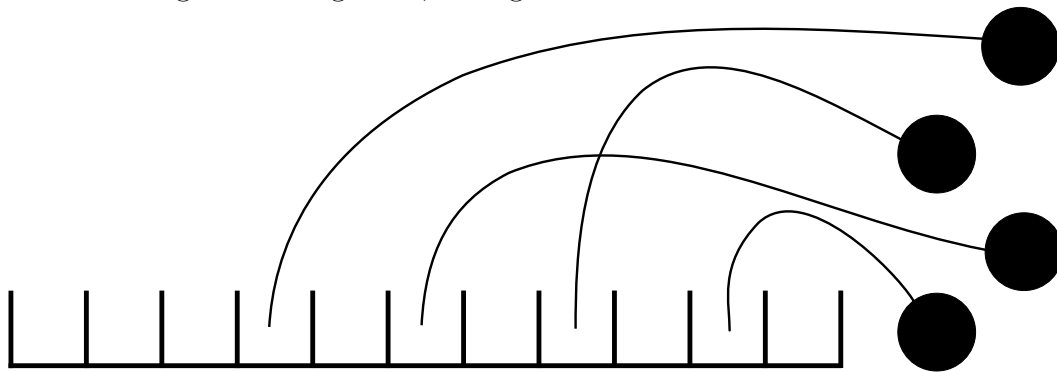
1	Zufallsgraphen	1
1.1	Vorbemerkung: Wahrscheinlichkeitsverteilungen	1
1.2	Definition: Graph	2
1.3	Klassische Zufallsgraphentheorie	3
1.4	Maße für Graphen - reale Netzwerke	4
1.5	Unterschiede zwischen Netzwerken und Gittern	7
1.6	Maße für Graphen - Erdős-Rényi-Netzwerk	9
2	Perkolation	11
2.1	Physikalische Anwendungen	12
2.2	Perkolation in 1D	13
2.3	Zusammenfassung (1D)	14
2.4	Bethe-Gitter	15
2.5	Mittlere Cluster-Größe S	16
2.6	Stärke des Netzwerks	17
2.7	Schlussfolgerung	17
3	Verallgemeinerte Zufallsgraphen	19
3.1	Zufallsgraphen mit beliebiger Grad-Verteilung	19
3.2	Methode der erzeugenden Funktionen	21
3.3	Exponentieller Graph	23
3.4	Potenzverteilung mit Cutoff	23
3.5	Empirische Gradverteilung	25
3.6	Clustergrößenverteilung	25
3.7	Gradverteilung eines Graphens mit Potenzverteilung (Exponent τ)	30
4	Small-World Netzwerke	32
4.1	Watts-Strogatz Modell	32

4.2	Cluster-Koeffizient	36
4.3	Gradverteilung im Watts-Strogatz Modell	37
4.4	Durchschnittliche Verbindungslänge l	38
5	Wachsende Netzwerke	40
5.1	Barabasi und Albert Modell	40
5.2	Exakte Ereignisse für Das BA-Modell	41
5.3	Gradverteilung	42
5.3.1	Modell A	43
5.3.2	Modell B	43
5.3.3	Zusammenfassung	44
5.4	Mastergleichungen	45
5.4.1	Gleichgewichtsnetzwerke	45
5.4.2	Wachsende exponentielle Netzwerke	47
5.5	Barabasi Albert Modell	48
6	Biologische Netzwerke	51

1 Zufallsgraphen

1.1 Vorbemerkung: Wahrscheinlichkeitsverteilungen

Wieviele Möglichkeiten gibt es, k Kugeln auf N Fächer zu verteilen?



Dieser Wert ist durch den Binomialkoeffizienten gegeben:

$$\binom{N}{k} = N(N-1)\dots(N-(k+1))\frac{1}{k!} = \frac{N!}{(N-k)!k!}$$

Wir werfen k -mal eine Münze. Die Wahrscheinlichkeit für Kopf nennen wir p . Wie groß ist die Wahrscheinlichkeit für k -mal Kopf? (Binomialverteilung)

$$B(k) = \binom{N}{k} p^k (1-p)^{N-k}$$

Der Grenzfall $N \rightarrow \infty$ und $p \rightarrow 0$, mit $pN = \text{const.} = \langle k \rangle = \lambda$, ergibt die Poisson-Verteilung:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

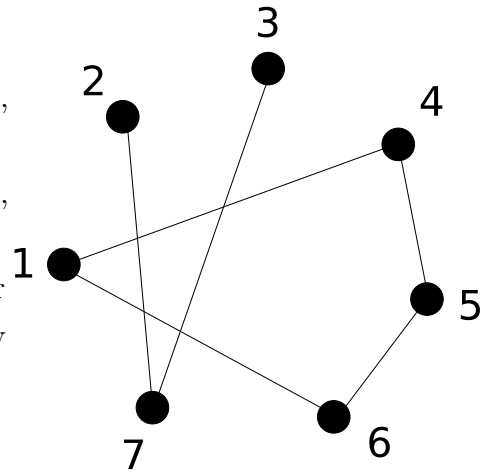
1.2 Definition: Graph

Ein Graph ist ein Paar von Mengen $G = \{P, E\}$.

P ist die Menge von N Knoten (oder auch nodes, vertices, points).

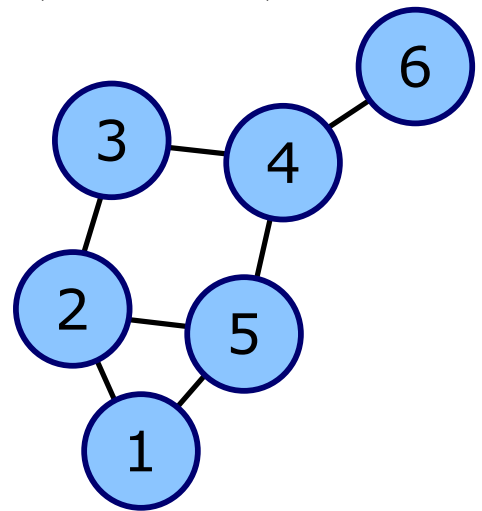
E ist die Menge von n Kanten (oder auch edges, links, lines).

Jeder Graph kann durch eine "Kontaktmatrix" der Dimension $N \times N$ repräsentiert werden. ("adjacency matrix")

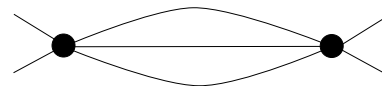
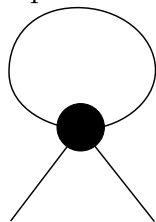


Eine solche Kontaktmatrix ist links unten dargestellt. Die Einträge können entweder Null oder Eins sein, da wir keine Tadpoles oder Melons (Def. siehe unten) zulassen.

$$a = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$



Folgende Graphenelemente werden nicht betrachtet:



Grund: Im Limes $N \rightarrow \infty$ sind Tadpoles(links) und Melons(rechts) irrelevant. Des weiteren werden wir nur ungerichtete Graphen behandeln, d.h. die Kontaktmatrix ist immer symmetrisch

1.3 Klassische Zufallsgraphentheorie

(Erdős und Rényi 1959)

Ergänzende Literatur: "Random Graphs" Bollobás 1985

2 Varianten:

1. "mikrokanonische" Gesamtheit $G(N, n)$

N Knoten werden mit n Kanten verbunden. Die Kanten verbinden zufällig ausgesuchte Knotenpaare.

Die Zahl der maximal möglichen Kanten ist:

$$n_t = \frac{N(N-1)}{2} = \binom{N}{2}$$

Jeder der möglichen Graphen hat die gleiche Wahrscheinlichkeit realisiert zu werden.

Diese ist:

$$p = \frac{1}{\binom{n_t}{n}}$$

2. "kanonische" Gesamtheit $G(N, p)$

Für jede der n_t möglichen Kanten wird mit der Wahrscheinlichkeit p erstellt.

Damit ist die Wahrscheinlichkeit für einen Graphen:

$$\sum \binom{n_t}{n} p^n (1-p)^{n_t-n}$$

Bei festem N kann man jetzt p variieren ("Evolution")

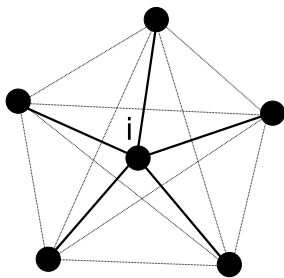
Bsp: $N=10$ und $n_t = 45$

2. Clusterkoeffizient

Der Grad k_i des Knotens ist die Zahl der mit dem Knoten verbundenen Kanten. (In dem Beispiel links ist $k_i = 5$.)

Die maximale Zahl der Kanten in der Nachbarschaft ist:

$$E_i^{max} = \frac{k_i(k_i - 1)}{2}$$



Hier ist E_i eine Messgröße (Zahl der grauen Links). Der lokale Clusterkoeffizient ist nun definiert durch:

$$C_i = \frac{E_i}{E_i^{max}} \quad 0 \leq C_i \leq 1$$

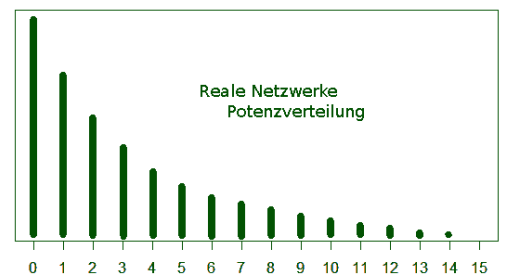
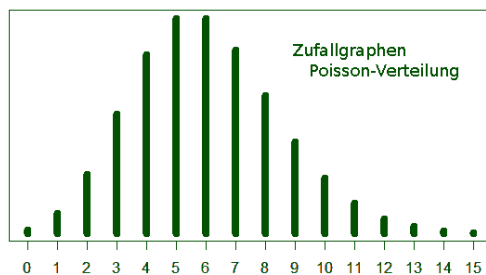
Der globale Clusterkoeffizient für einen gesamten Graphen ist:

$$C = \langle C_i \rangle = \frac{1}{N} \sum_i C_i$$

3. Gradverteilung

Die Gradverteilung entspricht der Funktion $p(k)$. Sie gibt die Wahrscheinlichkeit an einen Knoten des Grades k im Netzwerk zu haben.

Die folgenden Abbildungen zeigen zwei Beispiele für Gradverteilungen:



4. Häufigkeit von Subgraphen

5. Spektrum des Graphen

Eigenwerte λ_i der Kontaktmatrix

Spektraldichte:

$$\rho = \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i)$$

1 Zufallsgraphen

- Für das kanonische Ensemble $G(N, p)$ gilt:

$$\langle n \rangle = p \frac{N(N-1)}{2} \quad \langle k \rangle = p(N-1) \approx pN \quad C = p$$

Vergleich mit realen Netzwerken:

Netzwerk	N	$\langle k \rangle = pN$	l	l_{rand}^1	C	C_{rand}
www	150.000	35	3,1	3,35	0,1	0,0002
Schauspieler	225.000	61	37	30	0,79	0,0003
Physiker (LANL)	52.000	9,7	5,9	4,79	0,43	$2 \cdot 10^{-4}$
Biomed (MEDLINE)	1.520.000	18,1	4,6	4,9	0,06	$1 \cdot 10^{-5}$
Hochenergiephysik (SPIRES)	56.000	173	4,0	2,21	0,7	0,003
	$p = \frac{\langle k \rangle}{N} \ll 1$		ähnlich		starke Unterscheidung	

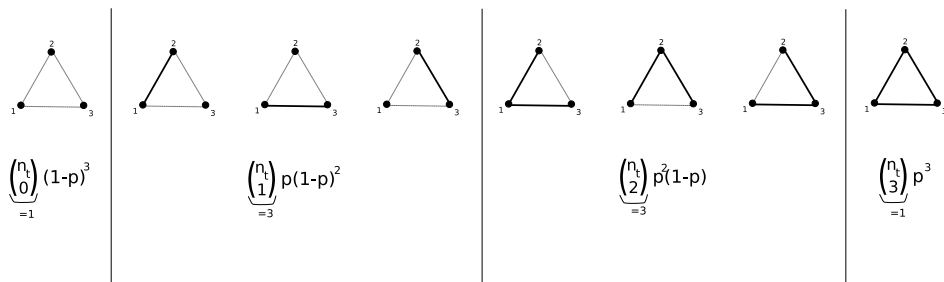
Zufallsnetzwerke clustern kaum - keine Korrelation.

- Betrachte ein Beispiel für das kanonische Ensemble $G(N, p)$

$$G(3, p) \implies n_t = \frac{3 \cdot 2}{2} = \binom{3}{2}$$

Die Zahl der Graphen ist:

$$\sum_{n=0}^{n_t} \binom{n_t}{n} = 2^{n_t} = 8$$



Es gilt immer:

$$\sum_{n=0}^{n_t} p_n = 1$$

1.5 Unterschiede zwischen Netzwerken und Gittern

Bei Netzwerken hängt die lokale Umgebung von N ab, bei Gittern nicht! \Rightarrow mögliche Schwellwerte in p werden durch Schwellwertfunktionen $p_c(N)$ ersetzt.

Die Strategie ist die Grenzwerte $N \rightarrow \infty$; $p \rightarrow 0$ zu betrachten und das Verhalten von:

$$\frac{p(N)}{p_c(N)} \rightarrow 0 \text{ oder } \infty$$

zu berechnen.

Beispiel: Erscheinen von Subgraphen mit k Knoten und l Kanten. (z.B. Bäume, Zyklen, vernetzte Graphen)

Mittlere Zahl dieser Subgraphen:

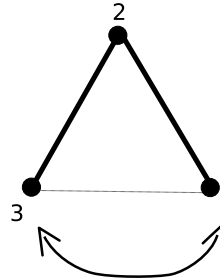
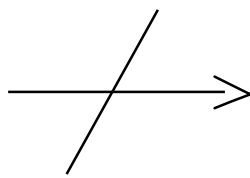
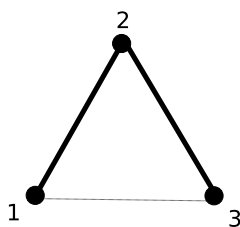
$$\langle r \rangle = \binom{N}{k} p^k \frac{k!}{a} \approx \frac{N^k p^l}{a}$$

$\binom{N}{k}$ - gibt die Anzahl der Möglichkeiten für Links an

p^l - Wahrscheinlichkeit für Link

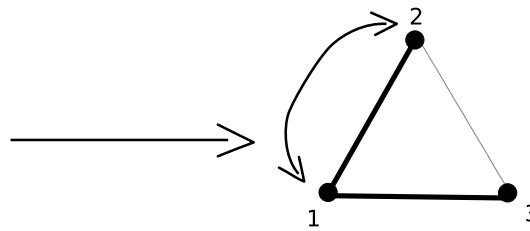
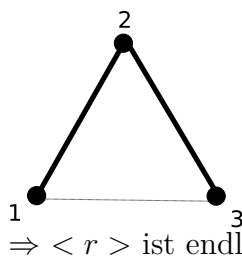
$\frac{k!}{a}$ - Permutationen müssen für Isomorphismen korrigiert werden

Motivation für den $\frac{k!}{a}$ Term. Welche Permutationen müssen gezählt werden?



Diese Permutation führt zu der selben Kontaktmatrix. Deshalb wird sie nicht mitgezählt.

1 Zufallsgraphen



Solche Permutationen führen zu einer neuen Kontaktmatrix und werden mitgezählt.

$$p_s(N) = C \cdot N^{-\frac{k}{l}}$$

daraus folgt:

$$\langle r \rangle = \text{const.} = \lambda = \frac{c^l}{a}$$

Bolobas 1985: Für strikte bilanzierte Graphen (insbesondere Bäume, Zyklen und vollständige Graphen)

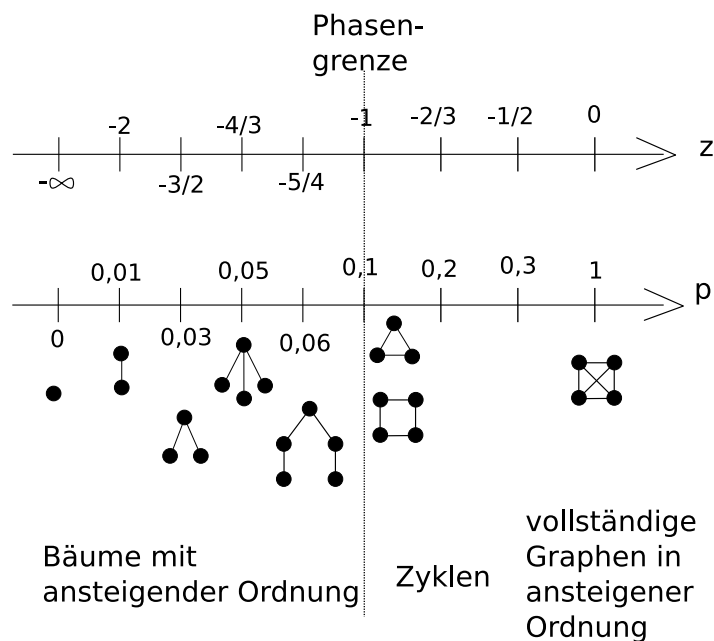
$$\lim_{N \rightarrow \infty} p(r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad \text{Poisson-Verteilung}$$

Bäume $l = k - 1 \rightarrow p_c = CN^{-\frac{k}{k-1}}$

Zyklen $l = k \rightarrow p_c = CN^{-1}$

vollst. Graph $l = \frac{k(k-1)}{2} \rightarrow p_c = CN^{-\frac{2}{k-1}}$

Betrachte $p = N^z \quad -\infty < z < 0$



Zentrales Ergebnis der Zufallsgraphentheorie (Erdős und Rényi 1960)

Die Netzwerkstruktur ändert sich sprunghaft bei $z = -1$ (dies entspricht $\langle k \rangle = 1$)
 $\langle k \rangle = 1$ ist unabhängig von N !

subkritisch **superkritisch**

isolierte Bäume Erscheinen einer gigantischen Komponente (“Perkolation”)

Die anderen Cluster sind Baumartig und verschmelzen zunehmend mit der gigantischen Komponente.

1.6 Maße für Graphen - Erdős-Rényi-Netzwerk

- Gradverteilung:
 exakt bekannt (Erdős und Rényi 1959)

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \xrightarrow{N \rightarrow \infty} \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

Poisson-Verteilung

⇒ Peak um Mittelwert $\langle k \rangle = pN$

- Durchmesser:
 Einfaches Argument: Ein Vertex hat $\langle k \rangle$ Nachbarn, der nächste ebenso, usw.
 Zahl der Nachbarn bei Entfernung l :

$$\langle k \rangle^l \stackrel{l=D}{\approx} N \quad \Rightarrow \quad D = \frac{\ln(N)}{\ln(\langle k \rangle)}$$

D ist nur schwach von der Systemgröße abhängig und typischerweise klein ⇒ “small world” Effekt.

→ typisch auch für reale Netzwerke.

→ kein gutes Maß um Erdős und Rényi mit Realen Netzwerken zu vergleichen.

- Cluster-Koeffizient

$$C = p = \frac{\langle k \rangle}{N}$$

Reale Netzwerke $C = \text{const.}$ (unabhängig von N) wie bei regulären Netzwerken (Gittern).

- Graphenspektrum (Spektraltheorie)
Superkritisch \Rightarrow Halbkreisverteilung; Wigner-Theorem; Zufallsmatrizen- Theorie

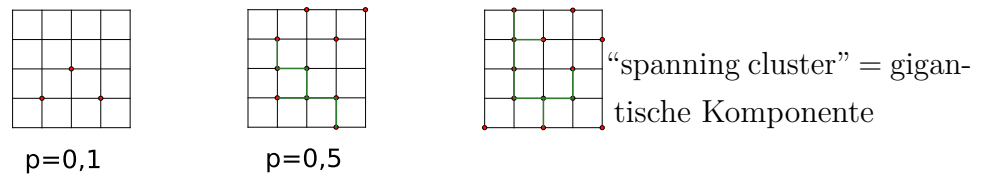
2 Perkolaton

Literatur: D.Stauber und A. Aharony "Introduction to percolation theory" Ind. rev. ed. 1994

Neuere B"ucher: Bolobas und Riordon 2002; Grimmelt 1999

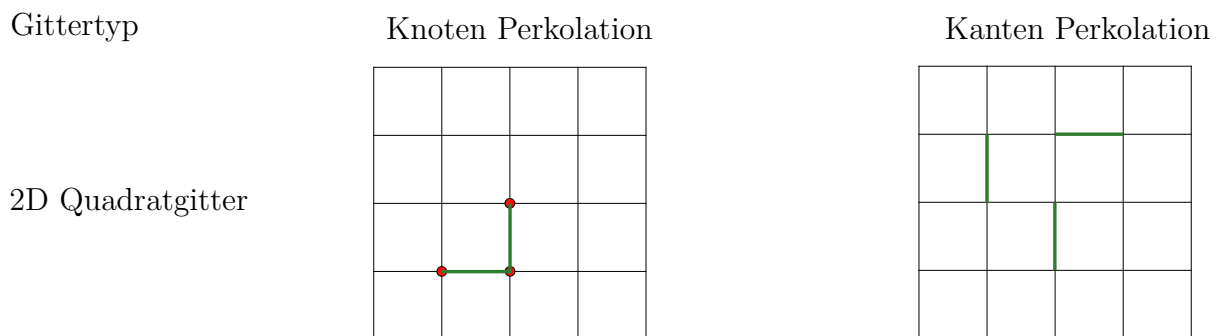
Betrachte einen Graphen (Gitter, Cayley-Baum, Zufallsgraphen) und klassifiziere jeden Knoten (alternativ: jede Kante) als "besetzt" oder "nicht besetzt" (Knoten- bzw. Kantenperkolaton).

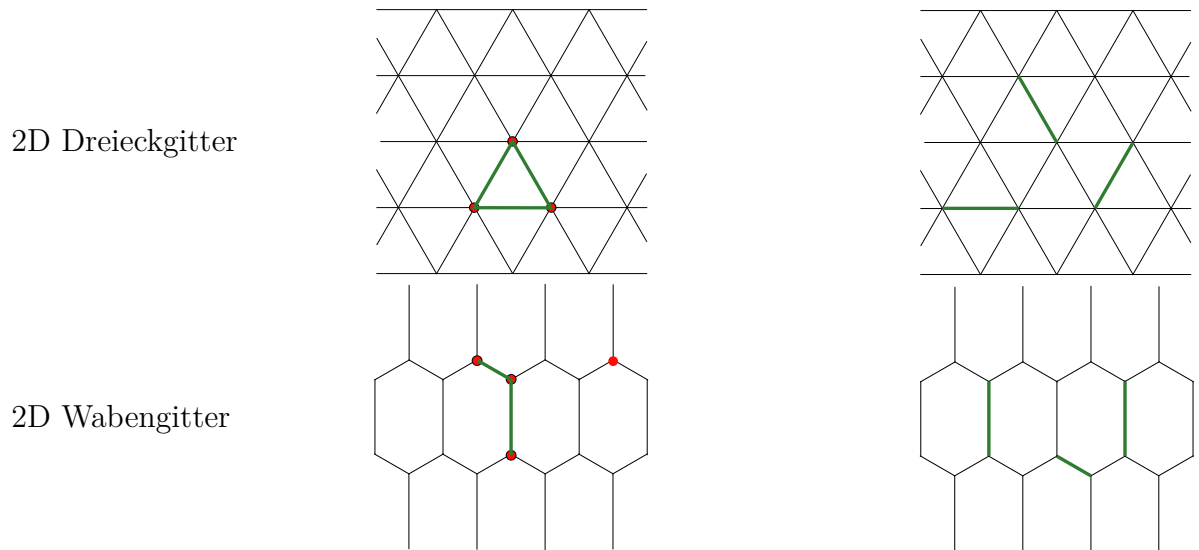
Beispiel 2D Quadratgitter, Knotenperkolaton.



Perkolatonsschwelle $p_c = 0,59$ (numerisch)

Eine exakte L"osung ist nicht bekannt. p_c h"angt von dem Gitter ab und ist f"ur Knoten-Perkolaton und Kanten-Perkolaton unterschiedlich.





Gitter	Knoten Perkolation	Kanten Perkolation
2D Quadratgitter	0,592746	0,5 (exakt)
2D Dreiecksgitter	0,5 (exakt)	$2\sin(\pi/19)$ (exakt)
2D Wabengitter	0,6962	$1 - 2\sin(\pi/18)$ (exakt)
3D einfach kubisch	0,3116	0,2488
4D Hyperwürfel	0,197	0,16
5D Hyperwürfel	0,141	0,118

Allgemein gilt hier:

$$p_c(\text{Knoten}) > p_c(\text{Kanten})$$

und p_c steigt mit fallender Dimension.

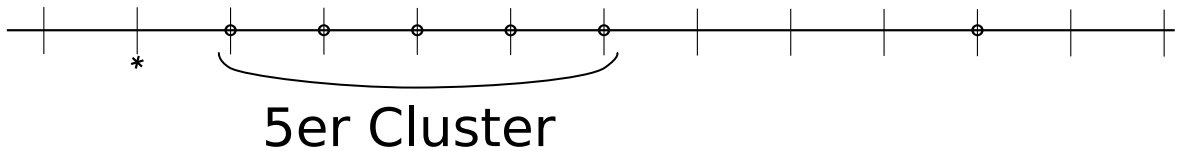
$d \rightarrow \infty$ (unendlich dimensionale Systeme) ähneln dem Zufallsgraphen ($p_c \rightarrow 0$)

2.1 Physikalische Anwendungen

- Ausbreitung von Waldbränden
- Diffusion in ungeordneten Medien (z.B. H_2 in Speichermedien)
- Ausbreitung von Krankheiten, Viren, Gerüchten und Meinungen in der Gesellschaft.

- Suchstrategie (Labyrinth, Bot, etc.)
- Anomale Diffusion in Zellen. (Diffusion mit Hindernissen)
- Fluss von Flüssigkeiten durch poröse Materialien
- elektrische Leitfähigkeit
- Polymerisation von Gelen

2.2 Perkolaton in 1D



Ein Knoten ist mit der Wahrscheinlichkeit p besetzt. Die kritische Wahrscheinlichkeit ist $p_c = 1$, da eine unbesetzte Stelle ausreicht, um das System zu unterbrechen.

Clustergröße $s \rightarrow$ Verteilung ? \rightarrow mittlere Clustergröße S

- Wahrscheinlichkeit, dass der Punkt * einen s -Cluster links begrenzt?

$$n_s = (1-p)p^s(1-p)$$

ist die Zahl der s -Cluster pro Knoten.

- Wahrscheinlichkeit, dass irgendein Knoten Teil eines s -Clusters ist?

$$n_s \cdot s$$

- Wahrscheinlichkeit, dass irgendein Knoten zu irgendeinem Cluster gehört?

$$\sum_{s=1}^{\infty} n_s s = (1-p)^2 \left(p \frac{d}{dp} \right) \underbrace{\sum_{s=1}^{\infty} p^s}_{\text{geo. Reihe}} = (1-p)^2 \left(p \frac{d}{dp} \right) \frac{p}{1-p} = (1-p)^2 p \frac{(1-p) + p}{(1-p)^2} = p$$

Wie erwartet.

- Wahrscheinlichkeit, dass ein besetzter Knoten zu einem s -Cluster gehört?

$$W_s = \frac{n_s s}{\sum n_s s} \quad \text{oder} \quad p W_s = n_s s$$

- Mittlere Clustergröße

$$S = \sum W_s s = \frac{\sum n_s s^2}{\sum n_s s} = \frac{1}{p} (1-p)^2 \left(p \frac{d}{dp} \right)^2 \sum_{s=1}^{\infty} p^s = \frac{1+p}{1-p} = \frac{1+p}{p_c - p} = S$$

mittlere Clustergröße S divergiert bei p_c mit einem Potenzgesetz.

Andere Systeme

$$S \sim \frac{1}{(p - p_c)^\alpha} \longrightarrow \text{krit. Phänomene}$$

- Korrelationsfunktion = Wahrscheinlichkeit, dass ein Knoten im Abstand r zum gleichen Cluster gehört.

$$g(0) = 1; \quad g(1) = p; \quad g(2) = p^2 \longrightarrow g(r) = p^r$$

$$g(r) = e^{-\frac{r}{\zeta}} \quad \text{mit} \quad \zeta = -\frac{1}{\ln(p)} = -\frac{1}{\ln(1 - p_c + p)} \approx \frac{1}{p_c - p}$$

Die Korrelationslänge divergiert bei p_c .

2.3 Zusammenfassung (1D)

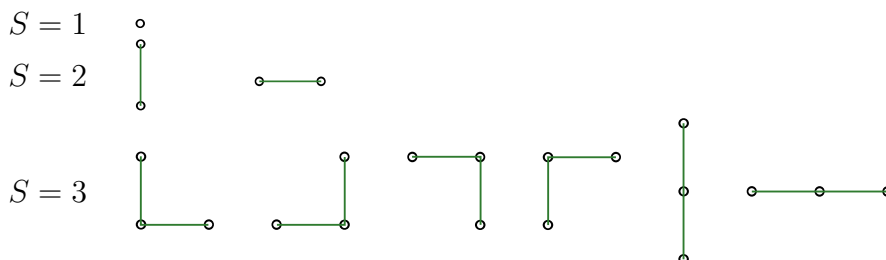
1D ist erstaunlich reichhaltig und kann exakt gelöst werden.

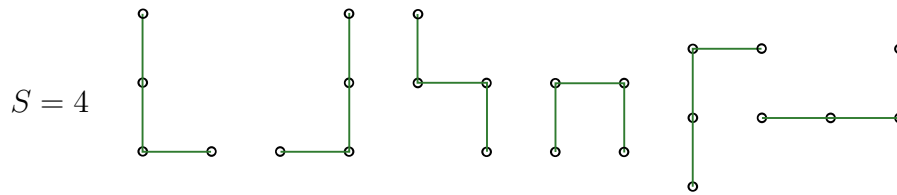
- Perkolationschwelle bei $p_c = 1$
- S und ζ divergieren mit Potenzgesetz bei p_c

Höhere Dimensionen: numerisch

geschicktes Abzählen von "Gittertieren"

z.B. 2D Quadratgitter:



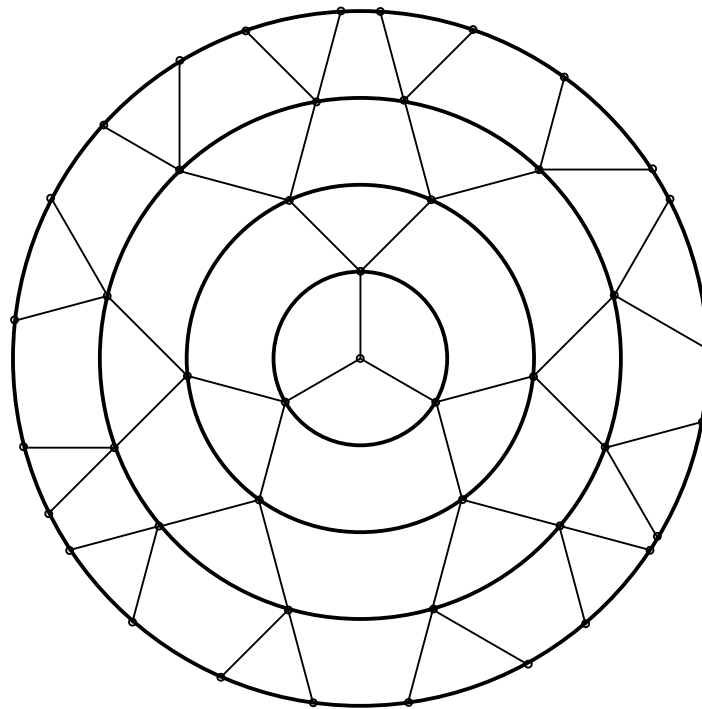


2.4 Bethe-Gitter

Ein elegantes Modell für den Übergang zu einem Zufallsgraphen ist das Bethe-Gitter (auch Cayley-Baum) → exakt lösbar.

Verzweigungsprozess: z Nachbarn \equiv Koordination $z - 1$

$z = 3$ und $r = 4$ Generationen:



Zahl der Knoten:

$$N = 1 + 3(1 + 2 + 4 + \dots + 2^{r-1}) = 1 + 3 \frac{1 + 2^r}{1 - 2} = 3 \cdot 2^r - 2$$

Allgemein $z \neq 3$:

$$N = 1 + z[1 + (z - 1) + (z - 1)^2 + \dots + (z - 1)^{r-1}] = 1 + z \frac{1 - (z - 1)^r}{1 - (z - 1)} = 1 + z \frac{(z - 1)^r - 1}{z - 2}$$

$$\frac{\text{Oberfläche}}{\text{Volumen}} = \frac{z(z-1)^{z-1}}{N} \rightarrow \frac{z-2}{z-1} \stackrel{z=3}{=} \frac{1}{2} = \text{const.}$$

das dieser Wert konstant ist ist typisch für eine Hyperkugel mit $D \rightarrow \infty$

$$A = R^{D-1} \quad V = R^D$$

$$\Rightarrow A = V^{\frac{D-1}{D}} = V^{1-\frac{1}{D}} \xrightarrow{D \rightarrow \infty} V$$

$$\Rightarrow \frac{A}{V} = \text{const. für } D \rightarrow \infty$$

Cayley-Baum entspricht $D \rightarrow \infty$

\Rightarrow Cayley-Baum entspricht einem unendlich-dimensionalen System. \Rightarrow Relevanz für Zufallsnetzwerke mit $N \rightarrow \infty$

Die Perkolationschwelle wird bei einem Cayley-Baum wie folgt bestimmt:
Suche Pfad von der Mitte ins Unendliche. Für jeden Schritt gibt es im Durchschnitt $p(z-1)$ Nachbarn. Iteration ergibt zwei Möglichkeiten: Divergenz oder Verschwinden.

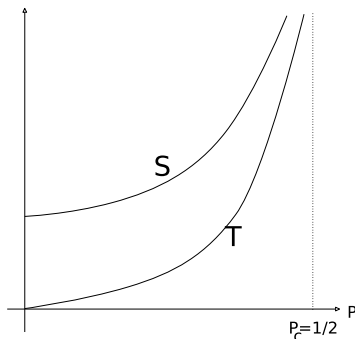
Schwelle: $p_c(z-1) = 1$

$$p_c = \frac{1}{z-1}$$

In diesem Fall sind Knoten und Kantenperkolaton äquivalent.

2.5 Mittlere Cluster-Größe S

Wobei der Ursprung mitgezählt wird und $z = 3$



$T =$ mittlere Clustergröße eines Zweiges.

\Rightarrow Unterzweig hat auch T

$$\Rightarrow T = (1-p) \cdot 0 + p(1+2T) \rightarrow T = \frac{p}{1-2p}$$

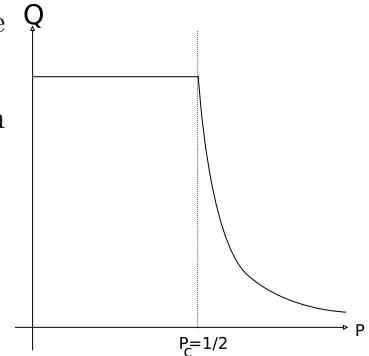
$$S = 1 + 3T = \frac{1+p}{1-2T}$$

2.6 Stärke des Netzwerks

P = Wahrscheinlichkeit, dass der Ursprung durch besetzte Gitterpunkte mit dem Rand verbunden ist.

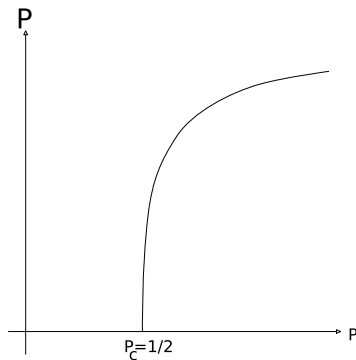
Q = Wahrscheinlichkeit, dass der Ursprung nicht mit dem Rand verbunden ist durch einen gegebenen Zweig.

$$\Rightarrow Q = (1 - p) + pQ^2 = \begin{cases} 1, & \text{für } p < p_c \\ \frac{1-p}{p}, & \text{für } p > p_c \end{cases}$$



Wahrscheinlichkeit, dass der Ursprung besetzt ist aber nicht mit dem Rand verbunden ist:

$$= p - P = pQ^2 \Rightarrow P = p(1 - Q^3) = \begin{cases} 0, & \text{für } p < p_c \\ p \left(1 - \left(\frac{1-p}{p} \right)^3 \right), & \text{für } p > p_c \end{cases}$$



P ist ein Ordnungsparameter für einen kontinuierlichen Übergang.

$P \sim p - p_c$ durch Entwicklung \Rightarrow kritischer Exponent = 1

2.7 Schlussfolgerung

- Beziehung zu Zufallsnetzwerken:

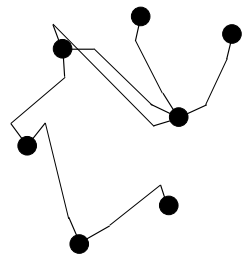
$$p_c = \frac{1}{z-1} \rightarrow p_c = \frac{1}{N} \quad \text{Schwellenwertfunktion}$$

- Perkolaton auf dem Cayley-Baum gehört zur gleichen Universitätsklasse wie Mean-Field-Perkolaton ($D \geq D_c = 6$) und ER-Zufallsgraphen.
- Achtung Unterschiede: Cayley-Baum hat keine Schleifen, im Gegensatz zum ER-Graphen oberhalb der Schwelle.
- Schleifen etc. können mit dem Cayley-Graphen nicht untersucht werden.

- Netzwerktheorie stellt generell andere Fragen als Perkolatonstheorie.

3 Verallgemeinerte Zufallsgraphen

Wir studieren Netzwerke, die vollständig zufällig sind, bis auf ihre Gradverteilung.



“Konfigurationsmodell”

m Schritte ohne Rückschritte

⇒ mittlere Zahl z_m der m nächsten Nachbarn. Im Prinzip gibt es jetzt Schleifen, aber deren Wahrscheinlichkeit ist $\sim O(1/N)$

3.1 Zufallsgraphen mit beliebiger Grad-Verteilung

1. starte mit Gradfolge $\{k_i\}$, z.B. durch Ziehen aus der gewünschten Gradverteilung p_k ; dies erzeugt an jedem Knoten einen Stern mit Strahlen.
2. verbinde Endpunkte der Strahlen zufällig.

Nun folgen wir einer der Kanten zum Nachbar B. Wie viele Nachbarn gibt es in l Schritten? Wähle zufälligen Vertex A → Zahl der Nachbarn $z_1 = \langle k \rangle = \sum_k k p_k$

Wahrscheinlichkeiten für Nachbarn bei B = $q_{k-1} \sim k p_k$. Wir schließen den Rückschritt aus. (Index um 1 erniedrigen)

$$q_{k-1} = \frac{k p_k}{\sum_j j p_j} \Rightarrow q_k = \frac{(k+1) p_{k+1}}{\langle k \rangle}$$

Die Zahl der Kanten auf denen man von B weiterlaufen kann, wenn man die Kante auf der man gekommen ist ausschließt, ist:

$$\Rightarrow \langle k \rangle_B = \sum_{k=0}^{\infty} k q_k = \frac{\sum (k+1) k p_{k+1}}{\langle k \rangle} = \frac{\sum k(k-1) p_k}{\langle k \rangle} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}$$

Die Mittlere Zahl der übernächsten Nachbarn von A:

$$z_2 = \langle k \rangle_B z_1 = \langle k^2 \rangle - \langle k \rangle$$

Erdős und Rényi Zufallsgraphen: Gradverteilung ist Poisson-Verteilt:

$$z_2 = \langle k^2 \rangle$$

Wie bei Cayley-Baum.

Bei realen Netzwerken wird z_2 durch $\langle k^2 \rangle$ bestimmt.

Argumente iterieren:

$$z_m = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} z_{m-1} = \left(\frac{z_2}{z_1} \right)^{m-1} z_1$$

(Phasen-) Übergang ist bei $z_2 = z_1$

$$\langle k^2 \rangle - \langle k \rangle = \langle k \rangle \Rightarrow 0 = \langle k^2 \rangle - 2 \langle k \rangle = \sum k(k-2)p_k = 0$$

Bedingung für den Phasenübergang in einem Zufallsgraphen mit beliebiger Gradverteilung. ER-Graph: (Poisson)

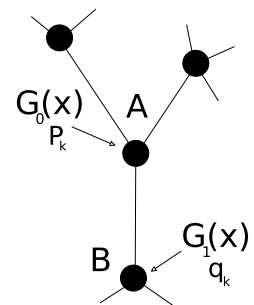
$$z_2 = \langle k \rangle^2 = z_1 = \langle k \rangle \rightarrow \langle k \rangle = 1$$

Clusterkoeffizient C ist klein aber er ist trotzdem größer als für ER.

$$\langle C \rangle = \frac{\langle k_i k_j \rangle}{Nz} = \frac{(\sum k q_k)^2}{Nz} = \frac{1}{Nz} \left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right)^2$$

$$\langle C \rangle = \underbrace{\frac{z}{N}}_p \underbrace{\left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle^2} \right)^2}_{> 1}$$

= 1, für ER, da $C = p$
 > 1, für reale Netzwerke



Da $z = pN$.

Beispiel: WWW

($N = 150000$; $z = 35$; $C = 0,11$)

$$C_{ER} = p = \frac{z}{N} = 2,3 \cdot 10^{-4}$$

$$C_{KM} = C_{ER} \cdot 210 = 0,05$$

Der Clusterkoeffizient ist für der Konfigurationsmodell deutlich größer als bei dem ER-Graphen, aber immer noch viel kleiner als bei realen Netzwerken.

3.2 Methode der erzeugenden Funktionen

Verwandle den Index k in eine kontinuierliche Variable x .

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \quad G_0(1) = 1$$

$G_0(x)$ "erzeugt" die Wahrscheinlichkeitsverteilung.

$$p_k = \frac{1}{k!} \left. \frac{d^k G_0}{dx^k} \right|_{x=0}$$

Die Momente folgen als:

$$z_1 = \langle k \rangle = G_0'(1)$$

$$\langle k^n \rangle = \left(x \frac{d}{dx} \right)^n G_0(x) \Big|_{x=1}$$

$G_0(x)$ zur n -ten Potenz erzeugt die n -fache Realisierung ("Potenzeigenschaft")

Beispiel: ($n = 2$)

$$(G_0(x))^2 = \sum_{k,j} p_j p_k x^{k+j} = p_0 p_0 + (p_0 p_1 + p_1 p_0) x + (p_0 p_2 + 2 p_1 p_1 + p_2 p_0) x^2$$

Der Vorfaktor von x^n ist die Wahrscheinlichkeit, dass die Summe genau n ergibt.

$$(G_0(x))^2 \Big|_{x=1} = 1^2 = 1$$

Was ist die erzeugende Funktion am nächsten Nachbarn?

$$G_1(x) = \sum_k q_k x^k = \sum_k \frac{(k+1)p_{k+1}}{\langle k \rangle} x^k = \sum_k \frac{k p_k}{\langle k \rangle} x^{k-1} = \frac{G_0'(x)}{G_0'(1)}$$

$$z_2 = z_1 \sum_k k q_k = z_1 G_1'(1) = z_1 \frac{G_0''(x)}{G_0'(1)} \Big|_{x=1} = G_0''(1)$$

Alternativ kann man die erzeugende Funktion für die Verteilung der nächsten Nachbarn herleiten.

$$G_2(x) = \sum_k p_k (G_0(x))^k = G_0(G_1(x))$$

Wir gehen über k Kanten nach außen und erzeugen k Realisierungen mit der Potenzierbarkeit:

$$z_2 = G'_2(1) = G'_0(G_1(1)) \cdot G'_1(1) = G'_0(1) \frac{G''_0(1)}{G'_0(1)} = G''_0(1)$$

Beispiel: ER-Graph

$$G_0(x) = \sum_{k=0}^N \binom{N-1}{k} p^k (1-p)^{(N-1)-k} x^k \stackrel{N \gg 1}{\approx} ((1-p) + px)^N = \left(1 + \frac{(x-1)z}{N}\right)^N$$

$$\underline{N \rightarrow \infty} \quad \boxed{e^{z(x-1)} = G_0(x)}$$

$$z = G'_0(1) = z \quad \text{Ok}$$

$$p_k = \frac{1}{k!} \frac{d^k}{dx^k} G_0(x) \Big|_{x=0} = \frac{1}{k!} z^k e^{-z} \quad \text{Poisson-Verteilung}$$

$$G_1(x) = \frac{G'_0(x)}{G'_0(1)} = \frac{ze^{z(x-1)}}{z} = G_0(x)$$

$$\boxed{z_2 = z_1 G'_1(1) = z_1^2}$$

Im ER-Graph macht es keinen Unterschied, ob man an einer Kante oder an einem Vertex startet.

3.3 Exponentieller Graph

Ein Graph der eine exponentielle Gradverteilung hat ist das Stromnetz (power grid).

$$\begin{aligned}
 p_k = ce^{-k/\kappa} \rightarrow 1 = \sum_k p_k &= c \sum (e^{-1/\kappa})^k = \frac{c}{1 - e^{-1/\kappa}} \\
 &\rightarrow p_k = (1 - e^{-1/\kappa}) e^{-k/\kappa} \\
 G_0(x) = (1 - e^{-1/\kappa}) \sum_k e^{-k/\kappa} x^k &= \frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}} \\
 \Rightarrow z_1 = G'_1(1) &= \frac{1}{1 - e^{-1/\kappa}}
 \end{aligned}$$

3.4 Potenzverteilung mit Cutoff

```

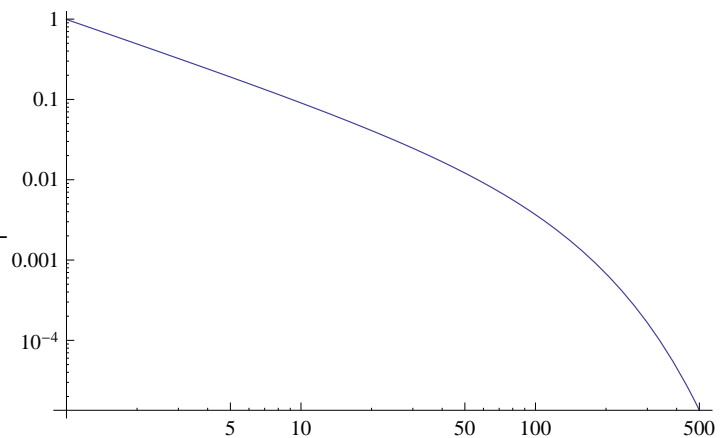
t = 1;
k = 100;
LogLogPlot[x^(-t) Exp[-x/k], {x, 1, 500}]

```

Die Cutoff-Länge ist κ .

$$p_k = ck^{-\tau} e^{-k/\kappa}$$

Reale Netzwerke folgen einem Potenzgesetz mit $1 \leq \tau \leq 3$



Beispiele:

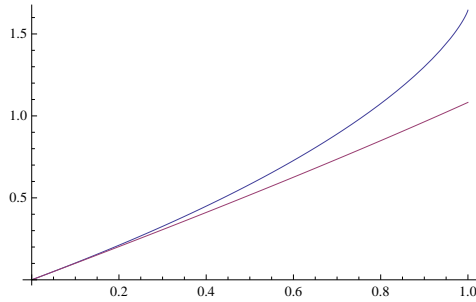
Netzwerk	Cutoff
WWW	1000
Internet	50
Mitautoren	400

$$1 = c \sum k^{-\tau} e^{-k/\kappa} = c Li_{\tau}(e^{-1/\kappa}) = G_0(1)$$

$Li_\tau(x)$ ist der τ -te Polylogarithmus

$$Li_\tau(x) = \sum_{k=1}^{\infty} k^{-\tau} x^k$$

`Plot[{PolyLog[2, x], PolyLog[4, x]}, {x, 0, 1}]`



$$Li_2(x) = x + \frac{x^2}{4} + \dots$$

$$Li_3(x) = x + \frac{x^2}{8} + \dots$$

$$Li_4(x) = x + \frac{x^2}{16} + \dots$$

In Mathematica kann man einen Polylogarithmus mit $PolyLog[\tau, x]$ verwenden. Die Ableitung ist:

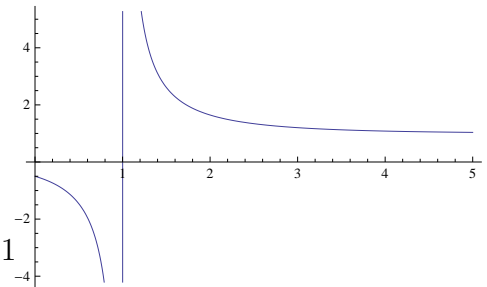
$$\frac{d}{dx} Li_\tau(x) = \sum_k k^{-\tau} k x^{k-1} = \frac{1}{x} Li_{\tau-1}(x)$$

$$Li_\tau(1) \stackrel{\kappa \rightarrow \infty}{=} \sum_k k^{-\tau} = \zeta(\tau)$$

$\zeta(x)$ ist die Riemannsche ζ -Funktion.

$$\zeta(1) = \infty; \zeta(2) = \frac{\pi^2}{6}; \zeta(3) = 1,2; \zeta(4) = \frac{\pi^2}{90}; \zeta(\infty) = 1$$

`Plot[Zeta[x], {x, 0, 5}]`



$$p_k = \frac{1}{Li_\tau(e^{-1/\kappa})} k^{-2} e^{-k/\kappa}$$

$$G_0(x) = \frac{Li_\tau(xe^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})} \xrightarrow{\kappa \rightarrow \infty} \frac{Li_\tau(x)}{\zeta(\tau)}$$

$$G_1(x) = \frac{g'_0(x)}{G'_0(1)} = \frac{1}{x} \frac{Li_\tau(xe^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})} \xrightarrow{\kappa \rightarrow \infty} \frac{1}{x} \frac{Li_{\tau-1}(x)}{\zeta(\tau-1)}$$

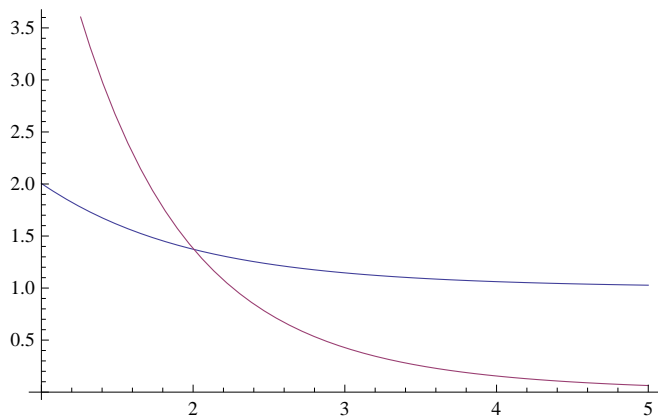
$$z_1 = G'_0(1) = \frac{Li_{\tau-1}(e^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})}$$

$$z_2 = G''_0(1) = \frac{Li_{\tau-2}(e^{-1/\kappa}) - Li_{\tau-1}(e^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})}$$

Cutoff nötig!

```

k = 3;
z1[t_] = PolyLog[t - 1, Exp[-1 / k]] / PolyLog[t, Exp[-1 / k]];
z2[t_] = (PolyLog[t - 2, Exp[-1 / k]] - PolyLog[t - 1, Exp[-1 / k]]) / PolyLog[t, Exp[-1 / k]];
Plot[{z1[t], z2[t]}, {t, 1, 5}]
    
```



3.5 Empirische Gradverteilung

Zahl der Verbindungen n_k mit Grad k

$$G_0(x) = \frac{\sum n_k x^k}{\sum n_k}$$

Beispiel: 1000 Personen Jede dieser Personen kennt 0,1,2,3,4 oder 5 andere.

$$\{n_k\} = \{86, 150, 363, 238, 109, 54\}$$

$$G_0(x) = \frac{1}{1000} (86 + 150x + 363x^2 + 238x^3 + 109x^4 + 54x^5)$$

3.6 Clustergrößenverteilung

ER-Graph: Es gibt einen Phasenübergang, denn oberhalb der Schwelle erscheint die gigantische Komponente.

Nun können zwei Größen definiert werden:

u : Anteil der Knoten, die nicht zur gigantischen Komponente gehören.

s : Anteil der Knoten, die zur gigantischen Komponente gehören.

$$u + s = 1$$

3 Verallgemeinerte Zufallsgraphen

u ist auch die Wahrscheinlichkeit, dass ein beliebiger Knoten nicht zur gigantischen Komponente gehört.

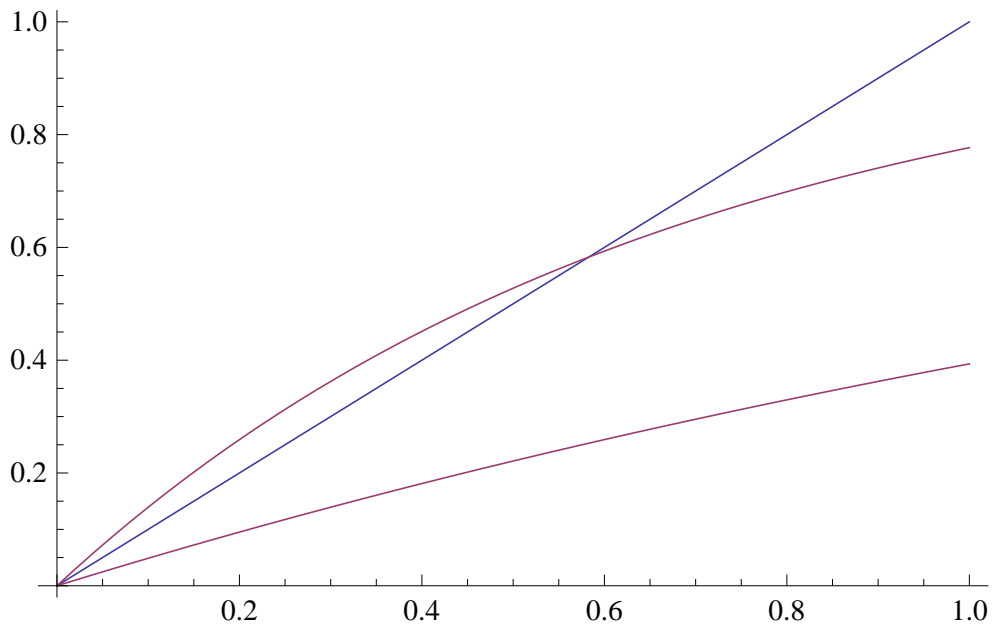
u ist auch die Wahrscheinlichkeit, dass keiner der Nachbarn zu der gigantischen Komponente gehört. Für den Grad k ist das auch u^k

$$u = \sum p_k u^k = e^{-z} \sum_{k=0} \frac{(zu)^k}{k!} = e^{z(u-1)}$$
$$s = 1 - u = 1 - e^{-zs} \quad z = pN$$

Die Gleichung für s kann nicht aufgelöst werden.

Sie kann jedoch graphisch gelöst werden:

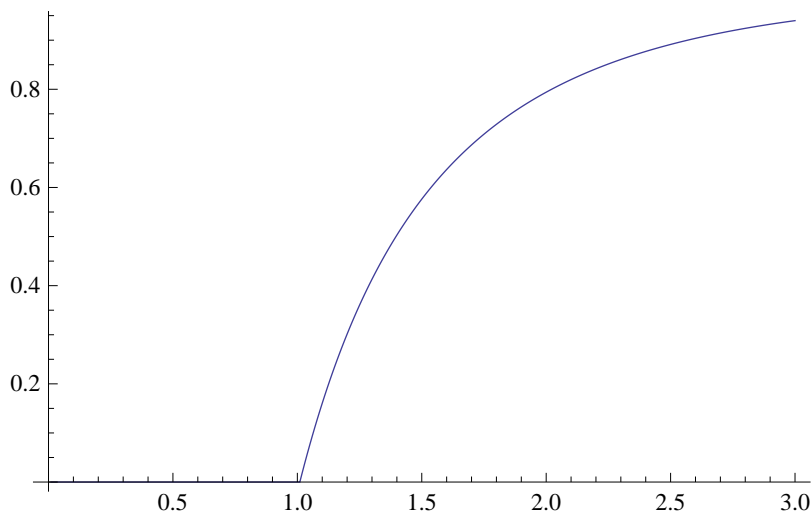
```
Plot[{x, Table[1 - Exp[-z * x], {z, 0.5, 1.5, 1}]}, {x, 0, 1}]
```



```

Num = 300; zmin = 0; zmax = 3; dz = (zmax - zmin) / Num;
A = Table[FindRoot[1 - Exp[-z * x] == x, {x, 0.1, 1}], {z, zmin, zmax, dz}];
B = Table[{t * dz, A[[t]][[1]][[2]]}, {t, 1, Num}];
ListLinePlot[B]

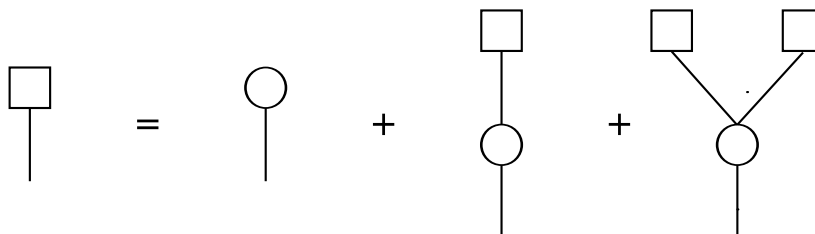
```



Diese Grafik zeigt die numerische Lösung. Bei $z=1$ ist ein Phasenübergang.

Clustergrößenverteilung für verallgemeinerten Zufallsgraph wird durch erzeugende Funktionen beschrieben. $H_1(x)$ für Größe der Cluster ohne gigantische Komponente und bei Eintreffen über Kante.

Graphische Entwicklung:

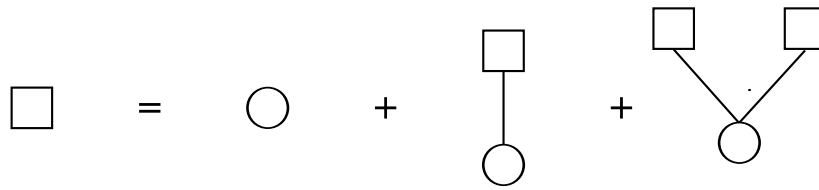


Daraus folgt eine selbstkonsistente Gleichung für $H_1(x)$.

$$H_1(x) = xq_0 + xq_1H_1(x) + xq_2H_1^2(x) + \dots$$

$$\boxed{H_1(x) = xG_1(H_1(x))} \tag{3.1}$$

$H_0(x)$ ist die erzeugende Funktion für die Clustergrößenverteilung ohne gigantische Komponente und bei Start an beliebigem Vertex.



$$\boxed{H_0(x) = xG_0(H_1(x))} \quad (3.2)$$

Rezept:

- berechne G_0 und G_1
- löse (3.1) für H_1
- löse (3.2) für H_0

Diese Rezept kann selbst für den ER-Graphen nicht exakt durchgeführt werden.

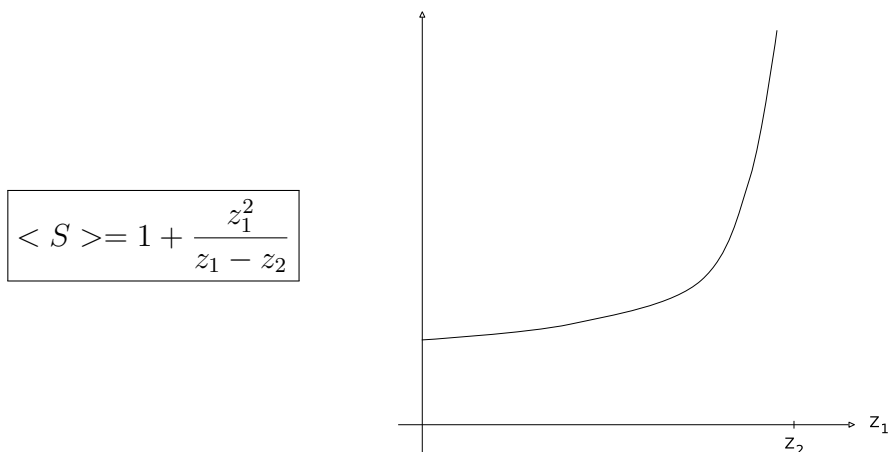
Für die mittlere Clustergöße gibt es das exakte Ergebnis:

$$\langle S \rangle = H'_0(1) = 1 + G'_0(1)H'_1(1)$$

(generierenden Funktionen sind unterhalb des Phasenübergangs normiert)

$$H'_1(1) \stackrel{(1)}{=} G_1(1) + G'_1(1)H'_1(1) \Rightarrow H'_1(1) = \frac{1}{1 - G'_1(1)}$$

$$\rightarrow \langle S \rangle = 1 + \frac{G'_0(1)}{1 - G'_1(1)} = 1 + \frac{z_1}{1 - \frac{z_2}{z_1}}$$



Oberhalb des Phasenübergangs: $H_0(1) < 1$ weil die gigantische Komponente ausgenommen wurde.

$$S = 1 - H_0(1) \stackrel{(3.2)}{=} 1 - G_0(u) \quad \text{mit} \quad u = H_1(1)$$

aus (3.1) folgt $u = G_1(u)$

Wiederholung der Rechnung für die mittlere Clustergröße.

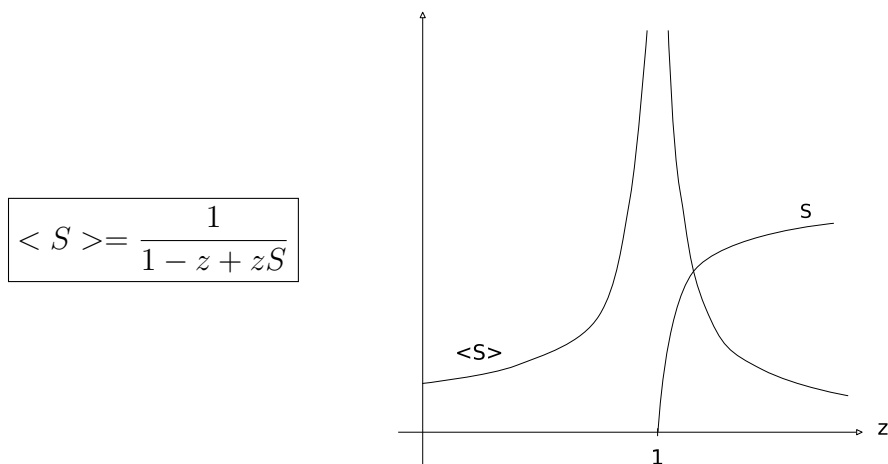
$$\begin{aligned} \langle s \rangle &= \frac{H'_0(1)}{H_0(1)} = \frac{1}{1-S} (G_0(1) + G'_0(u)H'_1(1)) \\ H'_1(1) &\stackrel{(3.1)}{=} G_1(u) + G'_0(u)H'_1(1) \Rightarrow H'_1(1) = \frac{G_1(u)}{1-G'_1(u)} \\ \rightarrow \langle s \rangle &= \frac{1}{1-S} \left[(1-S) + \frac{G'_0(u)G_1(u)}{1-G'_1(u)} \right] = 1 + \frac{[G'_0(u)]^2 [G'_0(1)]^{-1}}{(1-S)(1-G'_1(u))} \\ \langle s \rangle &= 1 + \frac{u^2 z_1}{(1-S)(1-G'_1(u))} \end{aligned}$$

für $u = 1, s = 0$ ist dies das alte Ergebnis. ER-Graph:

$$\begin{aligned} G_0(x) &= G_1(x) = e^{z(x-1)} \\ u &= G_1(u) = e^{z(u-1)} \\ S &= 1 - G_0(u) = 1 - u = 1 - e^{z(u-1)} \end{aligned}$$

$$\rightarrow \boxed{S = 1 - e^{-zS}} \quad \text{wie vorher}$$

$$\langle s \rangle = 1 + \frac{u^2 z}{(1-S)(1 - ze^{z(u-1)})} = 1 + \frac{z(1-S)}{(1 - ze^{z(u-1)})}$$



3.7 Gradverteilung eines Graphens mit Potenzverteilung (Exponent τ)

$\kappa \rightarrow 0$ kein Cutoff

$$u = G_1(u) = \frac{1 \text{ Li}_{\tau-1}(u)}{u \zeta(\tau-1)}$$

$$\tau \leq 2 \rightarrow u = 0; S = 1$$

Bei kleinen Exponenten ist jeder Knoten mit Wahrscheinlichkeit 1 in der gigantischen Komponente

$$\tau > 2 \rightarrow u > 0; S < 1$$

Die gigantische Komponente füllt nicht den ganzen Graphen aus.

z_m = Zahl der Nachbarn m Schritte entfernt - wird durch $G^{(m)}(x)$ erzeugt.

$$G^{(m)}(x) = \begin{cases} G_0(x) & m = 1 \\ G^{(m-1)}(G_1(x)) & m > 1 \end{cases}$$

Also

$$G^{(1)}(x) = G_0(x) \quad G^{(2)}(x) = G_0(G_1(x)) \quad G^{(3)}(x) = G_0(G_1(G_2(x)))$$

$$\begin{aligned} \Rightarrow z_m &= \frac{dG^{(m)}}{dx}(x=1) = (G^{(m-1)})'(1)G_1'(1) = \\ &= z_{m-1}G_1'(1) = [G_1']^{m-1} G_0'(1) = \left(\frac{z_2}{z_1}\right)^{m-1} z_1 \end{aligned}$$

Phasenübergang bei $z_1 = z_2$.

Abschätzung der mittlere Pfadlänge zwischen zwei zufällig ausgewählten Knoten: alle Knoten sollten nach l Schritten verbunden sein (oberhalb des Phasenübergangs)

$$1 + \sum_{m=1}^l z_m = SN \quad (S = 1 \text{ für Abschätzung})$$

$$N - 1 = \sum_{m=1}^l \left(\frac{z_2}{z_1}\right)^{m-1} z_1 = z_1 \frac{1 - \left(\frac{z_2}{z_1}\right)^l}{1 - \frac{z_2}{z_1}}$$

$$\Rightarrow l = \frac{\ln z_1^2 + (N - 1)(z_2 - z_1) - \ln z_1^2}{\ln z_2 / z_1}$$

$$N \gg z_1; z_2 = z_1 \Rightarrow l = \frac{\ln N / z_1}{\ln z_2 / z_1} + 1$$

l skaliert mit N wie $l = A + B \ln N$. Kleine-Welt-Eigenschaft gilt für alle verallgemeinerten Zufallsgraphen, reale Netzwerke haben aber oft andere Skalierung.

4 Small-World Netzwerke

Reale Netzwerke haben in der Regel die “Small-World” Eigenschaft. Dies bedeutet, dass die mittlere Pfadlänge mit dem Logarithmus der Anzahl der Netzteilnehmer geht.

$$l \sim \log(N)$$

Dieses Verhalten ist, wie bereits gesehen, typisch für Zufallsgraphen.

Aber: Sie haben in der Regel auch einen großen Clusterkoeffizienten, der oft unabhängig von N ist.

Solches Verhalten ist bei regulären Netzwerken (Gitter) vorhanden.

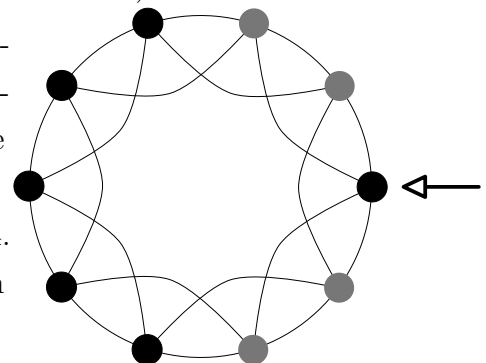
Die Frage, die uns in diesem Abschnitt beschäftigt, ist: Gibt es ein einfaches Modell, das ein reales Netzwerk mit kleinem l und großem C beschreibt?

4.1 Watts-Strogatz Modell

Das Watts-Strogatz Modell wurde erstmals in dem Artikel “Collective dynamics of ‘small-world’ networks”¹ in Nature 393, 440-442 (4 June 1998) beschrieben.

Es wird mit einem 1D Gitter mit periodischen Randbedingungen gestartet. Jeder Knoten hat K Nachbarn, $K/2$ Nachbarn nach rechts und ebenso viele nach links.

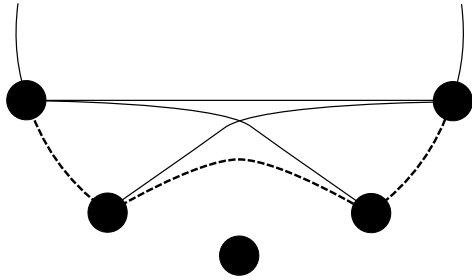
Das Bild rechts zeigt ein Beispiel mit $N=10$ und $K=4$. Der Pfeil zeigt auf den Punkt der mit allen grauen Punkten verbunden ist.



¹<http://www.nature.com/nature/journal/v393/n6684/pdf/393440a0.pdf>

Der Clusterkoeffizienten wird definiert über:

$$C = \frac{\text{Anzahl der Kanten in der Nachbarschaft}}{\text{max. Anzahl der Kanten}} = \frac{K_{real}}{K_{max}}$$



Für das oben angeführte Beispiel ist die Abzählung der Kanten in der Abbildung links zu sehen. Die gestrichelten Kanten sind die “Anzahl der Kanten” und die Summe aus den gestrichelten Kanten und den durchgezogenen ergibt die “max. Anzahl der Kanten”. In dem Beispiel erhalten wir $C = 3/6 = 1/2$.

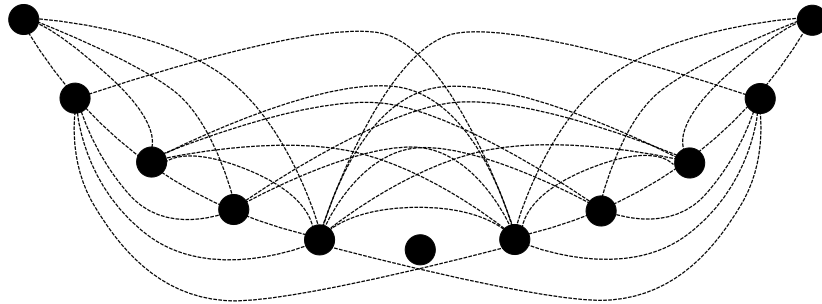
Für den allgemeinen Fall gilt: Die Punkte die rechts des markierten Punktes liegen bilden $K_{real}^{(r)}$ Kanten, die links liegen bilden $K_{real}^{(l)}$ Kanten.

$$K_{real}^{(r)} = \frac{K}{2} \left(\frac{K}{2} - 1 \right) = \frac{K - 2}{4}$$

$$K_{real}^{(l)} = \left(\frac{K}{2} - 1 \right) + \left(\frac{K}{2} - 2 \right) + \dots + 1 =$$

$$= \sum_{i=1}^{\frac{K}{2}-1} i = \frac{K(K - 1)}{8}$$

Die folgende Abbildung veranschaulicht die Aufsummation (Hier $K = 10$):



Berechne die mittlere Pfadlänge l :

Durch die periodischen Randbedingungen hat das Problem eine Symmetrie. Der am weitesten entfernte Punkt liegt bei $N/2$. Es gibt eine Symmetrie um die Entfernung $N/4$. Um diese Entfernung zu bewältigen benötigt man $K/2$ Schritte.

$$l = \frac{N}{2K}$$

l wächst also linear mit N . Dies nennt man: “large world” Eigenschaft.

Das Watts-Strogatz Modell besteht jetzt darin, dass man in dem regelmäßig aufgebauten System Unregelmäßigkeiten einbaut. Dazu geht man nach folgendem Schema vor:

- Wähle einen Startpunkt.
- Setze die Verbindung zum nächsten linken Nachbarn mit der Wahrscheinlichkeit p neu.
- Gehe im Uhrzeigersinn zum nächsten Punkt.
- Gelangt man wieder zum Startpunkt, so wiederholt man das Schema für den übernächsten Nachbarn.

Diese Schema wird also $K/2$ mal durchlaufen. Dabei ist darauf zu achten, dass keine doppelten Kanten (Melons) oder Kanten zum Ausgangspunkt (Tadpoles) gesetzt werden. Bei diesem Verfahren wird die gesamte Anzahl der Kanten nicht verändert. Es werden insgesamt $pNK/2$ Kanten neu gesetzt, diese werden “short-cuts” genannt. Da die linken Kanten ihren Ausgangspunkt behalten, hat jeder Punkt mindestens den Grad $K/2$. Es handelt sich also nicht um einen ER-Graphen.

Motivation:

- Soziologie: jeder hat ein lokales Netzwerk und zusätzlich ein paar weit entfernte Freunde.
- Polymere: Looping führt zu shortcuts (z.B. DNA)

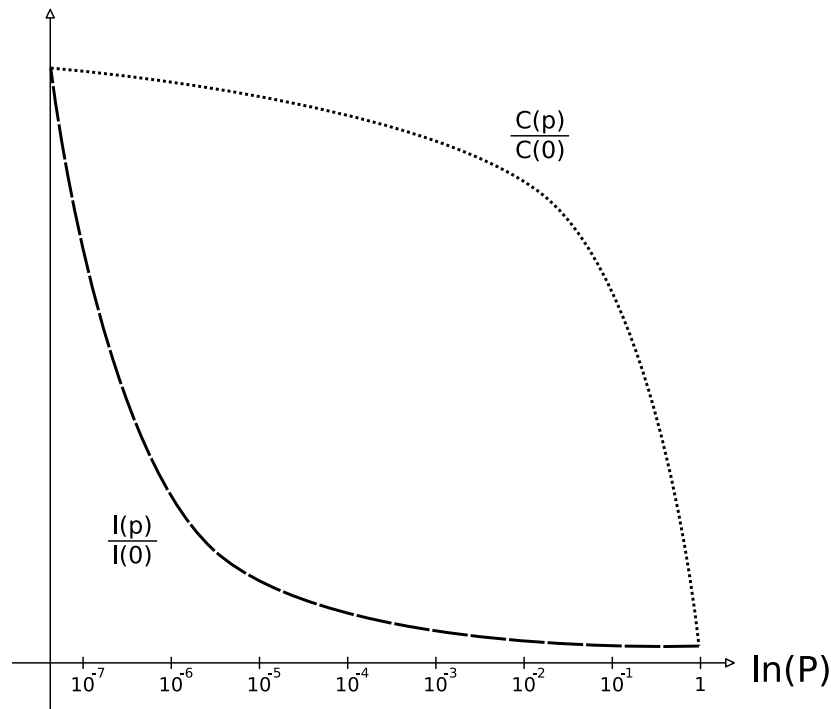
Bei $p = 1$ ist das Netzwerk (fast) zufällig.

$$l = \frac{\ln(N)}{\ln(K)} \quad C = p = \frac{K}{N}$$

Da Zufallsgraphen baumartig sind, kommt es zu der $K^l \sim N$ Abhängigkeit.

Es können zwei Grenzfälle unterschieden werden.

- Gitter , $l \sim N$, C groß
- Zufallsgraphen $l \sim \ln(N)$, C klein



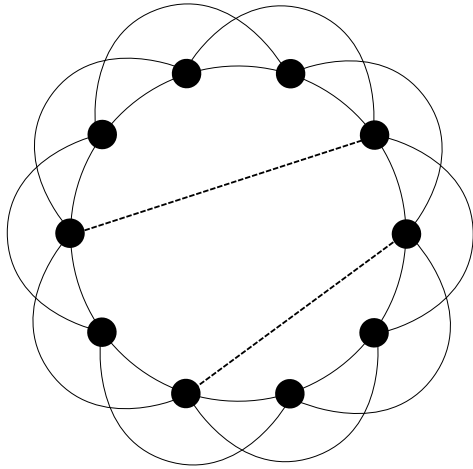
In dieser Abbildung (oben) wurden die Funktionen $l(p)$ und $C(p)$ qualitativ aufgetragen. Da l mit p viel schneller abfällt als C , kommt es über einen weiten Bereich zu einem Verhalten, was dem realer Netzwerke sehr ähnlich ist. In diesem Bereich sind die Voraussetzungen erfüllt, die am Anfang dieses Kapitels für die Simulation realer Netzwerke getroffen wurden. (C groß, l klein, “small-world”).

Das Watts-Strogatz Modell ist unnötig kompliziert. Es gibt folgende Probleme, die eine Realisierung erschweren:

- Nur die Endpunkte werden verlegt. → Jeder Punkt ist mindestens vom Grad $K/2$.
- Es werden keine Melons oder Tadpoles zugelassen. → Es kommt zu einer Inhomogenität.
- Das Netzwerk kann auseinanderfallen. → Es kommt zu Schwierigkeiten mit der mittleren Weglänge l .

Als Alternative wurde ein ähnliches Modell von Newman und Watts in dem Artikel “Scaling and percolation in the small-world network model”² in Phys. Rev. E 60, 7332 - 7342 (1999) vorgestellt. Hierbei werden nur neue Kanten dazugelegt, aber keine entfernt.

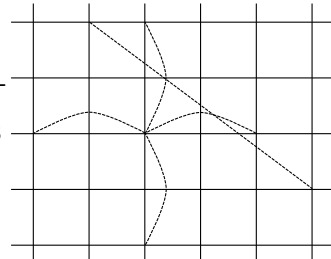
²http://prola.aps.org/abstract/PRE/v60/i6/p7332_1



p ist nun die Wahrscheinlichkeit, dass zusätzlich zu einer vorhandene Kante ein Shortcut eingeführt wird. Es gibt also $pNK/2$ neue Kanten. Der mittlere Grad ist nun $K(1 + p)$.

Für kleine p ist das Newman-Watts Modell äquivalent zu dem Watts-Strogatz Modell. Bei großem p ist es wesentlich einfacher zu analysieren - da l weiter definiert ist.

Diese Modelle kann man nun auf beliebige Dimensionen erweitern. Z.B. auf 2D wie in dem Bild rechts gezeigt ist.



Das Watts-Strogatz Modell lässt sich nicht analytisch lösen. Man muss Näherungen durchführen.

4.2 Cluster-Koeffizient

Für das Watts-Strogatz Modell gilt:

$$C(p = 0) = \frac{3(K - 2)}{4(K - 1)}$$

Für $p > 0$ sind Nachbarn weiterhin verbunden mit der Wahrscheinlichkeit $(1 - p)^3$. Die mittlere Zahl der Kanten beträgt $N_0(1 - p)^3$.

$$C(p) = \left\langle \frac{\text{Anzahl Kanten}}{\text{max. Anzahl Kanten}} \right\rangle \simeq \frac{\langle \text{Anzahl Kanten} \rangle}{\langle \text{max. Anzahl Kanten} \rangle} = C(0)(1 - p)^3$$

Im Newman-Watts Modell:

$$C = C(0) \frac{1}{1 + \frac{K}{K-1} p(p + 2)}$$

4.3 Gradverteilung im Watts-Strogatz Modell

Aus $p = 0$ folgt $k = K$.

Aus $p > 0$ folgt $\langle k \rangle = K$ - also wird die Gradverteilung breiter. Der minimale Grad ist $K/2$. Also kann man schreiben:

$$k_i = \frac{K}{2} + c_i \quad \text{mit} \quad c_i = c_i^{(1)} + c_i^{(2)} \geq 0$$

Es gilt $c_i^{(1)} \leq K/2$. Dies sind die Kanten die mit der Wahrscheinlichkeit $1 - p$ unverändert bleiben.

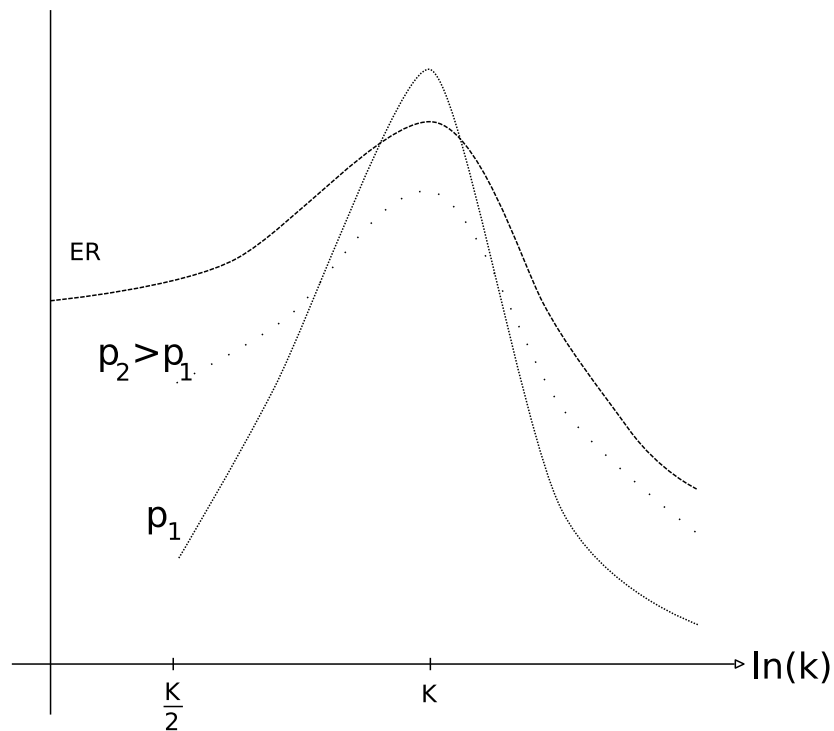
Die Konstante $c_i^{(2)}$ gibt die Anzahl der Kanten an, die zu i gelegt werden. Dies geschieht mit der Wahrscheinlichkeit $1/N$.

$$p(c_i^{(1)}) = \binom{\frac{K}{2}}{c_i^{(1)}} (1-p)^{c_i^{(1)}} p^{\frac{K}{2}-c_i^{(1)}}$$

$$p(c_i^{(2)}) = \binom{pN\frac{K}{2}}{c_i^{(2)}} \left(\frac{1}{N}\right)^{c_i^{(2)}} \left(1 - \frac{1}{N}\right)^{pN\frac{K}{2}-c_i^{(2)}} \xrightarrow{N \rightarrow \infty} \frac{\left(p\frac{K}{2}\right)^{c_i^{(2)}}}{c_i^{(2)}!} e^{-p\frac{K}{2}}$$

$$\xRightarrow{n=c_i^{(1)}} p(k) = \sum_{n=0}^{\text{Min}(k-\frac{K}{2}, \frac{K}{2})} \binom{\frac{K}{2}}{n} (1-p)^n p^{\frac{K}{2}-n} \frac{\left(p\frac{K}{2}\right)^{k-\frac{K}{2}-n}}{\left(k - \frac{K}{2} - n\right)!}$$

Das Watts-Strogatz Modell besitzt die gleichen qualitativen Eigenschaften wie das ER-Graphen. Also einen klaren Peak und darum herum einen exp Abfall.



4.4 Durchschnittliche Verbindungslänge l

WS-Ergebnisse sind nur numerisch zu erhalten. In der Originalarbeit wurde die Knotenzahl $N = 1000$ gewählt.

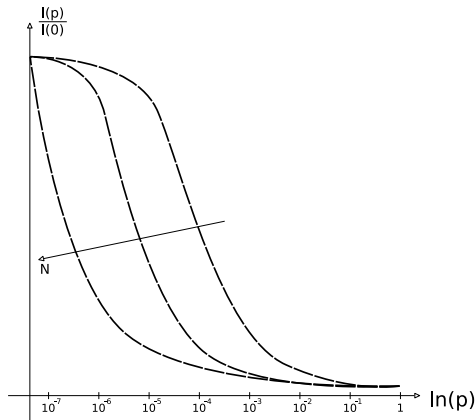
Mean Field Ergebnis von Newman et.al.

$$l = \frac{2N}{K} f\left(\frac{pNK}{2}\right)$$

Mit:

$$f(x) = \frac{1}{2\sqrt{x^2 + 2x}} \arctan\left(\frac{x}{\sqrt{x^2 + 2x}}\right) \sim \begin{cases} 1/4 & x \ll 1 \rightarrow l \sim N \text{ large world} \\ \frac{\log(2x)}{4x} & x \gg 1 \rightarrow l \sim \ln(N) \text{ small world} \end{cases}$$

Da das Netzwerk nicht skalenfrei ist, ist die Systemgröße relevant.



Skalenregime für kleines l und großes C :
 definiere typische Region zwischen zwei shortcuts.
 (Gebiet das durch die Störung nicht verändert wird.)

$$Nj = \frac{\text{Anzahl der Knoten}}{\text{Anzahl der shortcuts}} \sim \frac{N}{pN} = \frac{1}{p}$$

$$N^* \sim \frac{1}{p} \ll N \quad \text{small world Eigenschaft.}$$

Aber: Reale Netzwerke haben skalenfreie Gradverteilung.

5 Wachsende Netzwerke

Barabasi und Albert haben in Science 286, 509-512 (1999) den grundlegenden Artikel “Emergence of scaling in scaling networks”¹ zu diesem Thema geschrieben. Der Mechanismus um ein solches Netzwerk zu erstellen basiert auf folgenden Schritten:

1. Wachstum
2. “Preferential attachment”

Beispiele:

1. Eine neue Webseite enthält bevorzugt Links auf gut vernetzte Webseiten. (google etc.)
2. Eine neue Veröffentlichung enthält immer Verweise auf bestimmte Schlüssel-Veröffentlichungen.
3. Eine Person mit vielen Freunden wird noch mehr Freunde bekommen.

Das “the rich get richer” Prinzip war schon gut untersucht vor Barabasi und Albert.

- Merton 1968: “Mathäus-Effekt” (M 25-29)
- Simon 1955: “Gibrat-Prinzip”
- de Solla Price 1976: “Cumulative advantage”

5.1 Barabasi und Albert Modell

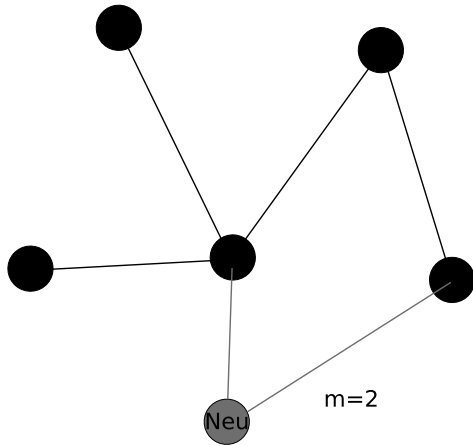
1. Wachstum:
Starte mit m_0 Kanten und füge in jedem Zeitschritt eine dazu.

¹<http://www.sciencemag.org/cgi/content/abstract/286/5439/509>

2. Preferential attachment:

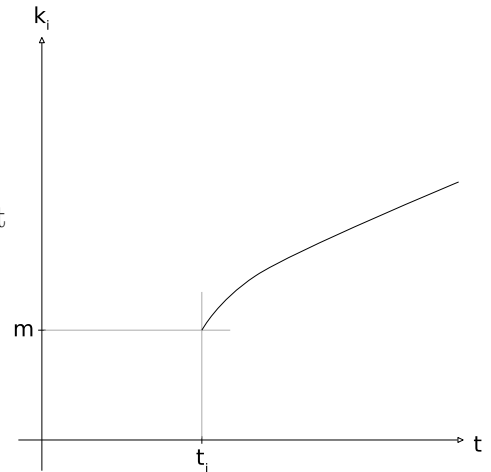
Der neue Knoten bekommt m Kanten, Verbindungen zu existierenden Knoten mit der Wahrscheinlichkeit

$$\Pi(k_i) = \frac{k_i}{\sum k_i}$$



Knoten	1	2	3	4	5
Grad	1	1	3	2	1

k_i startet zur Zeit t_i mit $k_i = m$ und wächst an mit der Zeit.



Trick: In einem Programm, das ein BA-Netz simuliert, kann man um die Auswahl des Zieles der Kanten zu erleichtern, indem man eine Liste erstellt, die den Index i k_i mal enthält.

$$L = (1, 2, 3, 3, 3, 3, 4, 4, 5) + (6, 6, 3, 2)$$

5.2 Exakte Ereignisse für Das BA-Modell

$$\langle C \rangle = \frac{m}{8N} (\ln(n))^2$$

Der Clusterkoeffizient ist klein und verschwindet mit großem N .

$$\langle l \rangle \sim \frac{\ln(N)}{\ln(\ln(n))}$$

Ist die mittlere Pfadlänge.

Mit diesen Ergebnissen kann man dem BA-Graphen die small-world Eigenschaft zuordnen.

5.3 Gradverteilung

Für die Gradverteilung gibt es bei den BA-Modell keine analytische Lösung. Es gibt verschiedene Näherungen.

Kontinuumstheorie: von Barabasi, Albert und Jeong

Betrachte k_i als kontinuierliche Variable.

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m \frac{k_i}{\sum k_i} = \frac{mk_i}{2mt} = \frac{k_i}{2t}$$

$$k_i(t_i) = m \Rightarrow k_i(t) = m\sqrt{\frac{t}{t_i}}$$

Der Grad jedes Knotens wächst mit einem Potenzgesetz.

Nun soll die Wahrscheinlichkeit berechnet werden, dass ein Knoten einen Grad $k_i < k$ besitzt.

$$p(k_i(t) < k) = p(t_i > \frac{m^2 t}{k^2})$$

Knoten werden gleichmäßig eingeführt:

$$p(t_i > t_0) = \frac{t_{tot} - t_0}{t_{tot}}$$

$$p(t_i > \frac{m^2 t}{k^2}) = \frac{m_0 + t - \frac{m^2 t}{k^2}}{m_0 + t}$$

$$p(k) = \frac{\partial p(k_i(t) < k)}{\partial k} = \frac{m^2 t}{m_0 + t} \frac{2}{k^3} t \xrightarrow{\infty} \frac{2m}{k^3}$$

Daraus ist abzulesen, dass die Gradverteilung einem Potenzgesetz mit $\gamma = 3$ folgt. Dies ist unabhängig von m . Dieses Ergebnis ist numerisch bestätigt.

Um zu zeigen, dass beide Regeln unersetzlich sind, werden zwei reduzierte Modelle betrachtet:

5.3.1 Modell A

Wachstum aber uniformes Anknüpfen.

$$\Pi(k_i) = \frac{1}{m_0 + t - 1}$$

$$\frac{\partial k_i}{\partial t} = \frac{m}{m_0 + t - 1} \quad \text{mit} \quad k_i(t_i) = m$$

$$k_i = m \cdot \ln \left[\left(\frac{m_0 + t - 1}{m_0 + t_i - 1} \right) + 1 \right]$$

$$p(k_i(t) < k) = p\left(t_i > (m_0 + t - 1)e^{1 - \frac{k}{m}} - m_0 + 1\right) = 1 - \frac{(m_0 + t - 1)e^{1 - \frac{k}{m}} - m_0 + 1}{m_0 + t}$$

$$p(k) = \frac{\partial}{\partial k} p(k_i(t) < k) = \frac{l}{m} e^{-\frac{k}{m}}$$

Modell A hat eine exponentielle Gradverteilung. Daraus folgt, dass Regel 2 für die Skalenfreiheit notwendig ist.

5.3.2 Modell B

Kein Wachstum aber proportionales Attachment. Bei diesem zweiten vereinfachtem Modell wird in jedem Zeitschritt ein Knoten ausgewählt, der Anfangspunkt einer neuen Kante wird. Der Zielknoten der neuen Kante wird so wie bisher gewählt. Das heißt die Wahrscheinlichkeit bei Knoten i zu landen ist gegeben durch Π .

$$\frac{\partial k_i}{\partial t} = \frac{1}{N} + \frac{k_i}{\sum k_i} \frac{N^2/2}{N(N-1)/2}$$

Der erste Term berücksichtigt den Prozess, dass ein Knoten als neuer Anfangspunkt gewählt werden kann. Der zweite Term ist die Änderung des Grades der Zielknoten. Dabei wurde $m = 1$ gesetzt.

$$\frac{\partial k_i}{\partial t} = \frac{1}{N} + \frac{N}{N-1} \frac{k_i}{2t}$$

Diese Differentialgleichung wird gelöst durch:

$$k_i(t) = 2 \frac{N-1}{N(N-2)} t + C t^{\frac{N}{2(N-1)}}$$

Für große Netzwerke wird dies zu ($N \rightarrow \infty$):

$$k_i(t) = \frac{2}{N} t + C \sqrt{t}$$

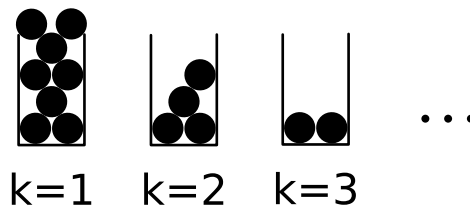
Da der Grad eines Knotens keinen stationären Wert erreicht $k_i \sim t$ macht es keinen Sinn eine Gradverteilung anzugeben.

5.3.3 Zusammenfassung

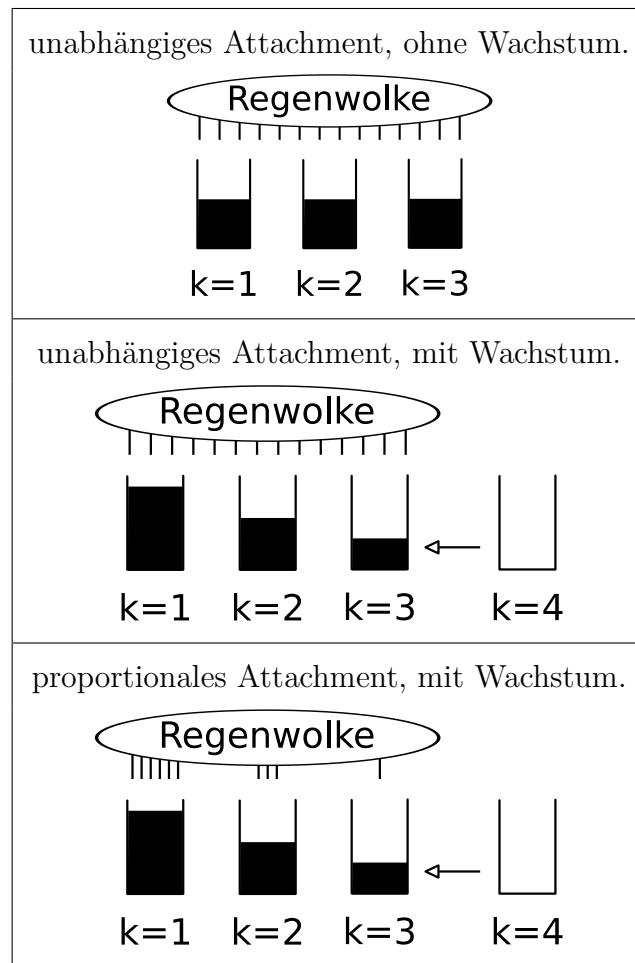
Eine kurze Übersicht über die Erkenntnis dieses Abschnitts:

Modell	$k_i(t)$	$p(k)$
BA-Modell	$\sim \sqrt{t}$	$\sim k^{-3}$
Modell A	$\sim \ln(t)$	$\sim e^{-\frac{k}{m}}$
Modell B	$\sim t$	-

Im folgendem wird noch eine “Regenanalogie” eingeführt. In den verschiedenen Modellen werden die Kanten verschieden auf die Knoten verteilt.



Die Kontinuumstheorie entspricht dem Einführen einer Füllhöhe, um die Anzahl der “Kugeln” in einem Behälter zu beschreiben. Die drei unterschiedlichen Modelle können so anschaulich als im Regen stehende Behälter beschrieben werden. Dies wird graphisch dargestellt:



5.4 Mastergleichungen

5.4.1 Gleichgewichtsnetzwerke

Die Anzahl der Knoten und Kanten ist konstant. Die Kanten werden jedoch zufällig neu gelegt.

Das Umlegen der Kanten folgt dem Schema:

1. Es wird zufällig eine Kante ausgewählt.
2. Einer der Endpunkte wird neu gelegt mit der Wahrscheinlichkeit $f(k)$

Kann man $f(k)$ so wählen, dass ein gewünschtes $p(k)$ resultiert? Das Ziel ist es eine Funktionalgleichung $p[f]$ zu erhalten.

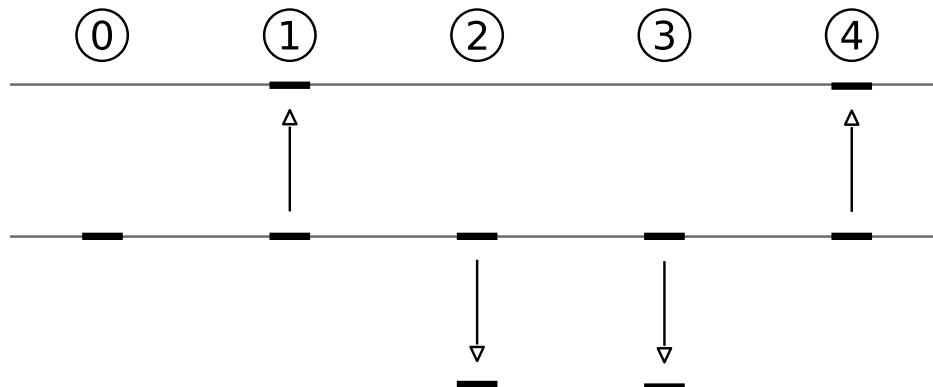
Die kann man mit der Mastergleichung erreichen.

Mastergleichung für $\langle N(k, t) \rangle = Np(k, t)$
 stationäre Verteilung $p(k)$

$$\sum_k \langle N(k, t) \rangle = N \sum_k p(k) = N$$

Aus Bilanzbetrachtungen kann man eine Differentialgleichung erhalten:

$$\begin{aligned} \langle N(k, t+1) \rangle = & \underbrace{\langle N(k, t) \rangle}_0 - \underbrace{\frac{f(k)}{N \langle f(k) \rangle} \langle N(k, t) \rangle}_1 + \\ & + \underbrace{\frac{f(k-1)}{N \langle f(k-1) \rangle} \langle N(k-1, t) \rangle}_2 - \underbrace{\frac{k}{N \langle k \rangle} \langle N(k, t) \rangle}_3 + \\ & + \underbrace{\frac{k+1}{N \langle k \rangle} \langle N(k+1, t) \rangle}_4 \end{aligned}$$



Diese Gleichung kann man auf $p(k, t)$ umschreiben und den Kontinuumslimites bezüglich t durchführen. Man erhält:

$$N \frac{\partial p}{\partial t} = \frac{1}{\langle f(k) \rangle} [-f(k)p(k, t) + f(k-1)p(k-1, t)] + \frac{1}{\langle k \rangle} [-kp(k, t) + (k+1)p(k+1, t)]$$

Eine stationäre Lösung existiert für (detailed balance)

$$\frac{f(k)}{\langle f(k) \rangle} p(k) - \frac{k-1}{\langle k \rangle} p(k+1) = 0$$

Wähle $\langle f(k) \rangle = \langle k \rangle$

$$f(k) = \frac{(k+1)p(k+1)}{p(k)}$$

Im Prinzip kann man so über einen Gleichgewichtsalgorithmus eine gewünschte Gradverteilung erreichen.

Exponentielle Gradverteilung: $f(k) = k + 1$

Eine skalenfreie Gradverteilung erreicht man mit:

$$f(k) = k \left(1 + \frac{1}{k}\right) \frac{k^{-\gamma} \left(1 + \frac{1}{k}\right)^{-\gamma}}{k^{-\gamma}} \simeq k + 1 + \gamma$$

Beachte: $\langle k \rangle = \int dk k \cdot k^{-\gamma} = \left[k^{-(\gamma-2)} \right]_{k_{min}}^{k_{max}} = \text{endl.}$

Daraus folgt dass: $\gamma \geq 2$

Aber: In Wirklichkeit kondensiert so ein Gleichgewichts skalenfreies Netzwerk. \Rightarrow Nicht-Gleichgewicht \Rightarrow BA-Modell

5.4.2 Wachsende exponentielle Netzwerke

t=1



Zur Zeit t : t Kanten und t Knoten.

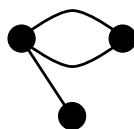
Im Unterschied zum Gleichgewichts-Netzwerk ist es wichtig, wann der Knoten eingeführt wurde: $t_i = s$

t=2



$$p(k, s, t) \rightarrow p(k, t) = \frac{1}{t} \sum_{s=1}^t p(k, s, t)$$

t=3



Mastergleichung:

$$p(k, s, t + 1) = \frac{1}{t} p(k - 1, s, t) + \left(1 - \frac{1}{t}\right) p(k, s, t)$$

Anfangsbedingungen: $p(k, s = [1, 2], t = 2) = \delta_{k,2}$

etc

Randbedingungen: $p(k, t, t \geq 2) = \delta_{k,1}$

Exakte Lösung: $p(k, s, t) = \frac{s}{t} \frac{1}{(k + 1)!} \left[\ln \left(\frac{t}{s} \right) \right]^{k+1}$

Diese exakte Lösung ist jedoch sehr unhandlich deswegen werden wir sie im Folgenden nicht weiter verwenden. Durch Multiplikation der Mastergleichung mit $(\frac{1}{t} \sum_{s=1}^t)$ ergibt sich:

$$\begin{aligned} \frac{1}{t} \sum_{s=1}^t p(k, s, t+1) &= \frac{1}{t} ((t+1)p(k, t+1) - p(k, t+1, t+1)) \\ &= \frac{1}{t} p(k-1, t) + (1 - \frac{1}{t}) p(k, t) \end{aligned}$$

$$(t+1)p(k, t+1) - tp(k, t) = p(k-1, t) - p(k, t) + \delta_{k,1}$$

Wir wenden den Kontinuumsliches an:

$$\frac{\partial}{\partial t} (tp(k, t)) = p(k-1, t) - p(k, t) + \delta_{k,1}$$

Mastergleichung für Gradverteilung, Unterschiede zur Mastergleichung für Gleichgewichts-Netzwerke.

1. Es werden nur Kanten hinzugefügt \rightarrow 2 Terme weniger
2. δ Term \rightarrow Kanten/Knoten werden hinzugefügt \rightarrow inhomogene Gleichung
3. $\frac{\partial}{\partial t}(tp)$ \rightarrow Netzwerk wächst

Stationärer Limes:

$$2p(k) - p(k-1) = \delta_{k,1}$$

Wir erhalten eine exponentielle Gradverteilung

$$p(k) = 2^{-k}$$

Dies entspricht den Ergebnis der Kontinuumstheorie.

5.5 Barabasi Albert Modell

Es wird das selbe Schema verwendet wie im letzten Abschnitt (wachsende exponentielle Netzwerke). Zum Zeitpunkt t gibt es t Kanten und t Knoten. Der totale Grad beträgt $2t$, da jede Kante einen Anfangs- und Endpunkt hat. Der durchschnittliche Grad ist

$\langle k \rangle = 2$

Mastergleichung:

$$p(k, s, t + 1) = \frac{k-1}{2t} p(k-1, s, t) + \left(1 - \frac{k}{2t}\right) p(k, s, t)$$

$$p(k, s = 1, 2, t = 2) = \delta_{k,2} \quad \text{Anfangsbedingung:}$$

$$p(k, s = t, t > 2) = \delta_{k,1} \quad \text{Randbedingung:}$$

Suche eine asymptotische Lösung: (d.h. $1 \ll s \ll t$)

$$p(k, s, t) = \sqrt{\frac{s}{t}} s^{-k} \sqrt{\frac{s}{t}}$$

Nun kann man die Mastergleichung mit \sum_s multiplizieren:

$$\frac{\partial}{\partial t} (tp(k, t)) = \frac{1}{2} [(k-1)p(k-1, t) - kp(k, t)] + \delta_{k,1}$$

Um den stationären Zustand zu berechnen setzt man $\partial_t p = 0$ ein:

$$p(k) - \frac{1}{2} [(k-1)p(k-1) - kp(k)] = \delta_{k,1}$$

Die Gradverteilung im BA-Modell ist nun:

$$p(k) = \frac{4}{k(k+1)(k+2)}$$

Für ein anderes m findet man:

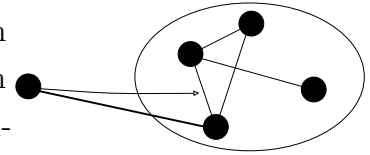
$$p(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

Daraus folgt, dass die Gradverteilung einem Potenzgesetz mit Grad $\gamma = 3$ folgt.

Dieses Modell kann in vielerlei Hinsicht erweitert werden.

Beispiel 1:

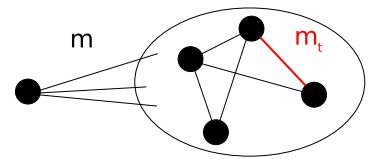
Neue Kante durch zufällige Selektion einer alten Kante. In diesem Modell gibt es ebenfalls eine Neigung der neuen Knoten sich mit gut verbundenen Netzteilnehmer zu verbinden. Durch Rechnung kann man erhalten: $f(k) \sim k$. Es ist also dem BA-Modell sehr ähnlich.



Beispiel 2:

Wähle zu den m neuen Kanten die den neuen Punkt verbinden auch $m' = cm$ neue Kanten zwischen den alten Knoten. Aus der Mastergleichung folgt:

$$\gamma = 2 + \frac{1}{1 + 2c}$$

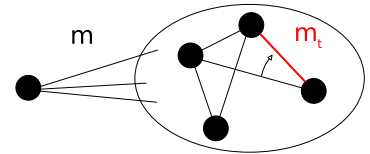


Wählt man c negativ so löscht man Kanten aus dem bestehenden Netzwerk. Hier muss man sich allerdings auf kleine c beschränken, da sonst das Netzwerk zerstört werden kann.

Beispiel 3:

Lege m_p alte Kanten in jedem Zeitschritt um.

$$\gamma = 2 + \frac{m - m_p}{m + m_p}$$



Schlussfolgerung: Das Hinzufügen neuer “Kanäle” zum BA-Modell erhält die Skalenfreiheit aber ändert γ .

Aber: Lineare Präferenz ($f(k) \sim k$ für $k \gg 1$) muss asymptotisch erhalten sein, sonst verliert man die skalenfreiheit (Gradverteilung ist kein Potenzgesetz mehr).

6 Biologische Netzwerke

Das biologische System „Zelle“ kann man als eine Ansammlung von mehreren gekoppelten Netzwerken betrachten. Die wichtigsten biologischen Netzwerke sind:

- Transkriptionsnetzwerke: Gene kodieren für Transkriptionsfaktoren, die dann andere Gene beeinflussen
- Protein-Protein-Interaktionsnetzwerke: Beispiel Hefe: $N = 6.000$ Proteine ergeben $N(N - 1)/2 = 18$ Millionen mögliche Verbindungen, von denen etwa 30.000 realisiert sind
- Metabolische Netzwerke
- Signaltransduktionsnetzwerke
- Neuronale Netzwerke

Jedes dieser Netzwerke kann dann wieder mit anderen Netzwerken zu einem Übernetzwerk gekoppelt werden.

Wichtige experimentelle Methoden um Protein-Daten zu erhalten:

- Gel-Elektrophoresis: Protein wird mit einem elektrischen Feld durch eine poröse Matrix gezogen und dadurch getrennt; z.B. SDS-PAGE (Tensid Sodium Dodecyl Sulfate, Matrix Polyacrylamide)
- 2D Gel-Elektrophoresis: erst Separation in pH-Gradienten, dann durch elektrisches Feld
- Affinitäts-Chromatographie: spezifische Bindung des Proteins an funktionalisierte Oberfläche
- Tandem affinity purification (TAP): Proteinkomplexe werden durch Bindung an ein Starterprotein ermittelt, das an einem Kügelchen hängt

- Massenspektrometrie (MS): Separation von geladenen und beschleunigten Bruchteilen in magnetischem Feld (zB matrix-assisted laser desorption/ionization time-of-flight, MALDI-TOF)

Nach der Entdeckung der skalenfreien Eigenschaften von verschiedenen realen Netzwerken (www, Autorenschaft, Schauspieler, etc) wurde gezeigt, dass auch biologische Netzwerke idR skalenfrei sind (metabolische Netzwerke: Jeong et al. Nature 2000, Protein-Protein-Interaktionsnetzwerk von Hefe: Jeong et al., Science 2001). Aber woher kommt hier diese Eigenschaft? Die einfachste Vermutung ist, dass der Mechanismus der Genduplikation zu preferential attachement führt (mit den Genen werden auch ihre Verbindungen kopiert, d.h. oft vorkommende Zielknoten werden verstärkt). Für das Drosophila Protein-Protein-Interaktionsnetzwerk wurde gezeigt, dass das duplication-mutation-complementation (DMC) model die Daten am besten beschreibt (besser als Erdős-Renyi, small-world, linear preferential attachement oder duplication-mutation without complementation) (Middendorf et al. PNAS 2005).

Biologische Netzwerke haben eine Besonderheit, die in den anderen realen Netzwerken nicht so wichtig zu sein scheint: sie haben ganz spezielle und hochvernetzte Subgraphen, die ganz spezielle Funktionen auszuführen scheinen. Das führt auf das Konzept der Netzwerkmotife, das von Uri Alon vom Weizmann Institut entscheidend entwickelt wurde (Buch Uri Alon, An Introduction to systems biology – design principles of biological circuits, Chapman & Hall/CRC 2007; Review Uri Alon, Networks motifs: theory and experimental approaches, Nature Reviews Genetics 2007). Z.B. wurde für das Transkriptionsnetzwerk von E Coli gezeigt, dass nur vier Netzwerkmotife statistisch signifikant auftreten (durch Vergleich ihrer Häufigkeit mit ER und verallgemeinerten Zufallsnetzwerken): negative autoregulation, feed-forward loop, single input module und dense overlapping regulons (Milo et al. Nature Genetics 2002). Um diese Ergebnisse zu erklären, reicht die Netzwerktheorie allerdings nicht mehr aus, dafür muss man die Dynamik dieser Motife untersuchen.