

Algorithmische Geometrie

Geometrische Zufallsgraphen – Euklidisch und hyperbolisch



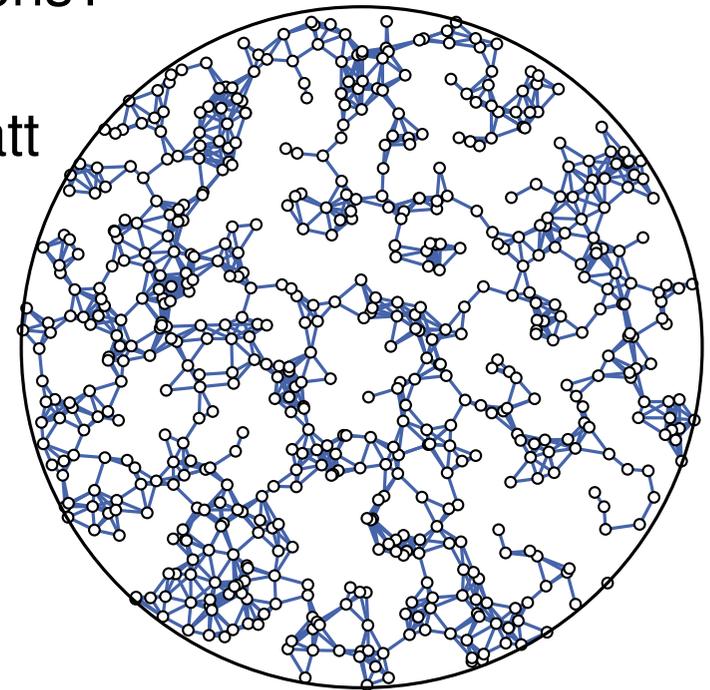
Geometrische Zufallsgraphen

Zufällige Generierung eines Graphen

- verteile n Knoten zufällig in einer Kreisscheibe mit Radius R
- verbinde zwei Knoten, wenn ihre Distanz maximal 1 ist

Fragen im Folgenden

- Wie kann man zufällige Punkte in einer Kreisscheibe generieren?
- Was ist der erwartete Knotengrad eines Knotens?
- Wie sollte man R wählen?
- Was passiert, wenn man die hyperbolische statt der Euklidischen Ebene benutzt?



Sampling der Knotenpositionen

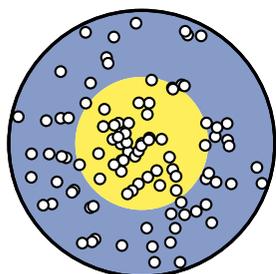
Koordinatensystem

- benutze Polarkoordinaten
- jeder Knoten v hat Winkel $\varphi_v \in [0, 2\pi)$ und Radius $r_v \in [0, R)$

Zufällige Positionen

- generiere einfach zwei Zufallszahlen: eine in $[0, 2\pi)$ und eine in $[0, R)$
- Problem: das verteilt die Punkte nicht „gleichmäßig“

Beispiel:

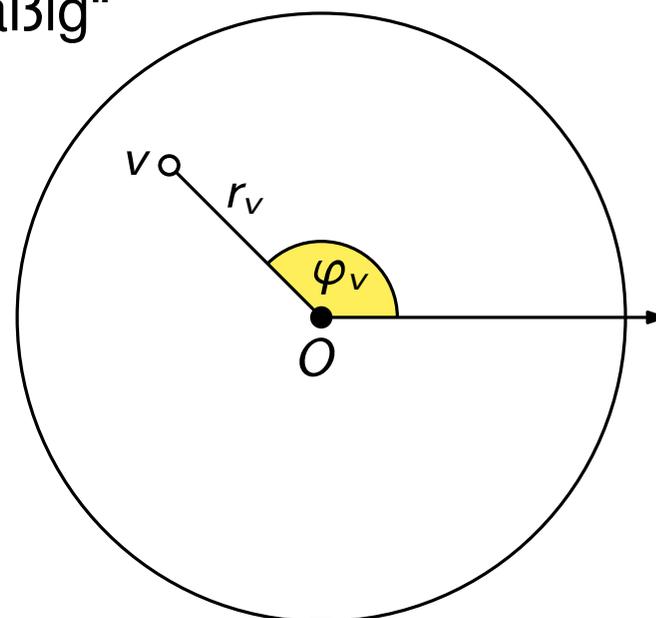


- 100 Knoten mit (φ, r) uniform gezogen aus $[0, 2\pi) \times [0, R)$
- Wahrscheinlichkeit für  und  ist gleich
- Fläche von  ist viel größer

- Ziel: für jede Region A ist die Wkt. für $v \in A$ proportional zur Fläche von A

- \rightarrow ziehe $r \in [0, R)$ mit $\mathbf{P}[r \leq x] = \frac{\pi x^2}{\pi R^2} = x^2/R^2$

- $F_r(x) = \mathbf{P}[r \leq x]$ ist die **Verteilungsfunktion** der Zufallsvariable r
(Englisch: **cumulative distribution function (CDF)**)



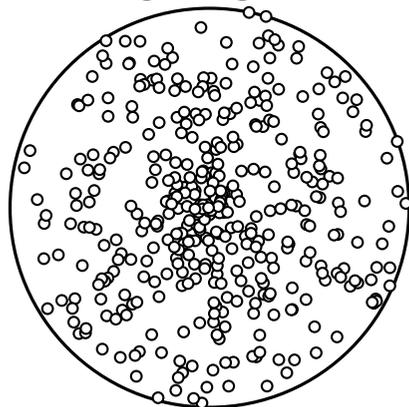
Sampling der radialen Koordinate

Basierend auf der uniformen Verteilung

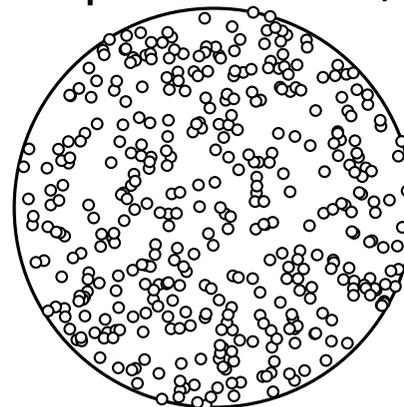
- ziehe $r \in [0, R)$ mit Verteilungsfunktion $F_r(x) = \mathbf{P}[r \leq x] = x^2/R^2$
- benutze dazu uniforme Zufallszahlen $z \in [0, 1)$
- es gilt: $F_z(x) = \mathbf{P}[z \leq x] = x$ (für $x \in [0, 1]$)
 - $\Rightarrow \mathbf{P}[z \leq x^2/R^2] = x^2/R^2$ (für $x^2/R^2 \in [0, 1]$)
 - $\Rightarrow \mathbf{P}[R\sqrt{z} \leq x] = x^2/R^2$ (für $x \in [0, R]$)
- also gilt: mit $r = R\sqrt{z}$ folgt r der gewünschten Verteilung
- es ist kein Zufall, dass $r(z)$ die Umkehrfunktion von $F_r(x)$ ist
(das funktioniert genauso auch für andere Verteilungen)

Beispiel

r uniform



r entsprechend $F_r(x)$



Erwarteter Knotengrad

Erwarteter Knotengrad von v

- Ziel: berechne Erwartungswert der Zufallsvariable $\text{deg}(v)$
- betrachte für jeden Knoten $u \neq v$ die Indikatorvariable $I_u = \begin{cases} 1, & \text{für } uv \in E \\ 0, & \text{sonst} \end{cases}$
- dann gilt:

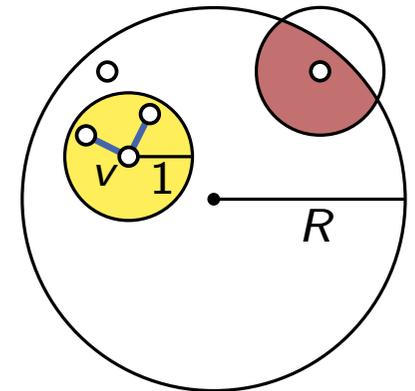
$$\text{deg}(v) = \sum_{u \neq v} I_u \Rightarrow \mathbf{E}[\text{deg}(v)] = \mathbf{E}\left[\sum_{u \neq v} I_u\right] = \sum_{u \neq v} \mathbf{E}[I_u] = (n-1)\mathbf{P}[uv \in E]$$

Wahrscheinlichkeit für $uv \in E$

- $uv \in E$ genau dann wenn u im Einheitskreis um v liegt

$$\mathbf{P}[uv \in E] = \frac{\text{Fläche Einheitskreis}}{\text{Gesamtfläche}} = \frac{\pi}{\pi R^2} = \frac{1}{R^2}$$

$$\Rightarrow \mathbf{E}[\text{deg}(v) \mid r_v \leq R-1] = \frac{n-1}{R^2}$$



Randfälle

- das stimmt natürlich nur wenn v nicht zu nah am Rand liegt
- eigentlich hätten wir gerne $\mathbf{E}[\text{deg}(v)]$ in Abhängigkeit von r_v
(\rightarrow bedingter Erwartungswert $\mathbf{E}[\text{deg}(v) \mid r_v]$)
- das kann man ausrechnen, muss man aber nicht
(wenn man sich nur für große n interessiert)

Abschätzung bei konstantem Durchschnittsgrad

Intuition

- sei der gewünschte Durchschnittsgrad d konstant
- für wachsendes n muss R auch wachsen
- Knoten mit Radius $> R - 1$ machen nicht viel aus

eben gezeigt:

$$\mathbf{E}[\deg(v) \mid r_v \leq R - 1] = \frac{n - 1}{R^2}$$

- setze zunächst $R = \sqrt{\frac{n-1}{d}} \Rightarrow \mathbf{E}[\deg(v) \mid r_v \leq R - 1] = d$

Obere und untere Schranken für $\mathbf{E}[\deg(v)]$

- es gilt: $\mathbf{E}[\deg(v)] = \mathbf{P}[r_v \leq R - 1] \cdot \mathbf{E}[\deg(v) \mid r_v \leq R - 1]$
(law of total expectation) $+ \mathbf{P}[r_v > R - 1] \cdot \mathbf{E}[\deg(v) \mid r_v > R - 1]$

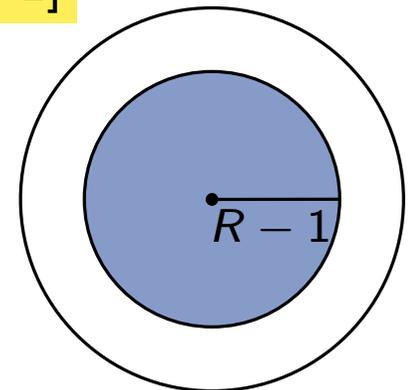
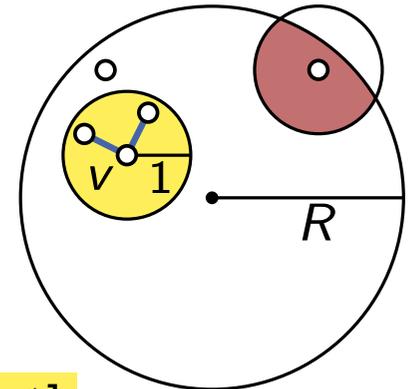
- es folgt direkt: $\mathbf{E}[\deg(v)] \leq d$ **Warum?**

- außerdem gilt: $\mathbf{E}[\deg(v)] \geq \mathbf{P}[r_v \leq R - 1] \cdot \mathbf{E}[\deg(v) \mid r_v \leq R - 1]$

$$= \frac{\pi(R - 1)^2}{\pi R^2} \cdot d = \left(1 - \frac{2}{R} + \frac{1}{R^2}\right) d$$

$$= d - \frac{2d^{3/2}}{\sqrt{n-1}} + \frac{d^2}{n-1}$$

- es folgt: $\mathbf{E}[\deg(v)] = d - O(n^{-1/2})$



Typische Eigenschaften echter Netzwerke

Überraschende Gemeinsamkeiten

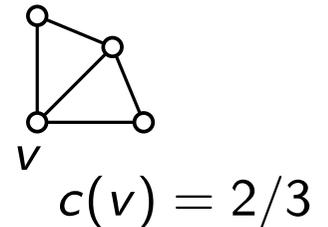
- Netzwerke aus unterschiedlichen Domänen unterscheiden sich strukturell (soziale Netzwerke, biologische Netzwerke, technische Netzwerke, Infrastruktur Netzwerke)
- es gibt dennoch Eigenschaften die sich viele (nicht alle!) Netzwerke teilen

Kleiner Durchschnittsgrad d

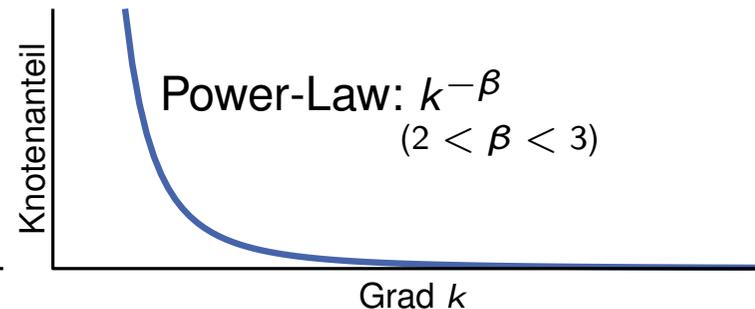
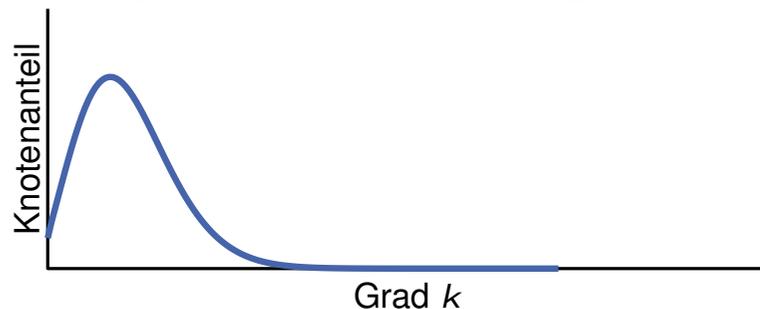
- typische Annahme: $d \in \Theta(1)$

Hohe Lokalität

- meist gemessen mit dem (lokalen) Clustering-Koeffizienten
- $c(v) = \mathbf{P}[uw \in E \text{ für zufällige Nachbarn } u, w \in N(v)]$



Homogene vs. Heterogene Gradverteilung



Kurze Wege (small world)

- der Durchmesser echter Netzwerke ist oft klein

siehe beispielsweise:

<https://www.sixdegreesofwikipedia.com/>

Euklidisch vs. hyperbolisch

Geometrische Zufallsgraphen in der Euklidischen Ebene

- typischer Name: *geometrische Zufallsgraphen*
- R kann so gewählt werden, dass $d \in \Theta(1)$
- hohe Lokalität: $c \in \Theta(1)$ (der Clustering-Koeffizient geht nicht gegen 0 für $n \rightarrow \infty$)
- homogene Gradverteilung: jeder Knoten hat etwa den gleichen Grad
- Durchmesser ist mit $\Theta(\sqrt{n})$ recht groß

Geometrische Zufallsgraphen in der hyperbolischen Ebene

- typischer Name: *hyperbolische Zufallsgraphen*
- $d \in \Theta(1)$
- hohe Lokalität: $c \in \Theta(1)$
- heterogene Gradverteilung: power law
- kleiner Durchmesser: $\Theta(\log n)$
- ist damit ein passendes Modell für viele Netzwerke

Hyperbolische Zufallsgraphen

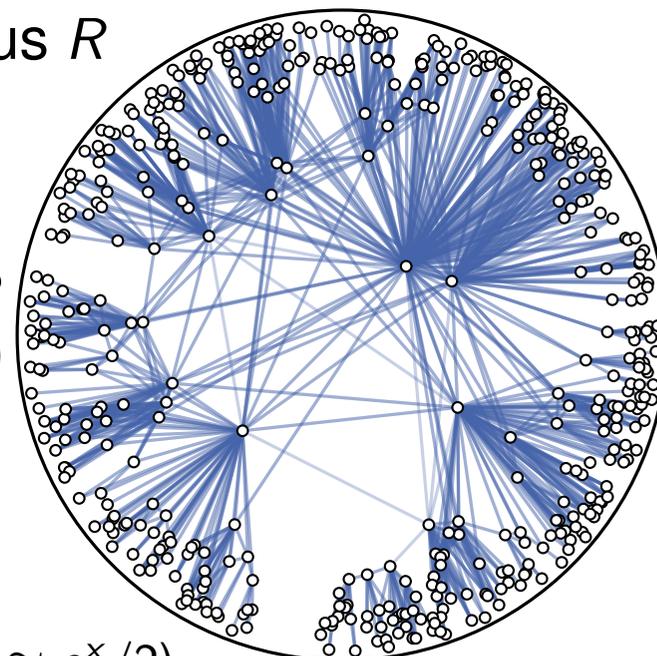
Zufällige Generierung eines Graphen

- verteile n Knoten zufällig in einer Disk mit Radius R
- verbinde zwei Knoten, wenn ihre Distanz $\leq R$

(natives Modell)

Fragen im Folgenden

- Welcher Verteilungsfunktion folgen die Radien?
(die Winkel sind natürlich wieder uniform in $[0, 2\pi)$)
- Was sind die erwarteten Knotengrade?
- Wie sollte R gewählt werden?



Disclaimer

- wir werden Terme häufig approximieren ($\sinh(x) \approx e^x/2$)
- typischerweise passen diese Approximationen sehr gut für großes n
- formal müsste man eine obere Schranke für den Fehler angeben
- dadurch wird die Vorlesung (hoffentlich) deutlich verständlicher
- wenn man etwas neues für hyperbolische Zufallsgraphen zu beweisen versucht, sollte man zunächst immer mit diesen Abschätzungen rechnen
(man muss ja erstmal eine Idee dafür bekommen, ob etwas stimmt oder nicht)

Verteilungsfunktion der Radien

Erinnerung

- Kreisfläche für Radius r : $2\pi(\cosh(r) - 1)$
- Ziel: für jede Region A ist die Wkt. für $v \in A$ proportional zur Fläche von A

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

Verteilungsfunktion

- ziehe $r \in [0, R)$ mit Verteilungsfunktion $F_r(x)$:

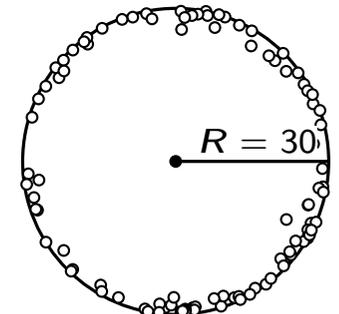
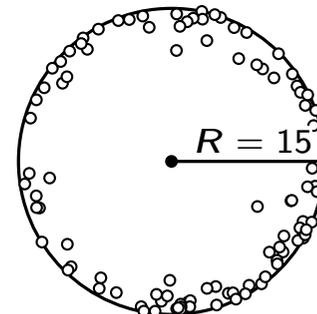
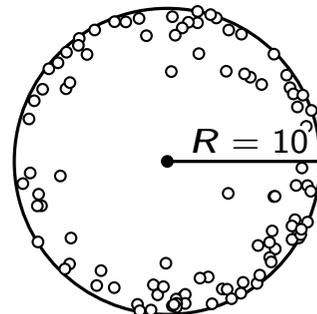
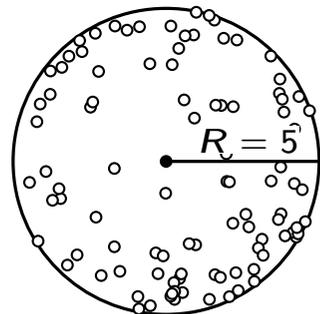
$$F_r(x) = \mathbf{P}[r \leq x] = \frac{2\pi(\cosh(x) - 1)}{2\pi(\cosh(R) - 1)} = \frac{\cosh(x) - 1}{\cosh(R) - 1} \approx e^{-(R-x)}$$

- also ziehe $z \in [0, 1)$ uniform und setze $r = \operatorname{arcosh}(z \cdot (\cosh(R) - 1) + 1)$
(Umkehrfunktion von $F_r(x)$)

Achtung

- wir können **nicht** einfach die Radien mit anderem R generieren und dann entsprechend skalieren
(im Euklidischen ging das noch)

- 100 Punkte für verschiedene R
(natives Modell)

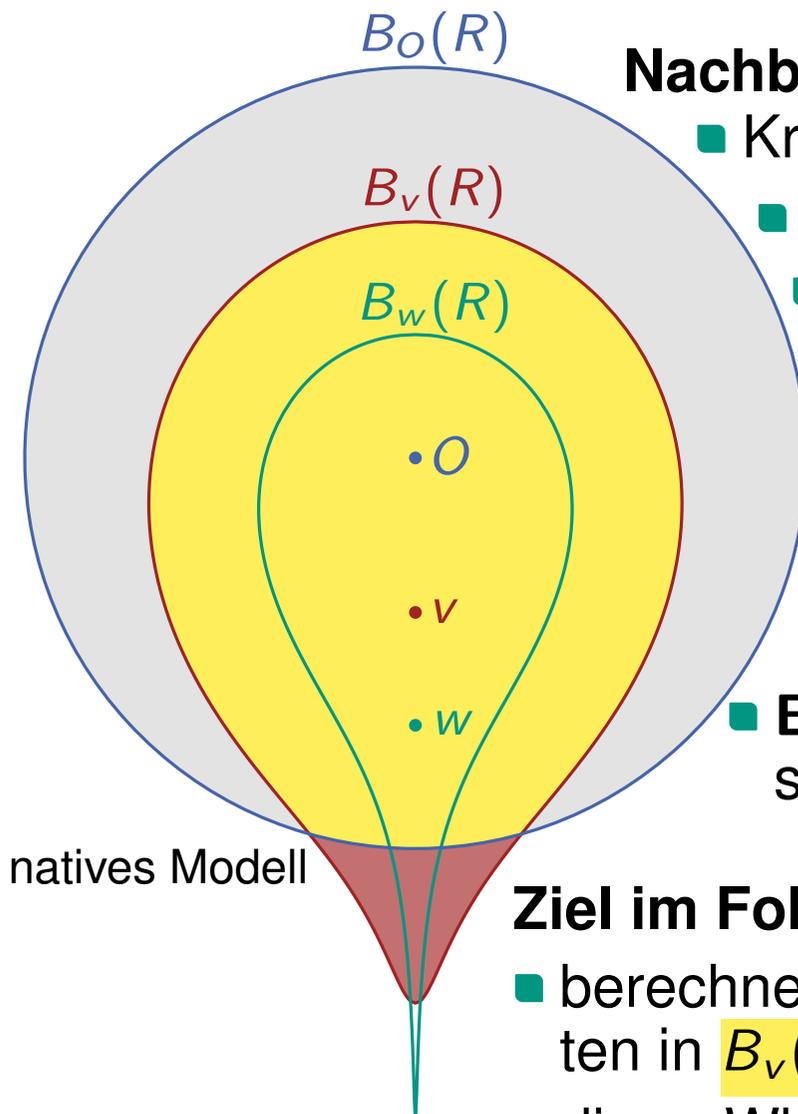


Knotengrade: Intuition

grundsätzliche Annahme
im Folgenden: $\varphi_v = 0$

Nachbarschaft eines Knotens v

- Kreis $B_v(R)$ mit Radius R um $v = (r_v, \varphi_v)$
- beachte: $r_v < R$ für alle Knoten v
- Wkt., für einen Knoten u dass $uv \in E$:
proportional zur Fläche von $B_v(R) \cap B_o(R)$
- $E[\deg(v) \mid r_v]$ ist damit ebenfalls proportional zu $B_v(R) \cap B_o(R)$
- $E[\deg(v) \mid r_v]$ schrumpft also, mit wachsendem r_v
- $E[\deg(v) \mid r_v]$ schrumpft stärker als das Bild suggeriert: $B_v(R) \setminus B_o(R)$ ist recht groß

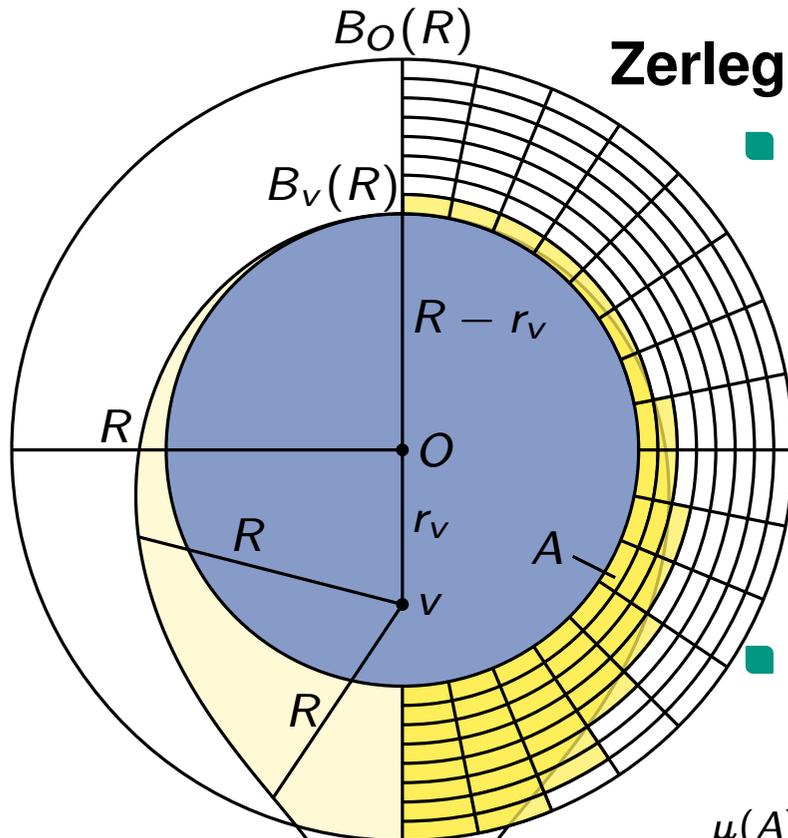


Ziel im Folgenden

- berechne Wkt., dass Knoten mit zufälligen Koordinaten in $B_v(R) \cap B_o(R)$ liegt (als von r_v abhängige Funktion)
- diese Wkt. heißt auch das **Maß** von $B_v(R) \cap B_o(R)$ bezeichnet mit $\mu(B_v(R) \cap B_o(R))$

Das Maß von $B_v(R) \cap B_o(R)$

grundsätzliche Annahme
im Folgenden: $\varphi_v = 0$



Zerlegung von $B_v(R) \cap B_o(R)$

- es gilt: $B_o(R - r_v) \subseteq B_v(R) \cap B_o(R)$
- für $\mu(B_o(R - r_v))$ wissen wir schon:
 $\mu(B_o(R - r_v)) = \mathbf{P}[r \leq R - r_v] = F_r(R - r_v) \approx e^{-r_v}$
- wegen Symmetrie: betrachte nur die Hälfte **A** von $B_v(R) \cap B_o(R) \setminus B_o(R - r_v)$
- zerlege **A** weiter in: Winkel der Breite $\Delta\varphi$ und Radien der Breite Δr
- Abschätzung für $\mu(A)$: summiere über Zellen

$$F_r(x) \approx e^{-(R-x)}$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

max. Winkeldifferenz zwischen Punkten mit Distanz $\leq R$ und Radien r_v, x_r

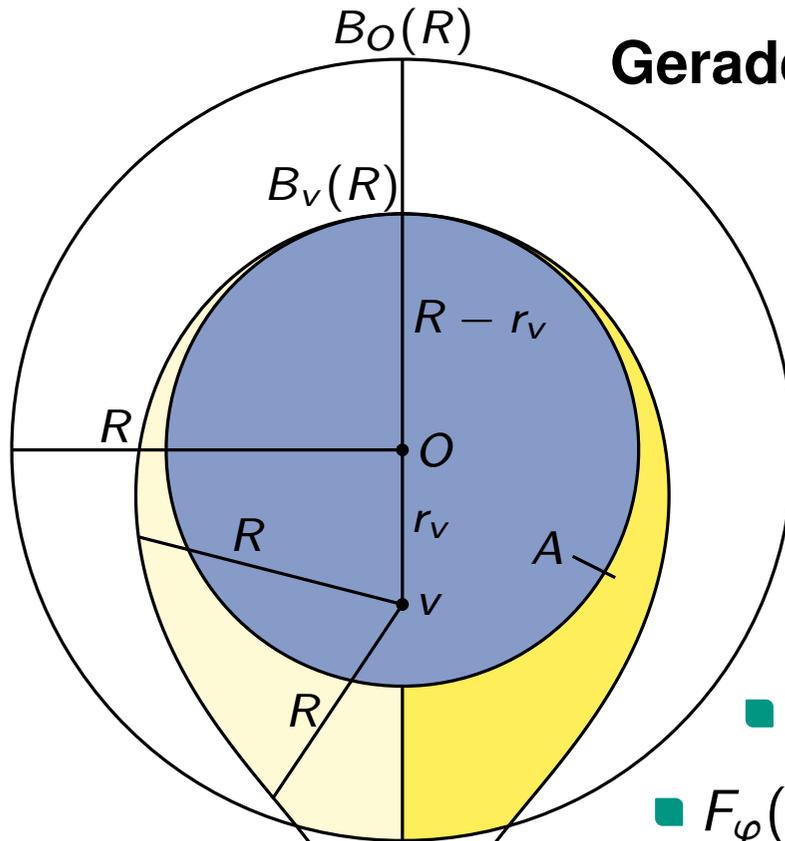
$$\mu(A) \leq \sum_{x_r=R-r_v}^{R-\Delta r} \sum_{x_\varphi=0}^{\theta(r_v, x_r)} \mathbf{P}[x_\varphi \leq \varphi \leq x_\varphi + \Delta\varphi] \cdot \mathbf{P}[x_r \leq r \leq x_r + \Delta r]$$

Summe über Vielfache von Δr bzw. $\Delta\varphi$

$$\rightarrow f_\varphi(x_\varphi) \cdot \Delta\varphi \quad \rightarrow f_r(x_r) \cdot \Delta x$$

- lasse Δr und $\Delta\varphi$ gegen 0 gehen
- $(F(x + \Delta x) - F(x))/\Delta x$ wird zur Ableitung $f(x)$
- $\sum f(x) \cdot \Delta x$ wird zum Integral $\int f(x) dx$

Das Maß von A: Vorbereitung



Gerade gesehen

$$\mu(A) = \int_{R-r_v}^R \int_0^{\theta(r_v, x_r)} f_\varphi(x_\varphi) dx_\varphi \cdot f_r(x_r) dx_r$$

- $f_\varphi(x)$ und $f_r(x)$ sind die Ableitungen der Verteilungsfunktionen $F_\varphi(x)$ bzw. $F_r(x)$
- $\theta(r_1, r_2)$: max. Winkeldifferenz zw. Punktepaar mit Abstand $\leq R$ und Radien r_1, r_2

Ableitungen der Verteilungsfunktionen

- $F_r(x) \approx e^{-(R-x)} \rightarrow f_r(x) \approx e^{-(R-x)}$
- genauer: $F_r(x) = \frac{\cosh(x)-1}{\cosh(R)-1} \Rightarrow f_r(x) = \frac{\sinh(x)}{\cosh(R)-1}$
- $F_\varphi(x) = \frac{x}{2\pi} \Rightarrow f_\varphi(x) = \frac{1}{2\pi}$

■ $f(x)$ heißt **Dichtefunktion**

(Probability Density Function (PDF))

■ quasi das kontinuierliche Äquivalent zur Wkt.

Maximale Winkeldifferenz

■ $\theta(r_1, r_2) \approx 2e^{(R-r_1-r_2)/2}$ (folgt aus Umstellen der Distanzfunktion)

$$F_r(x) \approx e^{-(R-x)}$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

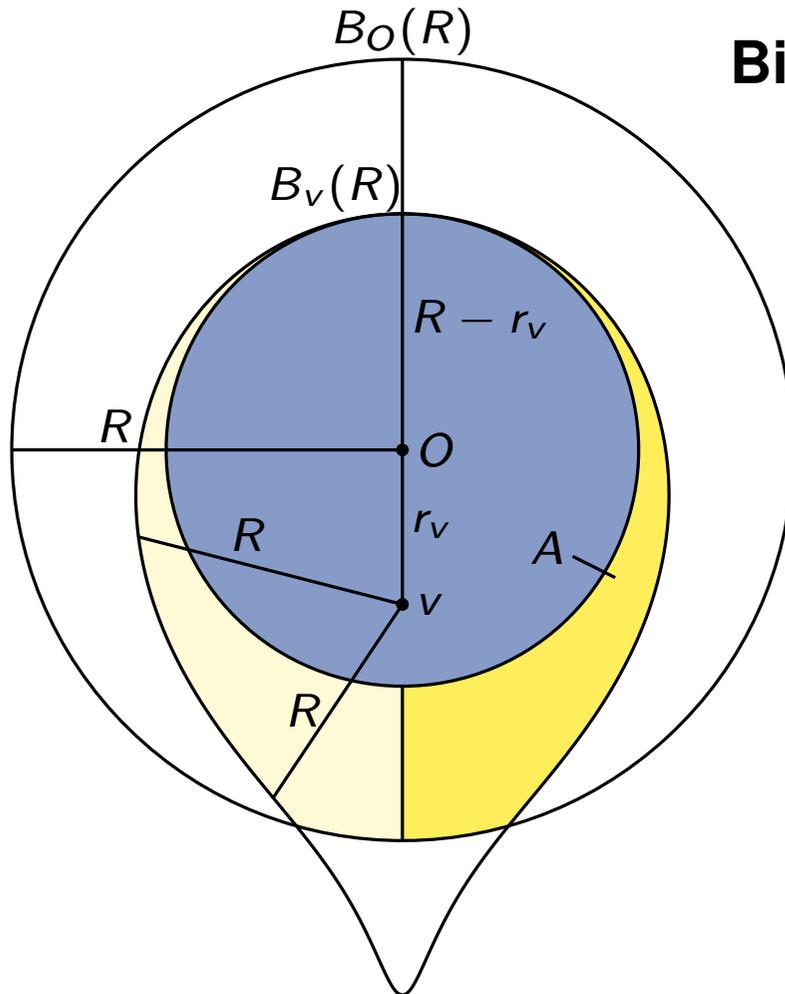
$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

Das Maß von A : Berechnung des Integrals

$$\begin{aligned}
 \mu(A) &= \int_{R-r_v}^R \int_0^{\theta(r_v, x_r)} f_\varphi(x_\varphi) dx_\varphi \cdot f_r(x_r) dx_r \\
 &= \int_{R-r_v}^R \frac{\theta(r_v, x_r)}{2\pi} \cdot f_r(x_r) dx_r \\
 &\approx \int_{R-r_v}^R \frac{2e^{(R-r_v-x_r)/2}}{2\pi} \cdot e^{-(R-x_r)} dx_r \\
 &= \frac{e^{-R/2-r_v/2}}{\pi} \int_{R-r_v}^R e^{x_r/2} dx_r = \frac{e^{-R/2-r_v/2}}{\pi} \left[2e^{x_r/2} \right]_{R-r_v}^R \\
 &= \frac{2e^{-R/2-r_v/2}}{\pi} \left(e^{R/2} - e^{(R-r_v)/2} \right) \\
 &= \frac{2e^{-r_v/2}}{\pi} - \frac{2e^{-r_v}}{\pi}
 \end{aligned}$$

$$\begin{aligned}
 f_r(x) &\approx e^{-(R-x)} \\
 f_\varphi(x) &= \frac{1}{2\pi} \\
 \theta(r_1, r_2) &\approx 2e^{(R-r_1-r_2)/2}
 \end{aligned}$$

Der Erwartet Knotengrad



Bisher ausgerechnet

$$\mu(B_O(R - r_v)) \approx e^{-r_v}$$

$$\mu(A) \approx \frac{2e^{-r_v/2}}{\pi} - \frac{2e^{-r_v}}{\pi}$$

Damit erhalten wir

$$\begin{aligned} \mathbb{E}[\deg(v) \mid r_v] &= (n - 1) \cdot \mu(B_v(R) \cap B_O(R)) \\ &= (n - 1) \cdot (2\mu(A) + \mu(B_O(R - r_v))) \\ &\approx n \cdot \left(\frac{4e^{-r_v/2}}{\pi} - \frac{4e^{-r_v}}{\pi} + e^{-r_v} \right) \\ &\approx n \cdot \frac{4e^{-r_v/2}}{\pi} \end{aligned}$$

Anmerkung: erwarteter Durchschnittsgrad

Was ist der erwartete Durchschnittsgrad des Graphen?

- erwarteter Durchschnittsgrad = erwarteter Grad eines zufälligen Knotens

- bisher gesehen: $\mathbf{E}[\text{deg}(v) \mid r_v] \approx \frac{4n}{\pi} e^{-r_v/2}$

- wäre der Radius diskret (d.h. es gibt abzählbare Menge \mathcal{R} von Radien):

$$\mathbf{E}[\text{deg}(v)] = \sum_{x \in \mathcal{R}} \mathbf{P}[r = x] \cdot \mathbf{E}[\text{deg}(v) \mid r_v = x]$$

- r ist aber kontinuierlich:

$$\mathbf{E}[\text{deg}(v)] = \int_0^R f_r(x) \cdot \mathbf{E}[\text{deg}(v) \mid r_v = x] dx$$

- rechnet man das aus, so erhält man: $\mathbf{E}[\text{deg}(v)] \approx \frac{8n}{\pi} e^{-R/2}$

Was ist ein guter Wert für R ?

- setze $R = 2 \log n + C$ für eine Konstante C

- dann gilt: $\mathbf{E}[\text{deg}(v)] \approx \frac{8}{\pi} e^{-C/2}$

- der Durchschnittsgrad ist also konstant und kann durch den Parameter C kontrolliert werden

Anmerkung: Gradverteilung

Wie viele Knoten mit welchem Grad?

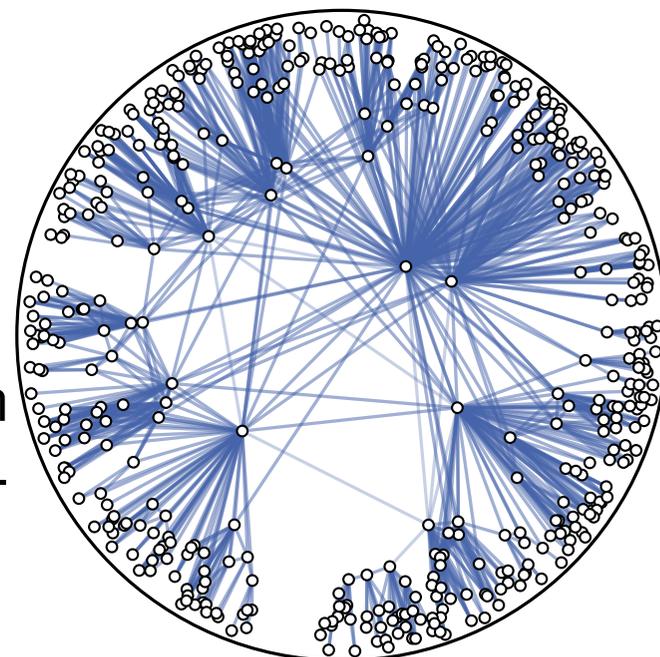
- bisher gesehen: $\mathbf{E}[\text{deg}(v) \mid r_v] \approx \frac{4n}{\pi} e^{-r_v/2}$
- die Knotengrade schrumpfen also rapide mit wachsendem Radius
- außerdem haben viel mehr Knoten großen Radius
- also: viele Knoten mit kleinem Grad, wenige mit großem
- tatsächlich erhält man ein power law mit Exponent 3

Gehen auch andere Exponenten?

- angepasste Verteilungsfunktion:

$$f(r) = \frac{\alpha \sinh(\alpha r)}{\cosh(\alpha R) - 1} \text{ statt } f(r) = \frac{\sinh(r)}{\cosh(R) - 1}$$

- $\alpha \in (1/2, 1] \Rightarrow$ power-law Exponent $\beta \in (2, 3]$
- $\beta \in (2, 3]$ ist gerade der interessanteste Bereich
- alle Berechnungen, die wir heute gesehen haben, funktionieren analog mit α



Zusammenfassung

Heute gesehen

- geometrische Zufallsgraphen, Euklidisch und hyperbolisch
- Berechnung grundlegender Eigenschaften: hier erwartete Knotengrade
- Euklidisch/hyperbolisch: homogene/heterogene Gradverteilung
- Geometrie sorgt für Lokalität/Clustering

Hyperbolische Zufallsgraphen: Ausblick

Parameter

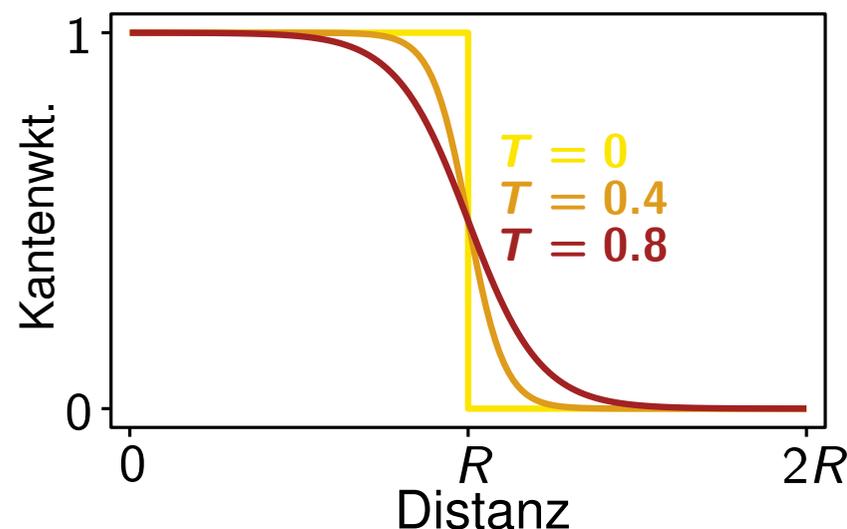
- Knotenzahl n , Kantenanzahl m (via R)
- power-law Exponent β
- Temperatur T

Strukturelle Eigenschaften

- power-law Gradverteilung ($k^{-\beta}$)
- Clusterkoeffizient > 0
- logarithmischer Durchmesser
- Komponentenstruktur: eine große Komponente
- Verteilung von Cliques
- Separatoren & Baumweite (sublinear)

Algorithmen

- effiziente Generierung: $O(n)$
- bessere Algorithmen dank kleinen Separatoren/Baumweite
- bidirektionale Breitensuche braucht $o(n)$
- Vertex Cover in polynomieller Zeit



Literaturhinweise

- **Random Geometric Graphs** (2003)
 Mathew Penrose doi.org/10.1093/acprof:oso/9780198506263.001.0001
- **Hyperbolic geometry of complex networks** (2010)
 Krioukov, Papadopoulos, Kitsak, Vahdat, Boguñá https://doi.org/10.1103/PhysRevE.82.036106
- **Random Hyperbolic Graphs: Degree Sequence and Clustering** (2012)
 Luca Gugelmann, Konstantinos Panagiotou, Ueli Peter https://doi.org/10.1007/978-3-642-31585-5_51
- **The diameter of KPKVB random graphs** (2019)
 Tobias Müller, Merlijn Staps https://doi.org/10.1017/apr.2019.23
- **On the Largest Component of a Hyperbolic Model of Complex Networks** (2015)
 Michel Bode, Nikolaos Fountoulakis, Tobias Müller https://doi.org/10.37236/4958
- **Hyperbolic Random Graphs: Separators and Treewidth** (2016)
 Thomas Bläsius, Tobias Friedrich, Anton Krohmer https://doi.org/10.4230/LIPIcs.ESA.2016.15
- **Sampling Geometric Inhomogeneous Random Graphs in Linear Time** (2017)
 Karl Bringmann, Ralph Keusch, Johannes Lengler https://doi.org/10.4230/LIPIcs.ESA.2017.20
- **Efficiently Generating Geometric Inhomogeneous and Hyperbolic Random Graphs** (2019)
 Bläsius, Friedrich, Katzmann, Meyer, Penschuck, Weyand https://doi.org/10.4230/LIPIcs.ESA.2019.21
- **Efficient Shortest Paths in Scale-Free Networks with Underlying Hyperbolic Geometry** (2018)
 Bläsius, Freiberger, Friedrich, Katzmann, Montenegro-Retana, Thieffry https://doi.org/10.4230/LIPIcs.ICALP.2018.20
- **Solving Vertex Cover in Polynomial Time on Hyperbolic Random Graphs** (2020)
 Bläsius, Fischbeck, Friedrich, Katzmann https://doi.org/10.4230/LIPIcs.STACS.2020.25