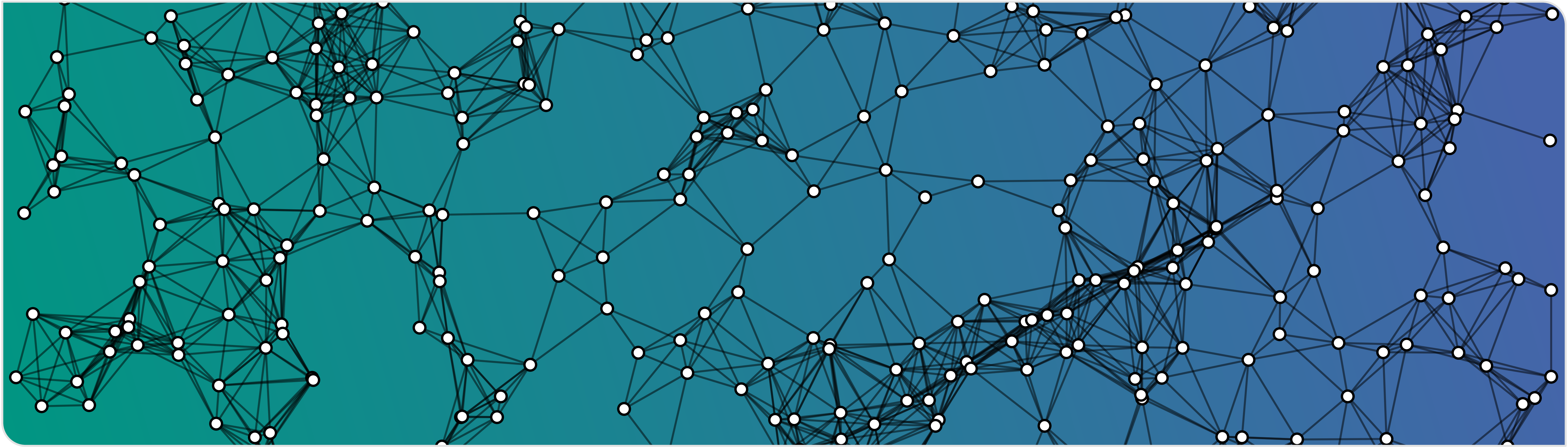
Karlsruher Institut für Technologie

Praktikum – Beating the Worst Case

Jean-Pierre von der Heydt und Marcus Wilhelm | 22.11.2023



Fragen zum Übungsblatt 1

Gradverteilung

- Wie unterscheiden sich die Gradverteilungen?
  - Gib es hierfür ein Maß?
- [Wie] habt ihr die Verteilungen visualisiert?

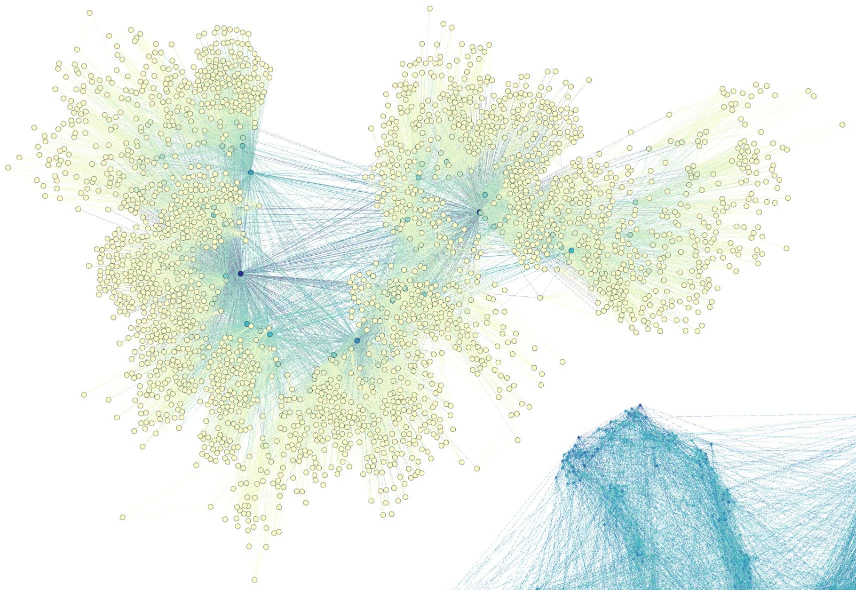
Lokalität

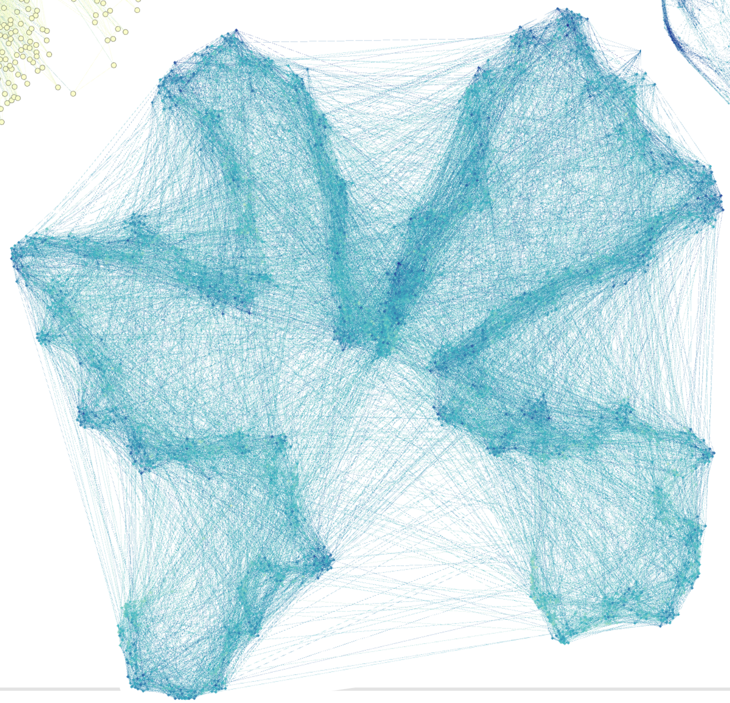
- Wie kann man Lokalität messen?
- Welche Maße habt ihr verwendet?

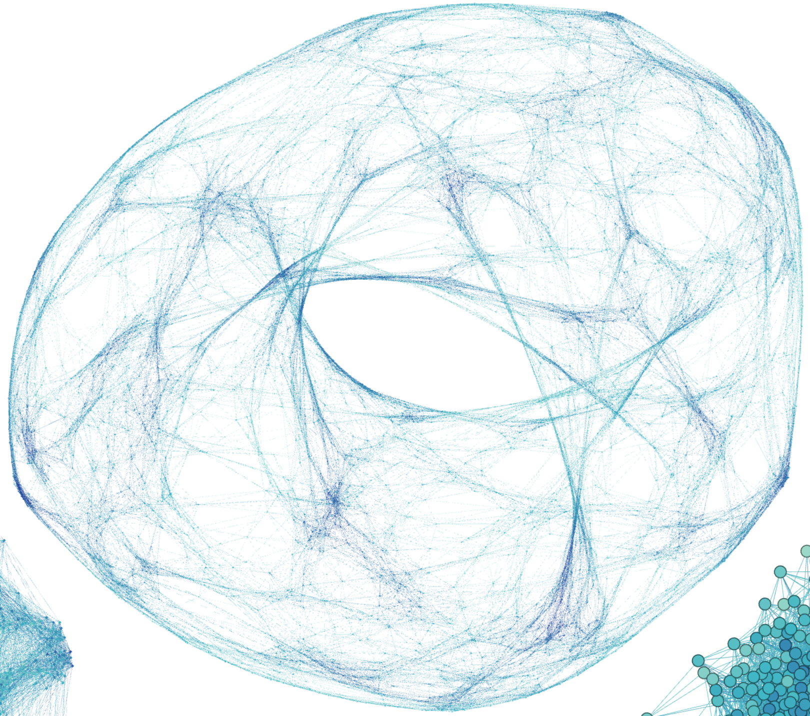
Allgemein

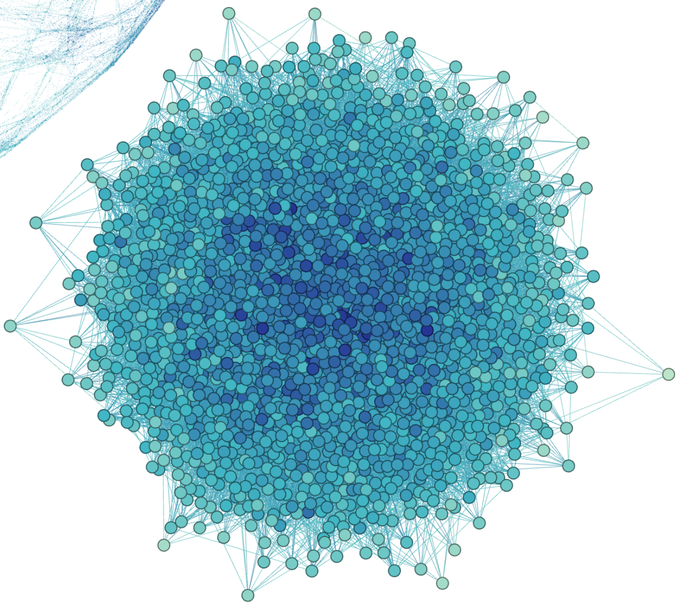
- Mit welchen Maßen lassen sich die vier Netzwerkklassen unterscheiden?
- [Wo / Wie] habt ihr Recherchiert?
- Habt ihr eine Vermutung, wie die Graphen erzeugt wurden?

Netzwerke aus Datensatz

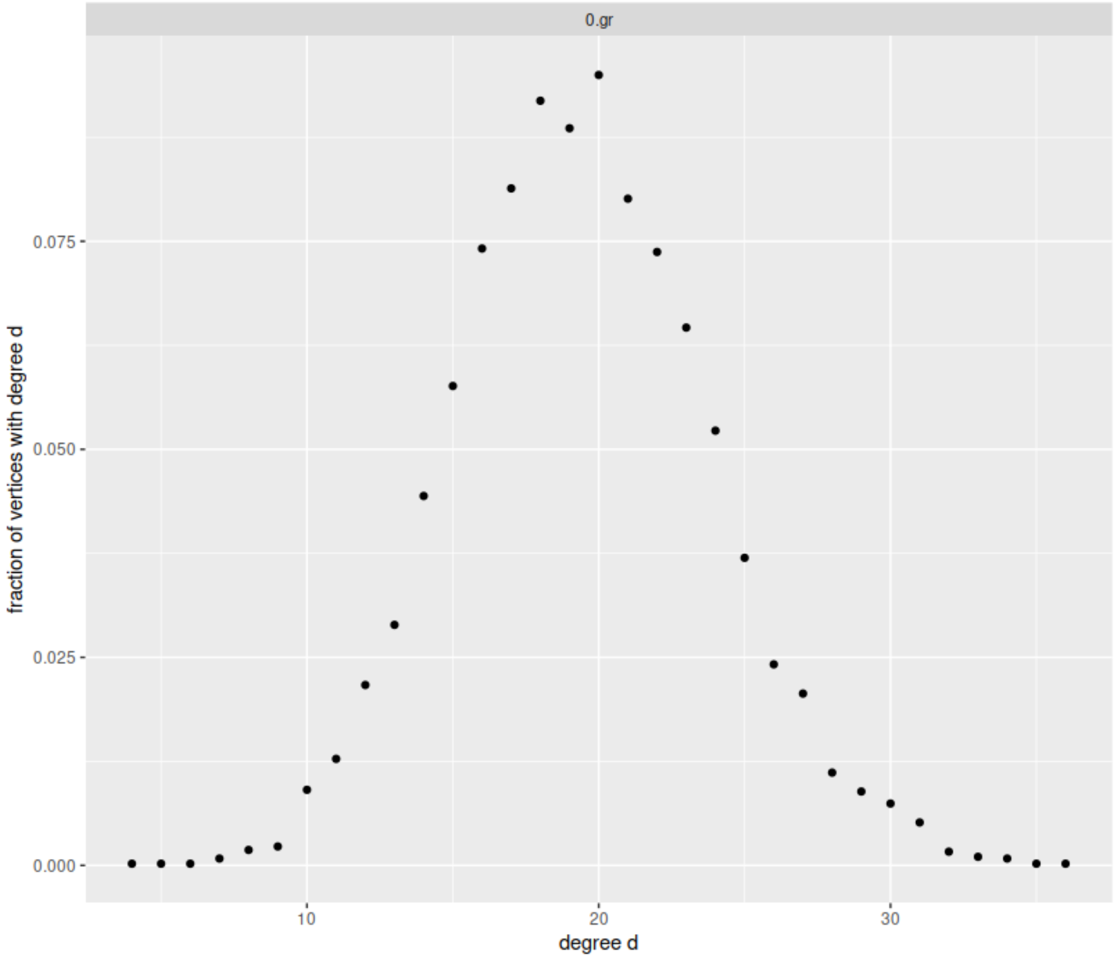






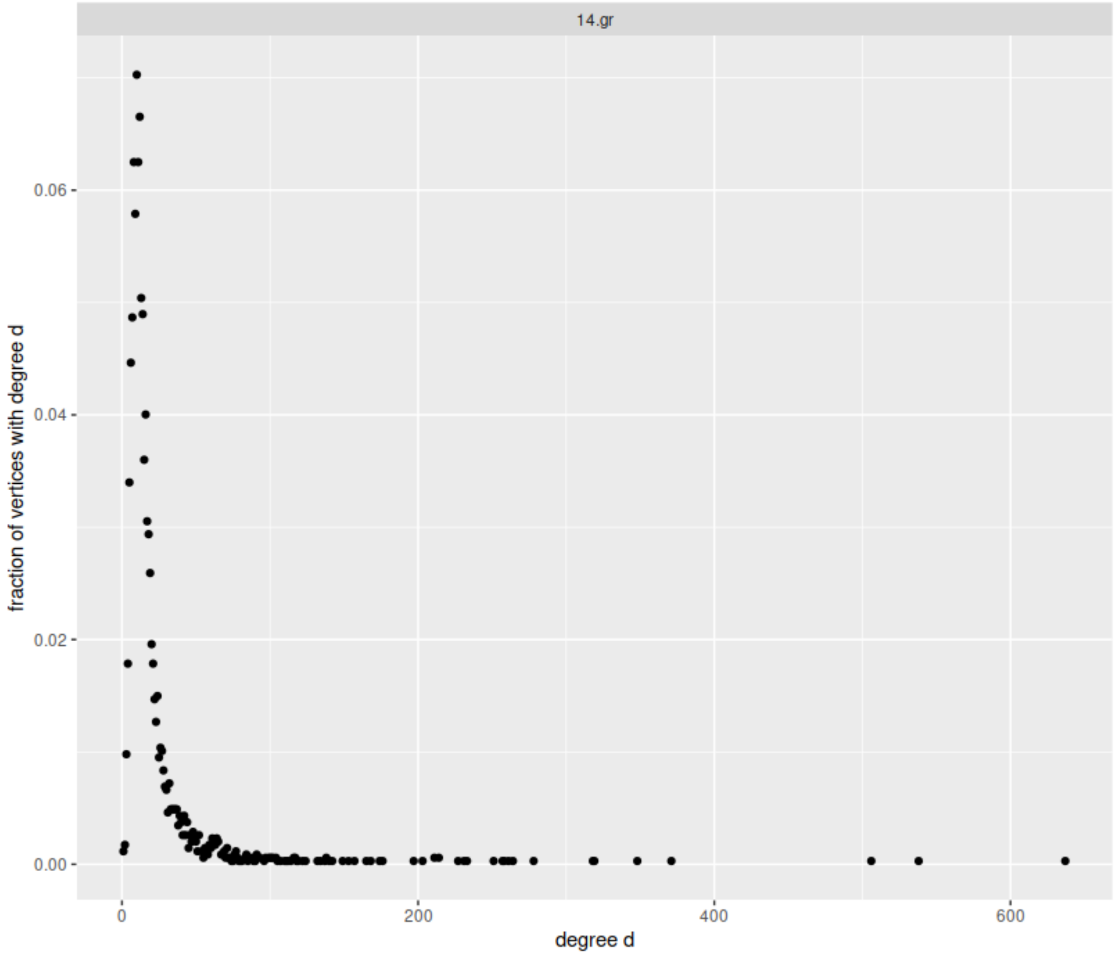


Homogene Gradverteilung



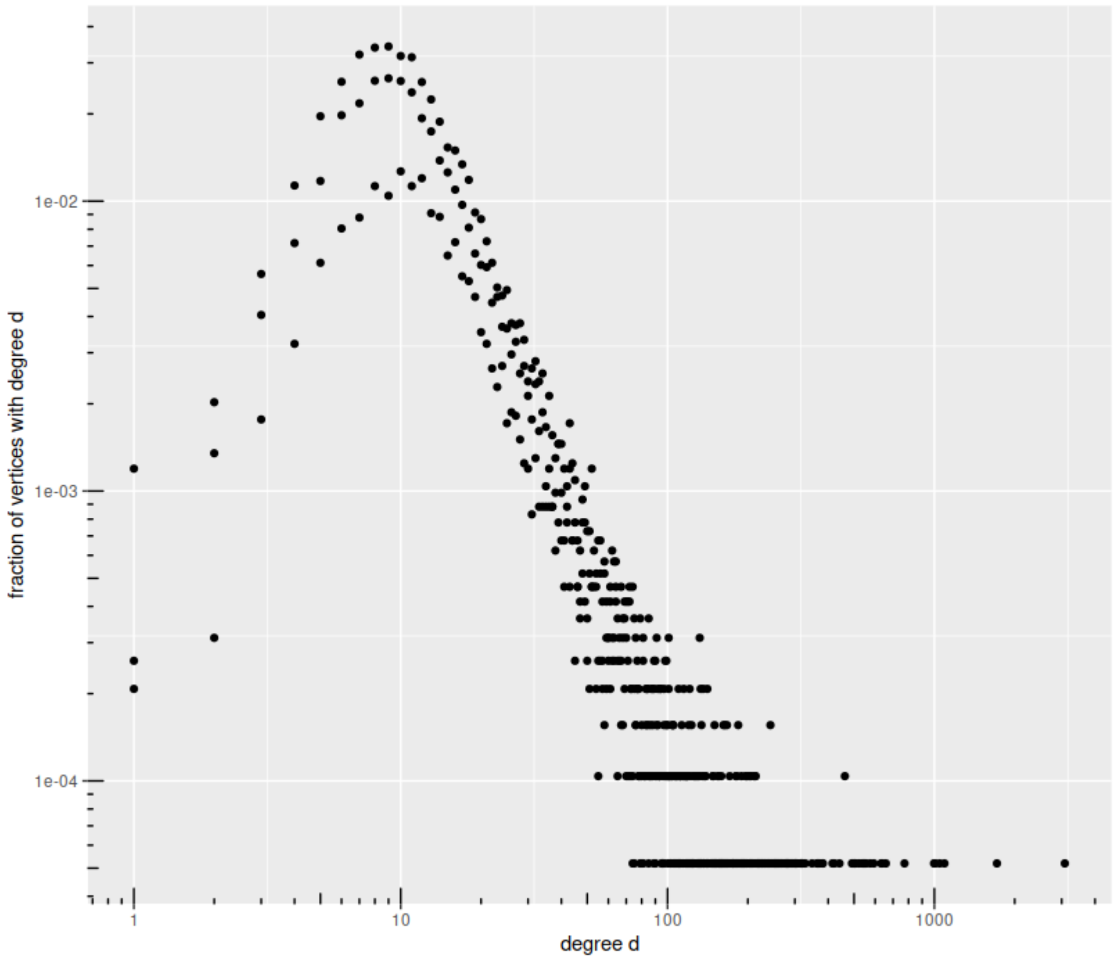
n: 4843
mean: 19.4
variance: 18.7
stdev: 4.33

Heterogene Gradverteilung



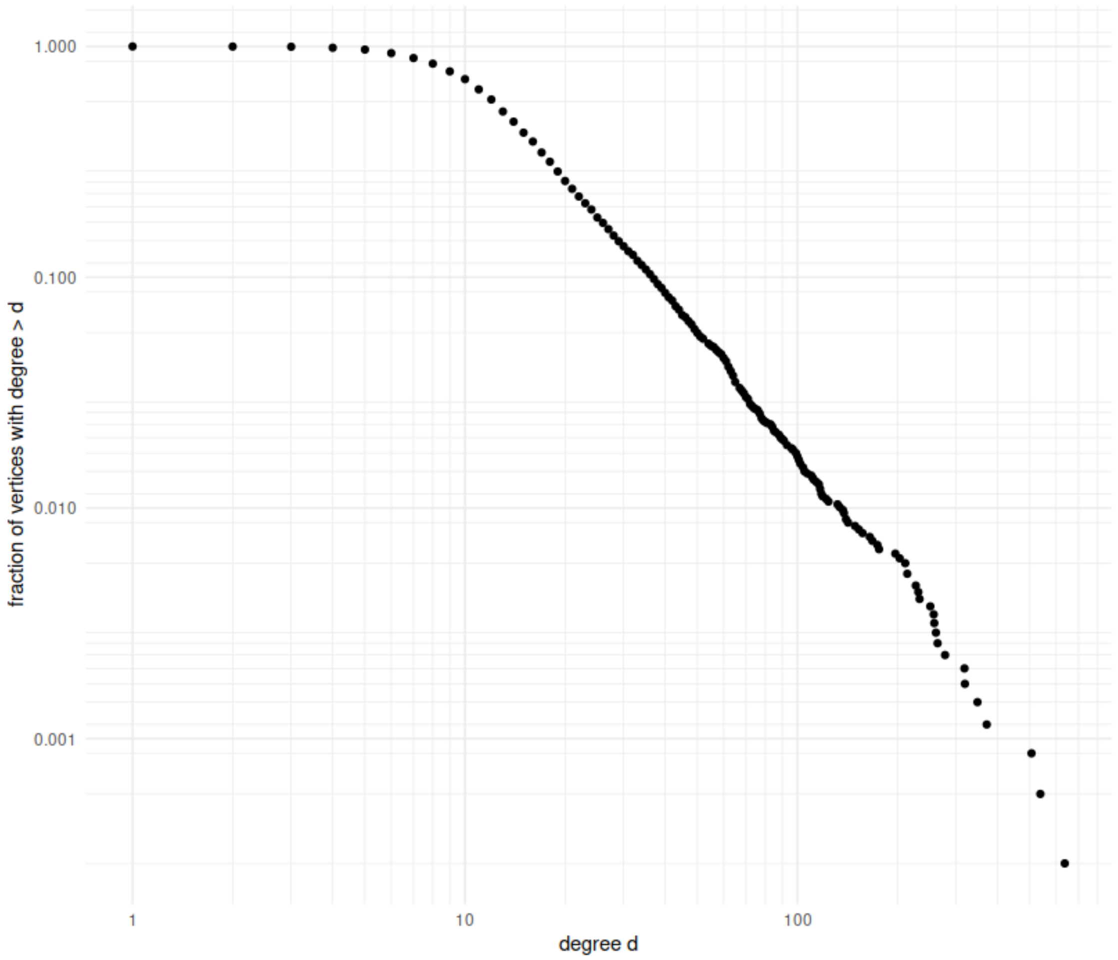
n: 3472
mean: 20.0
variance: 891
stdev: 29.8

Heterogene Gradverteilung



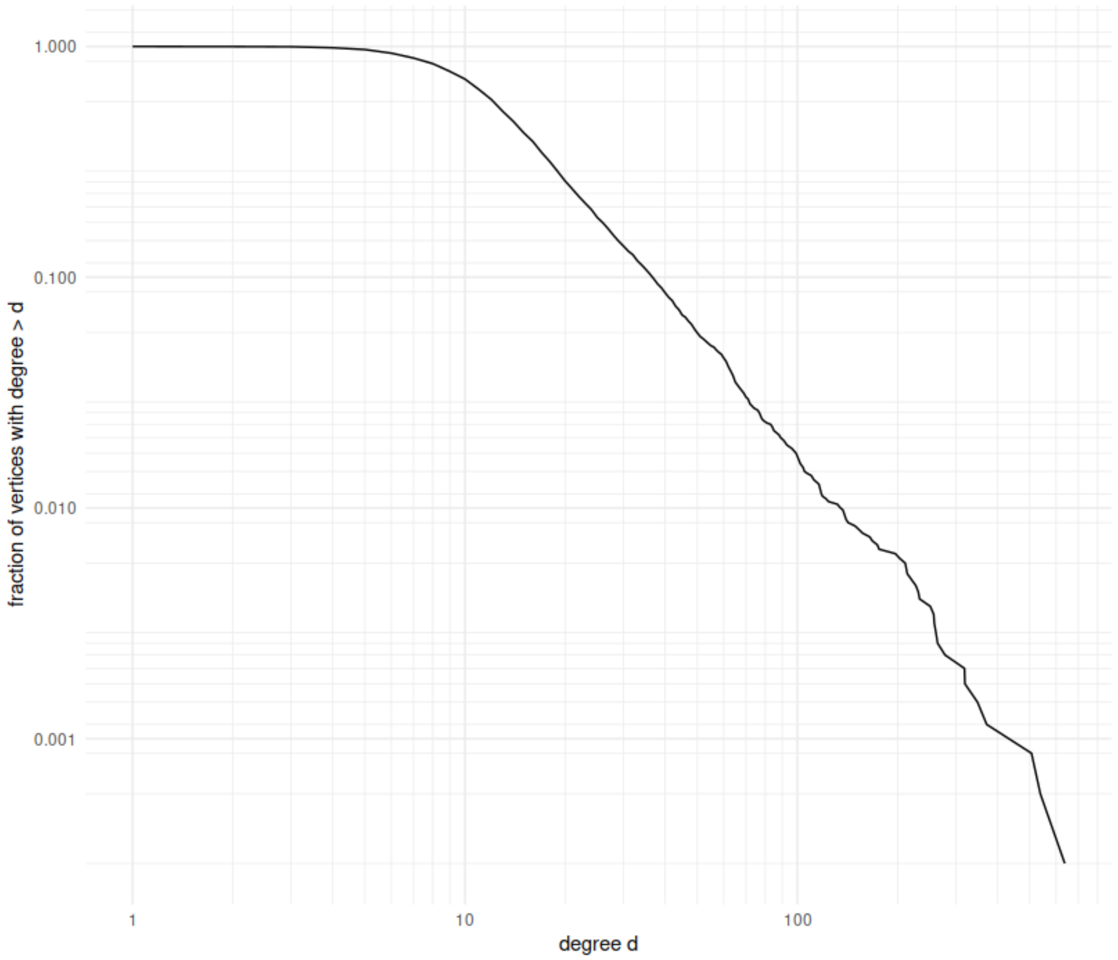
n: 3472
mean: 20.0
variance: 891
stdev: 29.8

Heterogene Gradverteilung



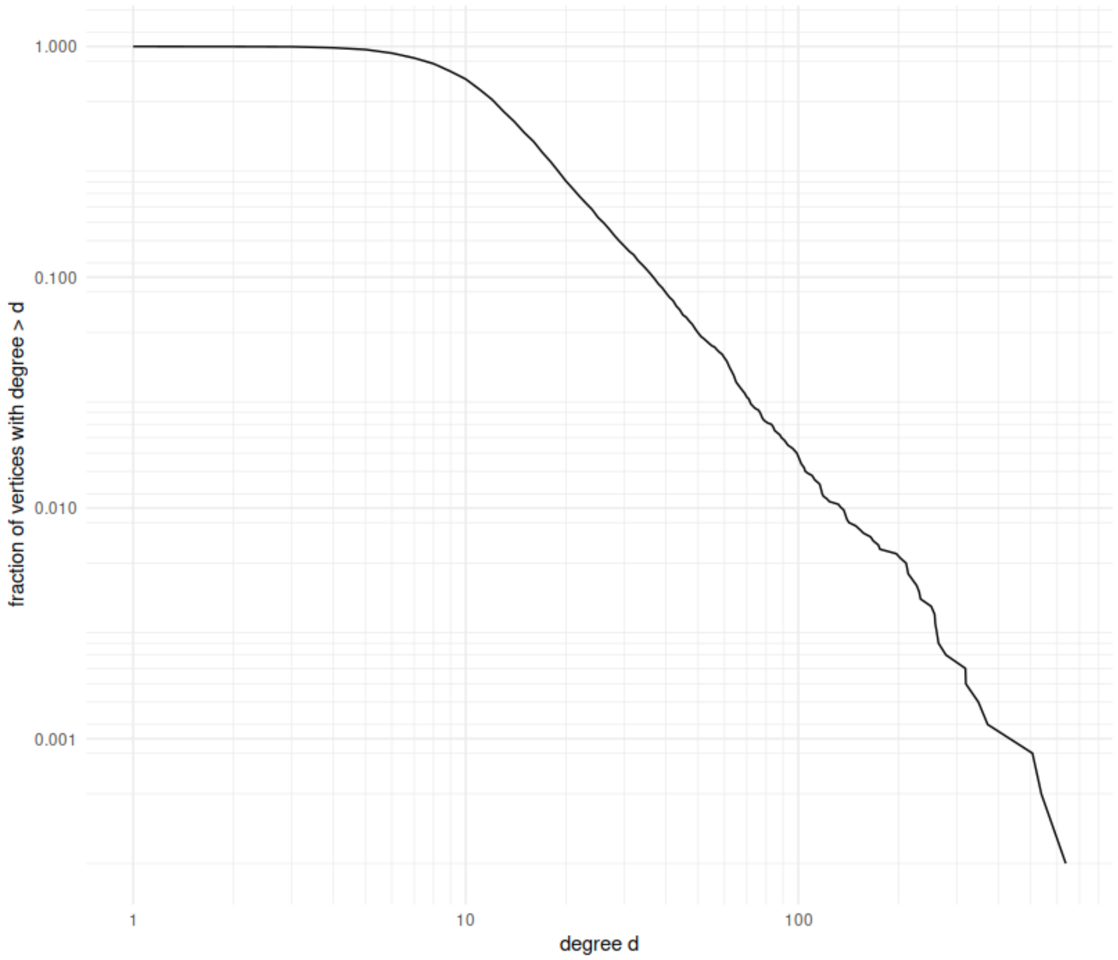
n: 3472
mean: 20.0
variance: 891
stdev: 29.8

Heterogene Gradverteilung

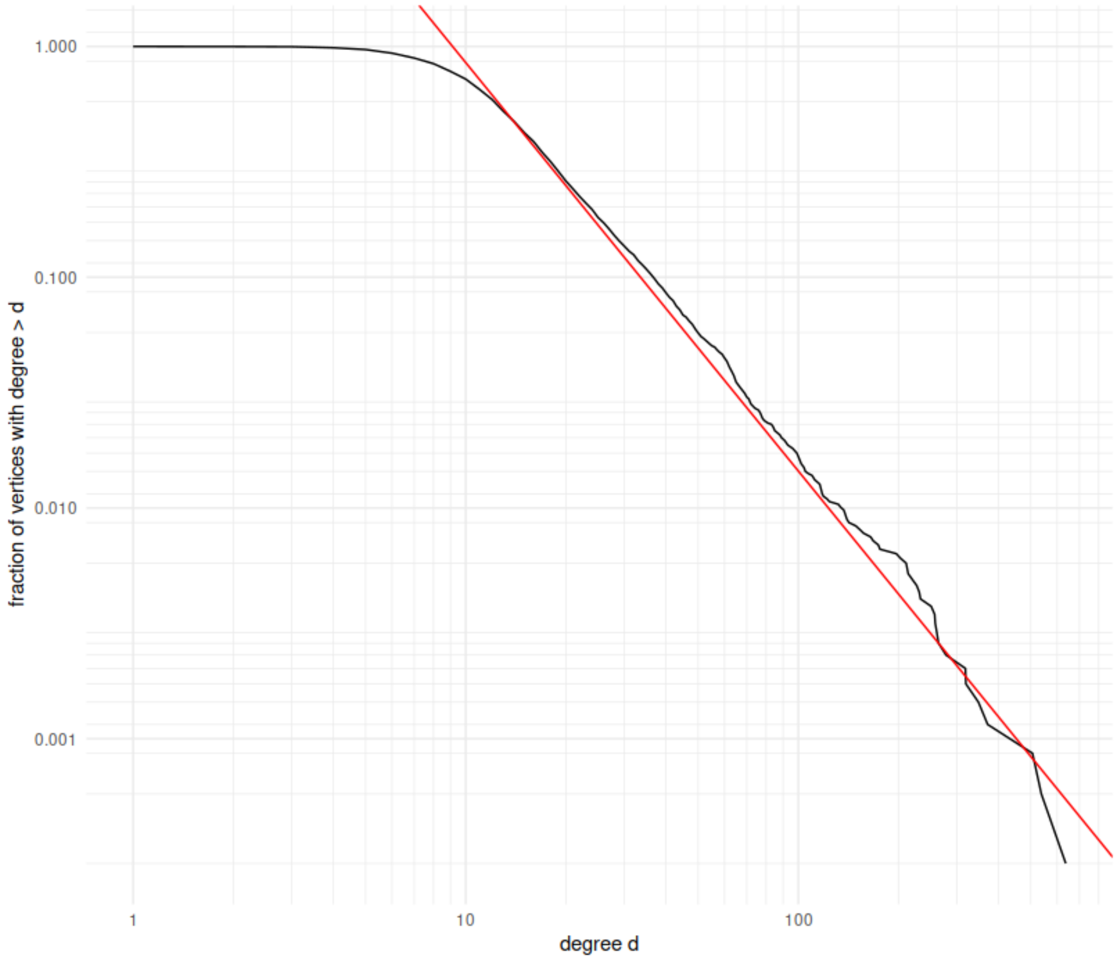


n: 3472
mean: 20.0
variance: 891
stdev: 29.8

Heterogene Gradverteilung – Details



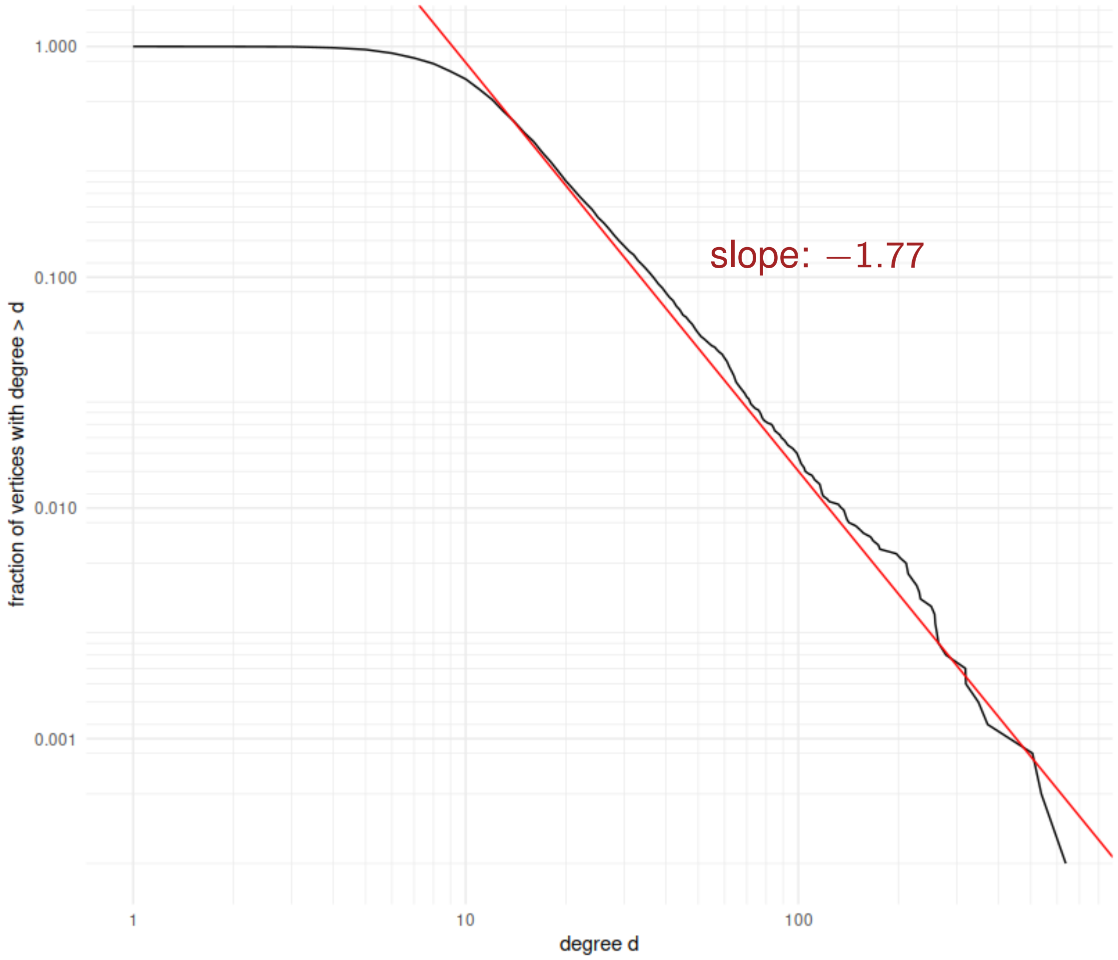
Heterogene Gradverteilung – Details



Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$

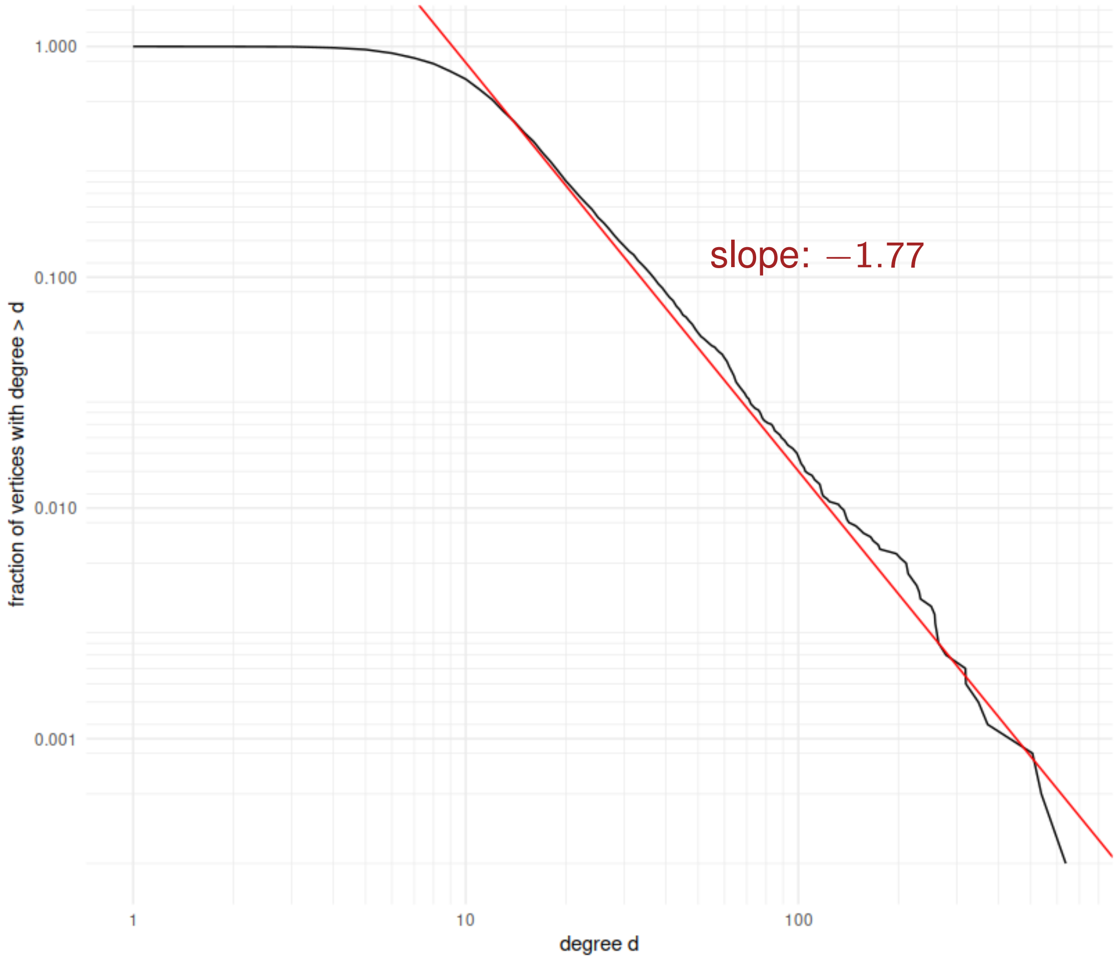


Heterogene Gradverteilung – Details

Karlsruher Institut für Technologie

Gerade Linie im log-log plot

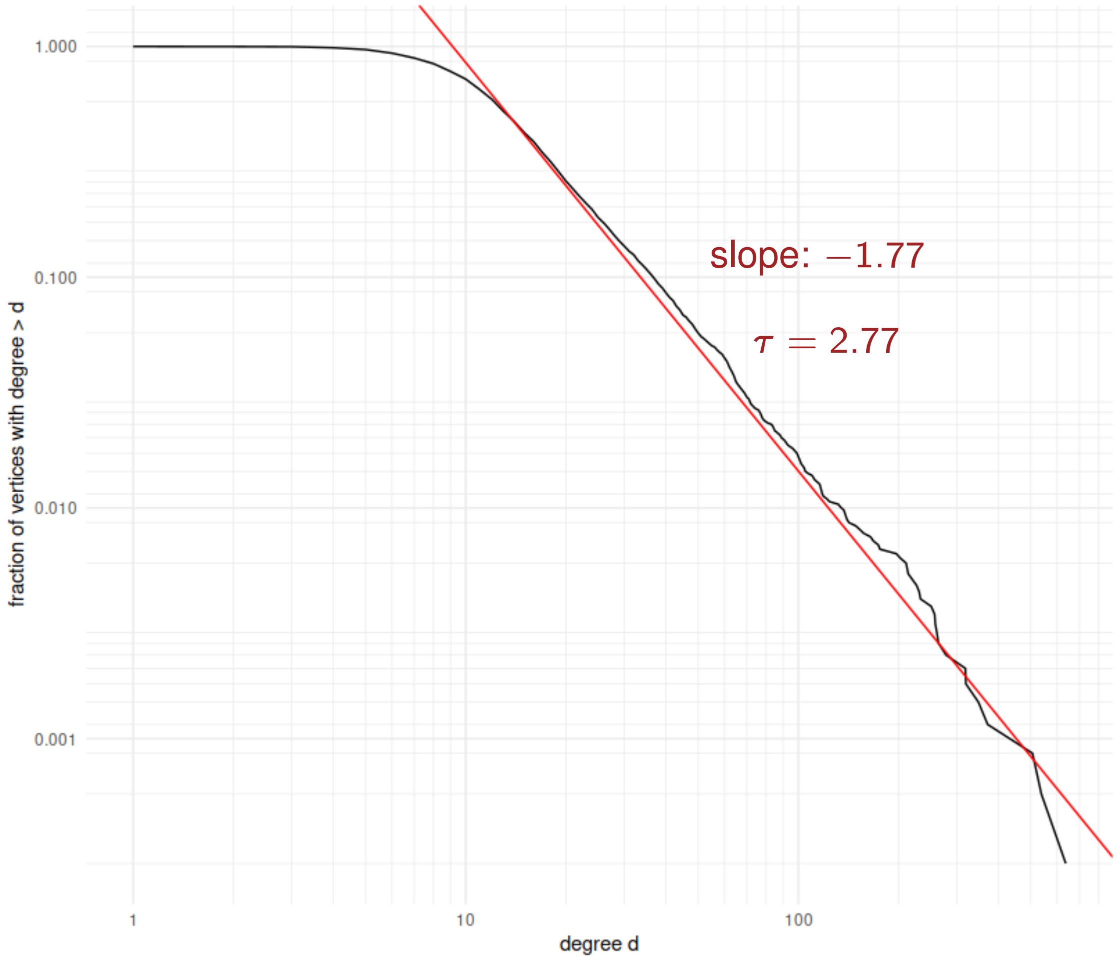
- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$



Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$



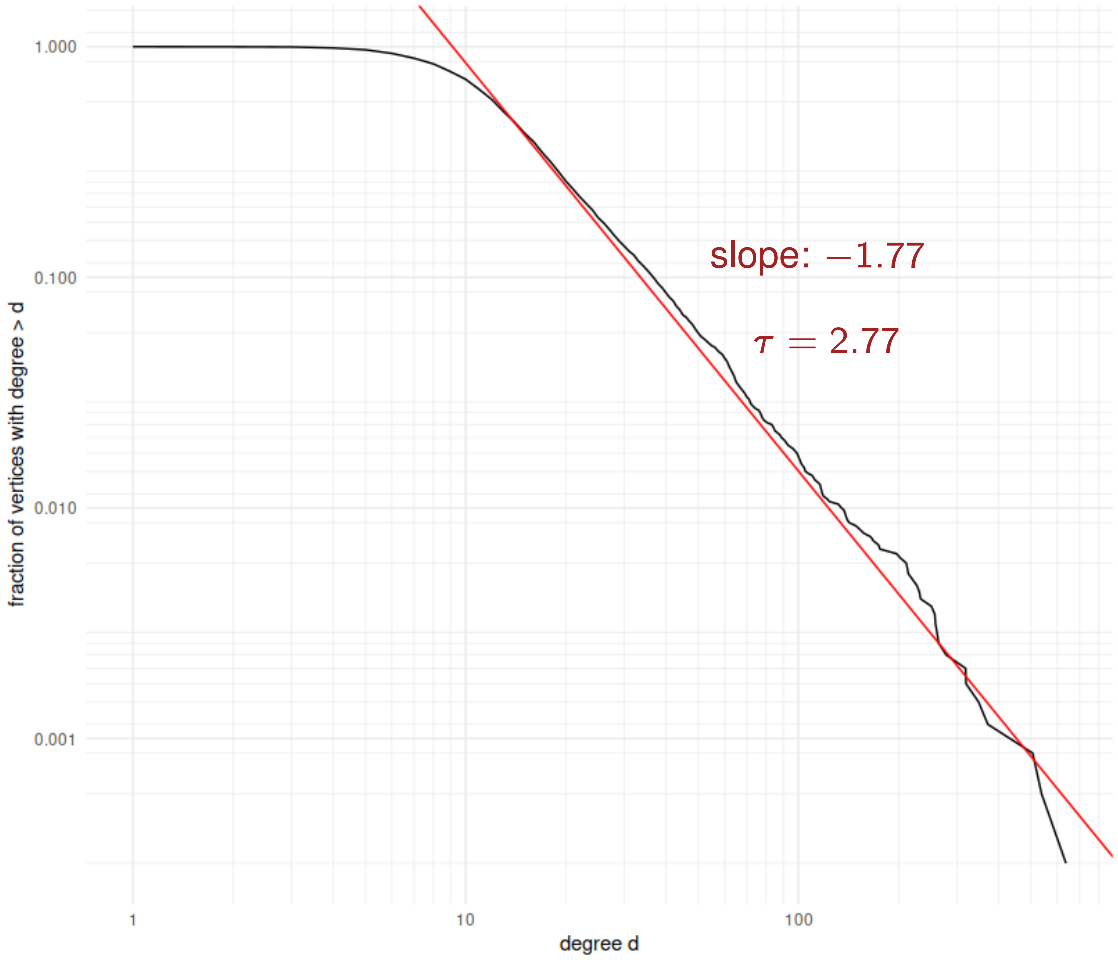
Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$

Definition: power-law distribution

- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent



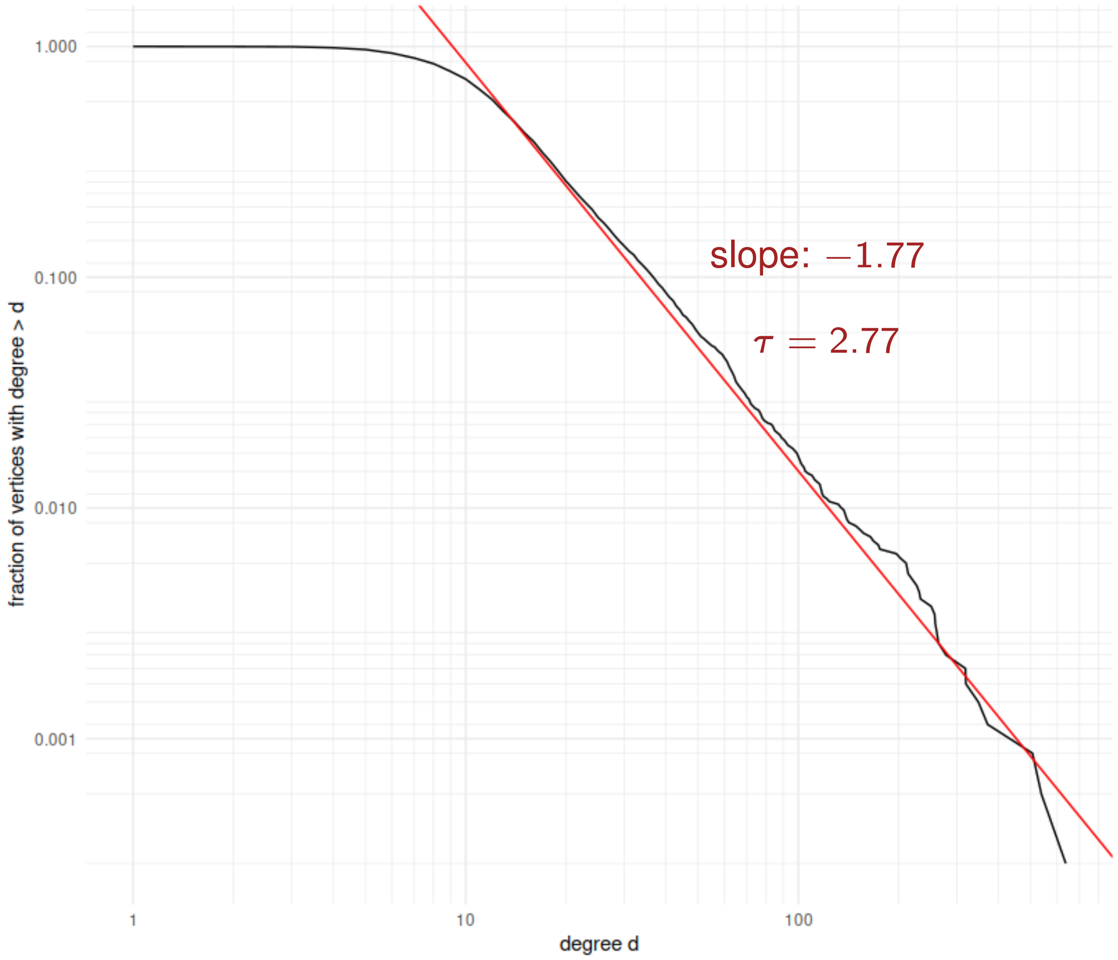
Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$

Definition: power-law distribution

- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent
- üblicherweise: $\tau \in (2, 3)$



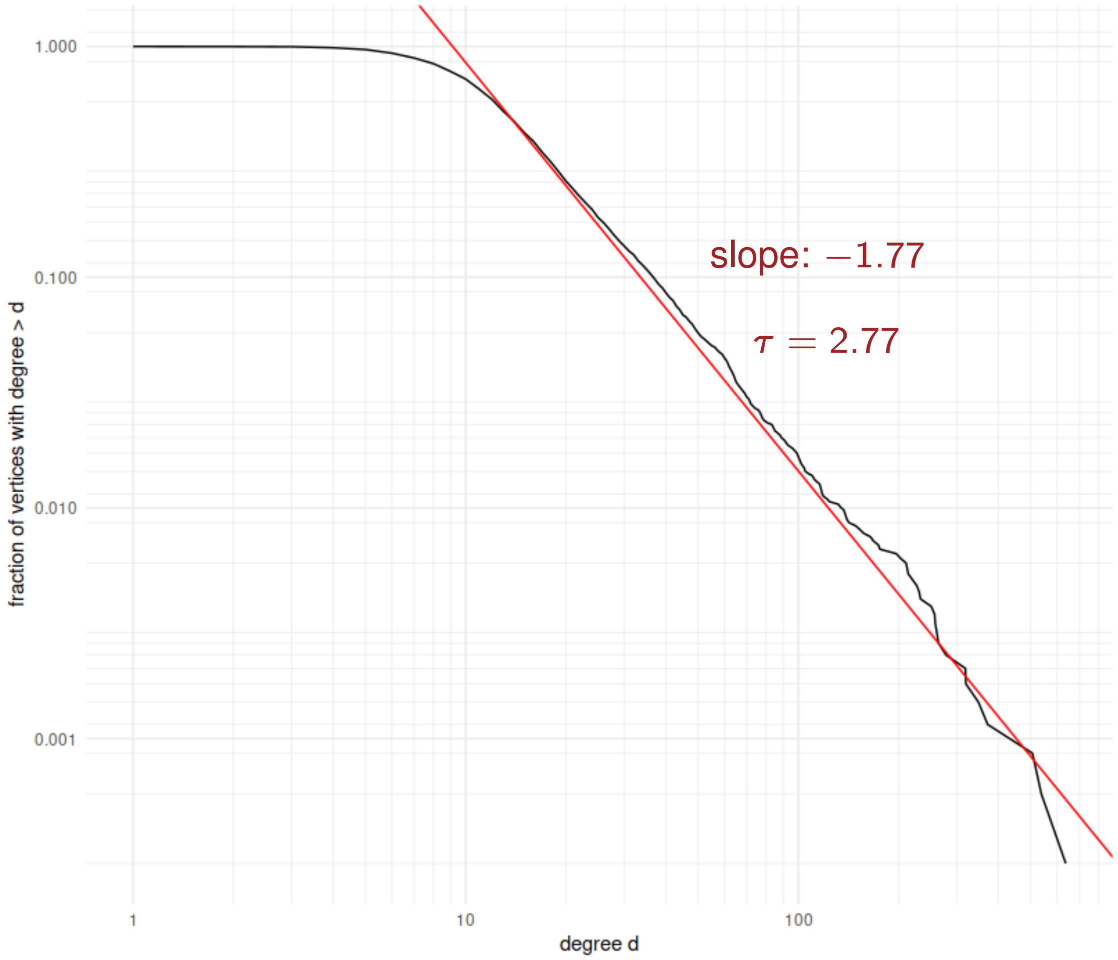
Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$

Definition: power-law distribution

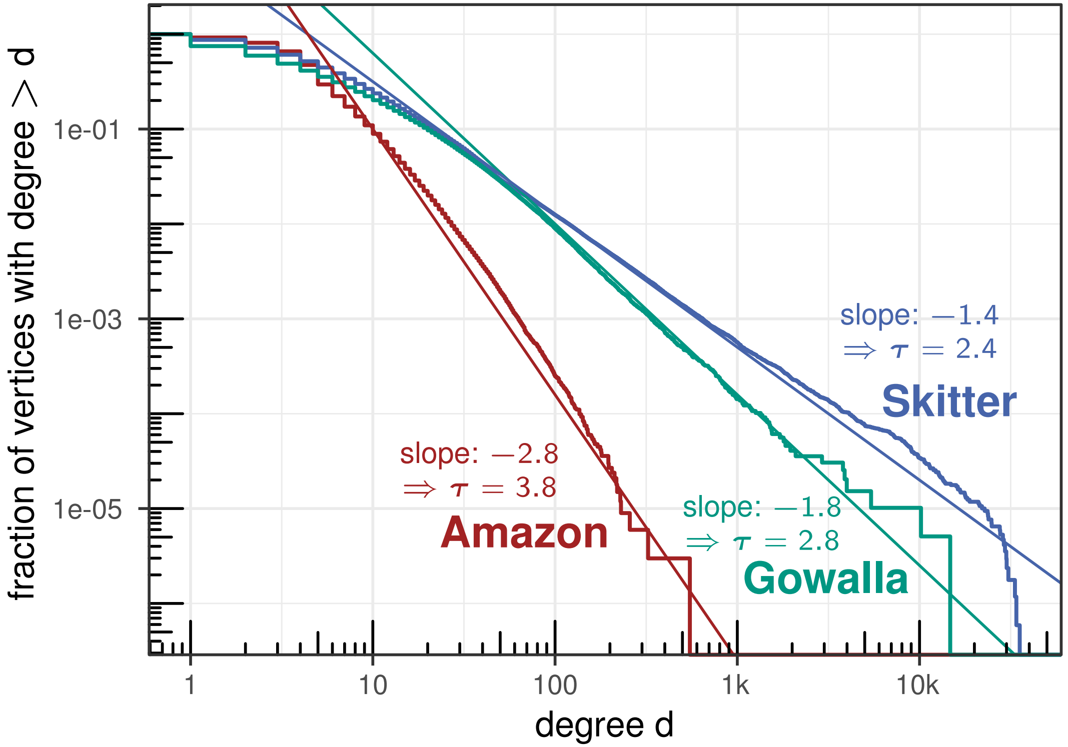
- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent
- üblicherweise: $\tau \in (2, 3)$
  - $\tau > 2 \rightarrow$ constant average degree
  - $\tau \in (2, 3] \rightarrow$ variance increases with n
  - $\tau > 3 \rightarrow$ constant variance



Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$



Definition: power-law distribution

- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent
- üblicherweise: $\tau \in (2, 3)$
  - $\tau > 2 \rightarrow$ constant average degree
  - $\tau \in (2, 3] \rightarrow$ variance increases with n
  - $\tau > 3 \rightarrow$ constant variance

Heterogene Gradverteilung – Details

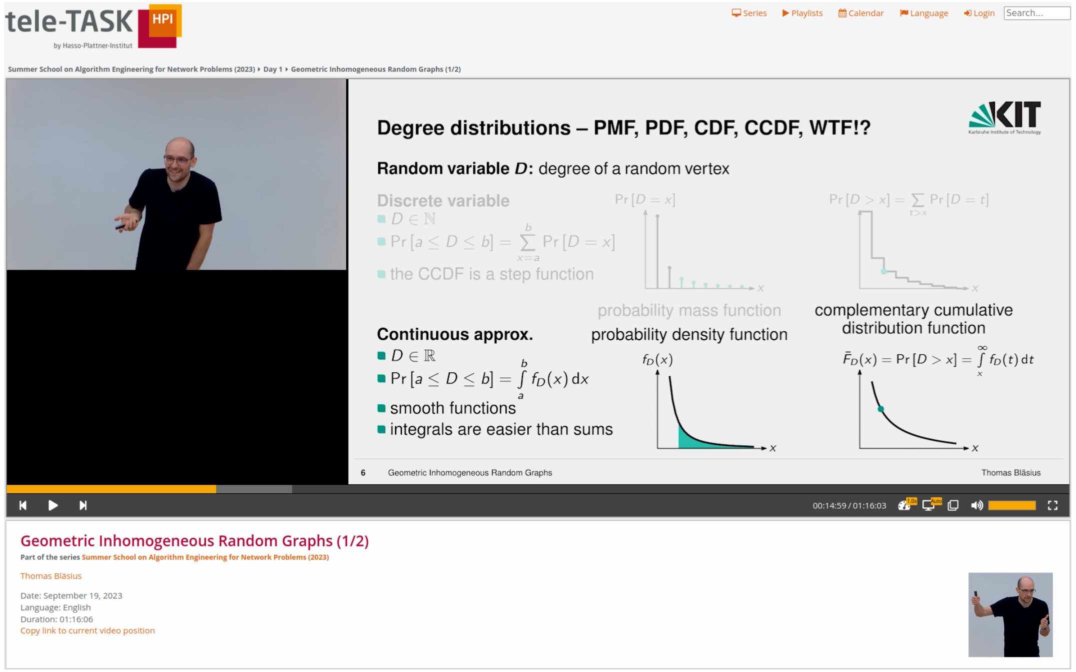
Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$

Definition: power-law distribution

- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent
- üblicherweise: $\tau \in (2, 3)$
  - $\tau > 2 \rightarrow$ constant average degree
  - $\tau \in (2, 3] \rightarrow$ variance increases with n
  - $\tau > 3 \rightarrow$ constant variance

weiterführend: Vortrag von Thomas

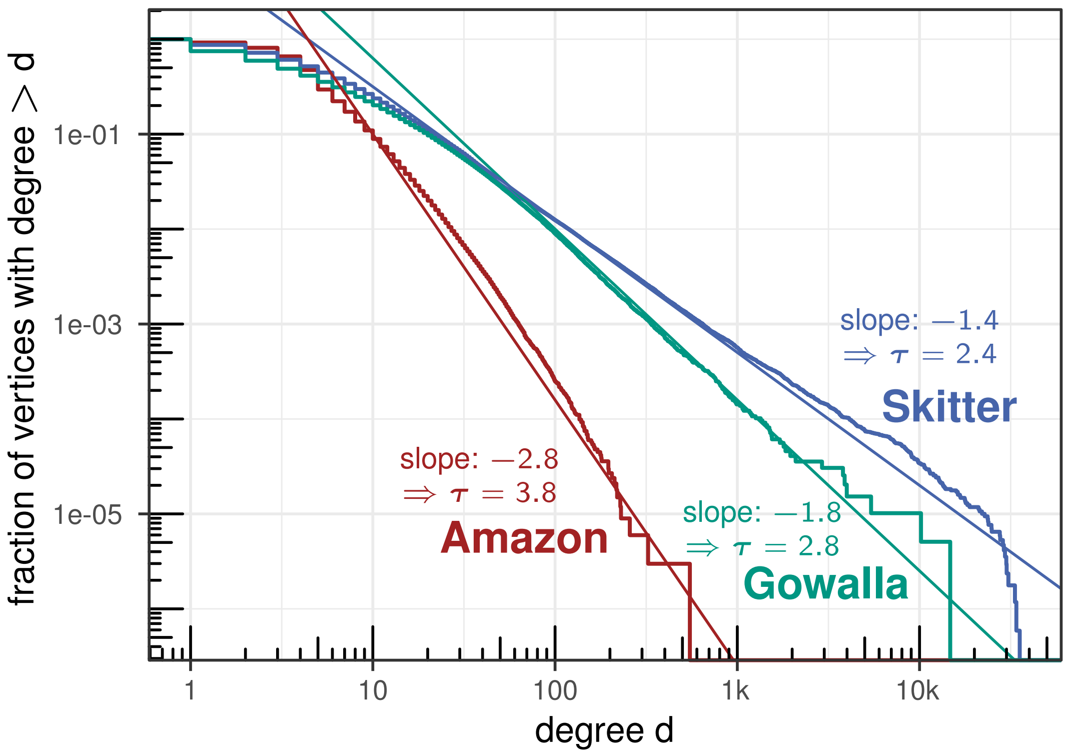


Noch weiterführender:
Scale-free networks well done
(Phys. Rev. Research 1, 033034)

Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$



Definition: power-law distribution

- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent
- üblicherweise: $\tau \in (2, 3)$
  - $\tau > 2 \rightarrow$ constant average degree
  - $\tau \in (2, 3] \rightarrow$ variance increases with n
  - $\tau > 3 \rightarrow$ constant variance

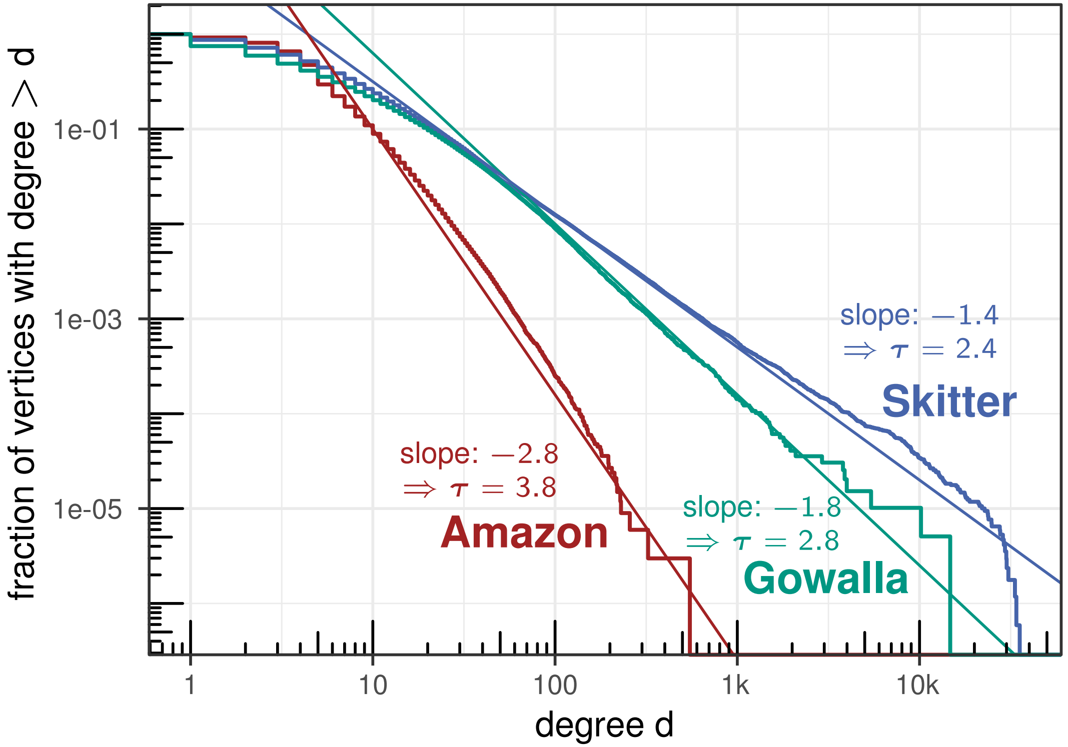
Heterogene Gradverteilung – Details

Gerade Linie im log-log plot

- gerade Linie: $y' = b - ax'$
- log-log plot: $\log y = b - a \log x$
- CCDF: $y = \bar{F}_D(x) = \Pr(D > x) = e^b x^{-a}$
- PDF: $f_D(x) = cx^{-\tau}$ mit $\tau = a + 1$

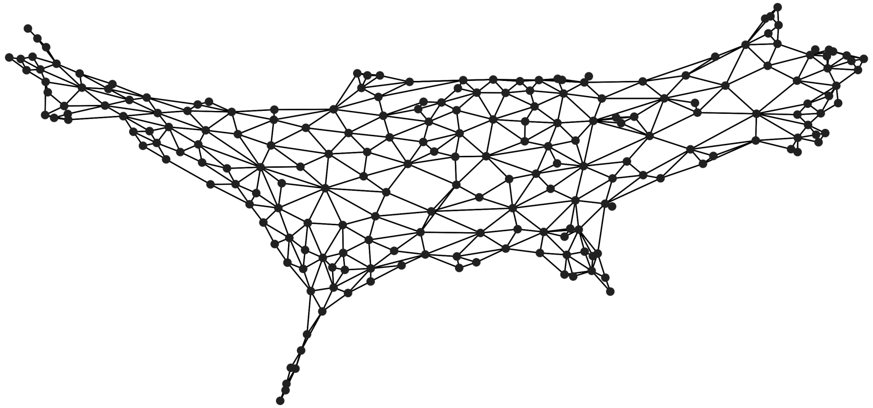
Definition: power-law distribution

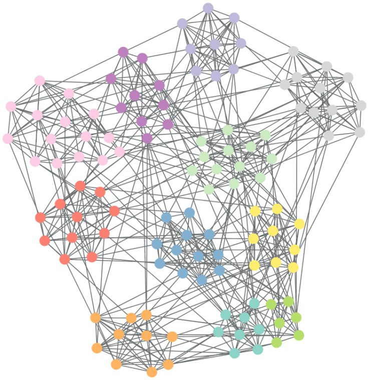
- Zufallsvariable D mit Dichte $f_D(x) = cx^{-\tau}$
- τ heißt auch power-law exponent
- üblicherweise: $\tau \in (2, 3)$
  - $\tau > 2 \rightarrow$ constant average degree
  - $\tau \in (2, 3] \rightarrow$ variance increases with n
  - $\tau > 3 \rightarrow$ constant variance

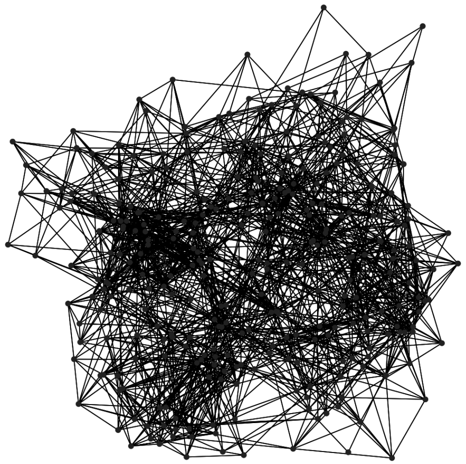


einfacheres Maß für Heterogenität: *coefficient of variaton:* $\frac{\text{std. dev.}}{\text{mean}}$

Lokalität Wiederholung





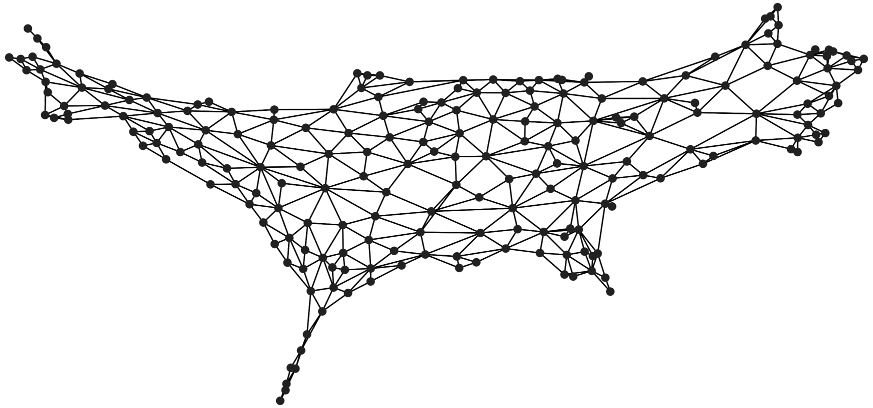


- Teil des US-Straßennetzes
- Knoten mit gemeinsamen Nachbarn sind oft selbst benachbart

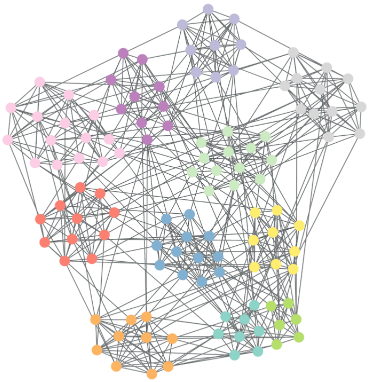
- Spiele von College football teams
- Kanten entsprechen community structure

- Erdős–Rényi Graph
- Jede mögliche Kante existiert mit Wahrscheinlichkeit p (unabhängig)

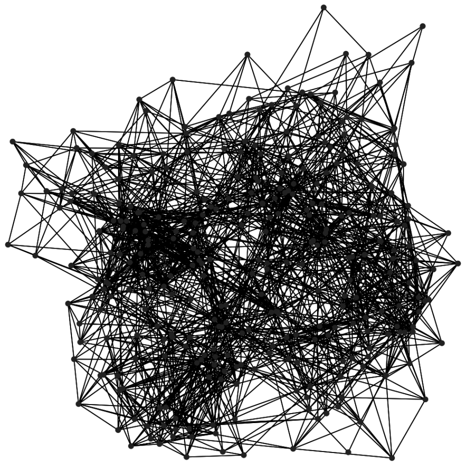
Lokalität Wiederholung



- Teil des US-Straßennetzes
- Knoten mit gemeinsamen Nachbarn sind oft selbst benachbart

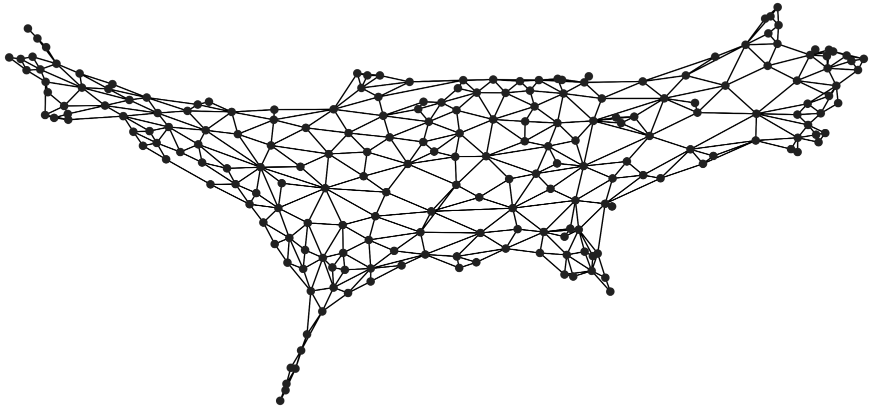


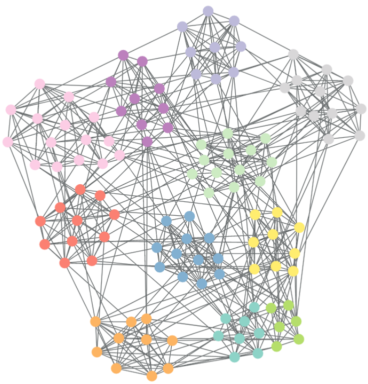
- Spiele von College football teams
- Kanten entsprechen community structure

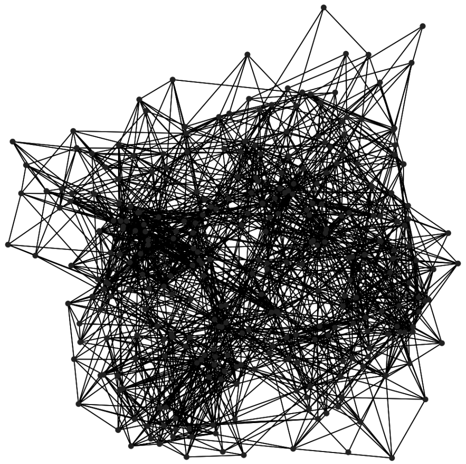


- Erdős–Rényi Graph
- Jede mögliche Kante existiert mit Wahrscheinlichkeit p (unabhängig)

Lokalität Wiederholung







- Teil des US-Straßennetzes
- Knoten mit gemeinsamen Nachbarn sind oft selbst benachbart

- Spiele von College football teams
- Kanten entsprechen community structure

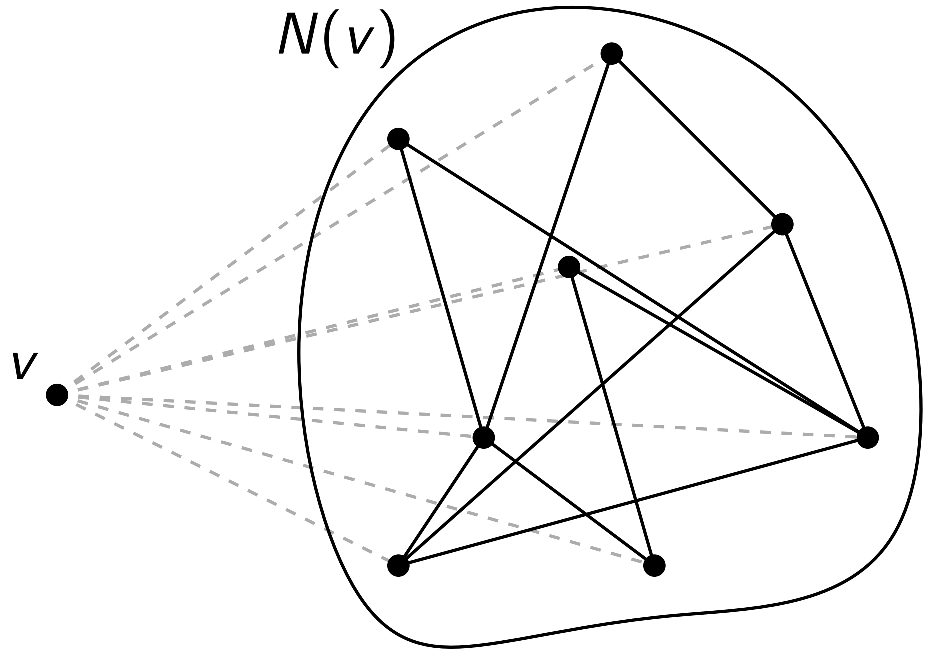
- Erdős–Rényi Graph
- Jede mögliche Kante existiert mit Wahrscheinlichkeit p (unabhängig)

- *„Lokalität“*, *„Clustering“*, *„Geometrie“*
- In vielen Echtwelt Netzwerken beobachtet (z.B. Straßen-, soziale Netzwerke)

- keine „Lokalität“
- high *“temperature”*

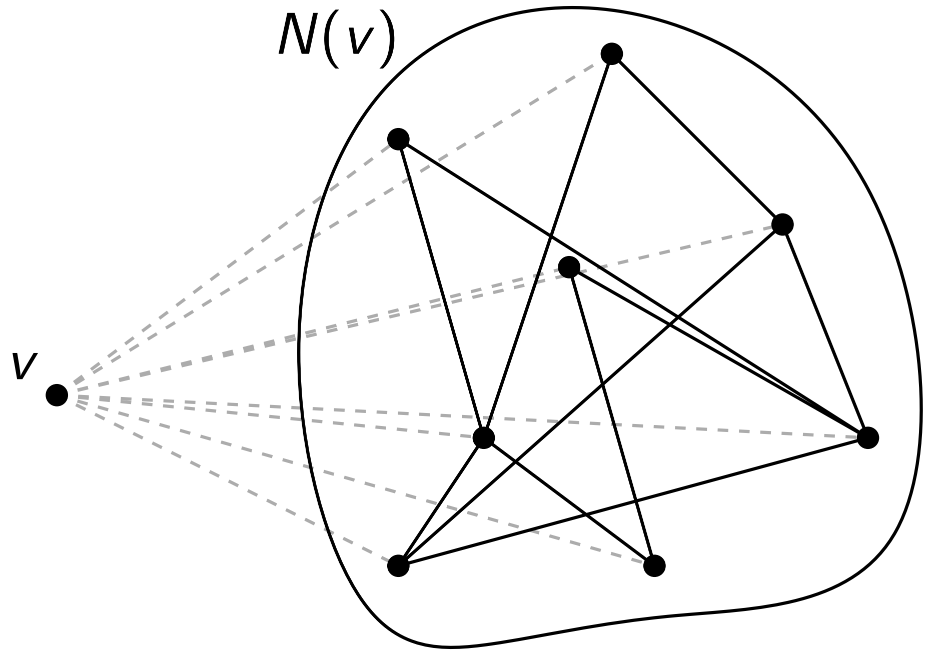
Local Clustering Coefficient

- Wie nah ist meiner Nachbarschaft an einer Clique dran?



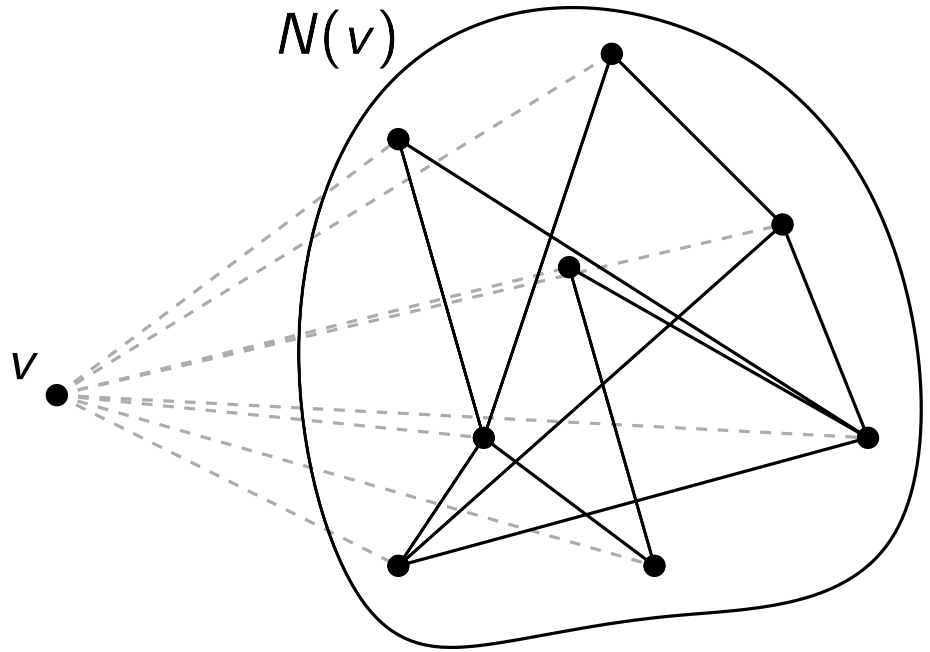
Local Clustering Coefficient

- Wie nah ist meiner Nachbarschaft an einer Clique dran?
- Wie viele Kanten kann es in $N(v)$ geben?



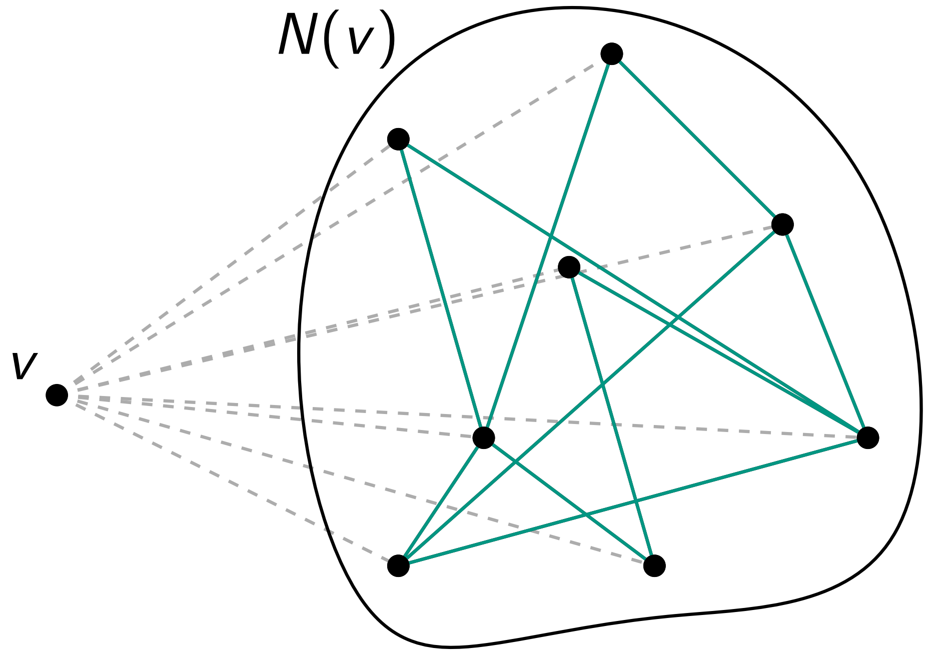
Local Clustering Coefficient

- Wie nah ist meiner Nachbarschaft an einer Clique dran?
- Wie viele Kanten kann es in $N(v)$ geben?
  - $\binom{\deg(v)}{2}$



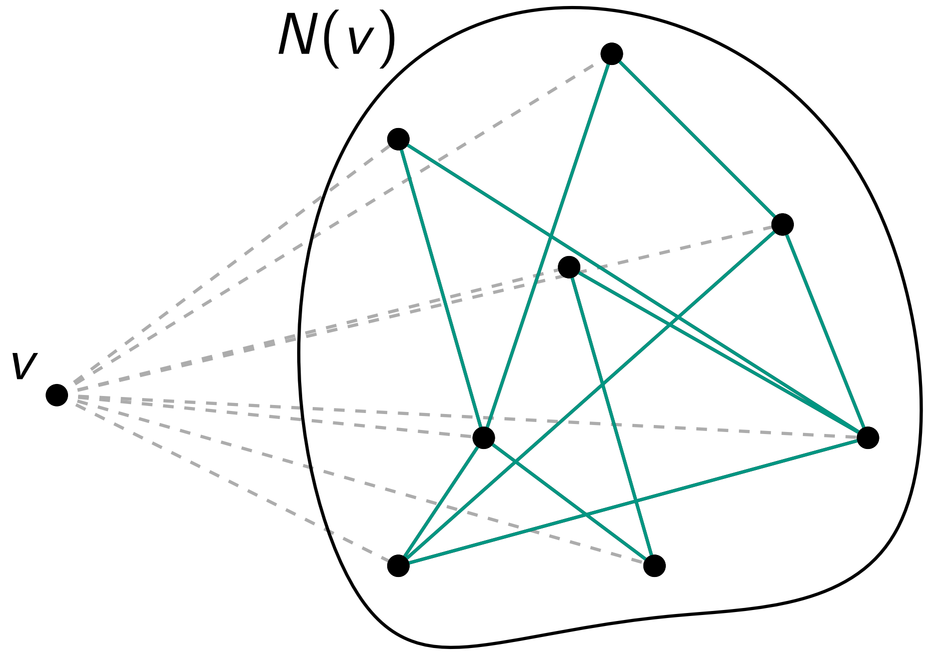
Local Clustering Coefficient

- Wie nah ist meiner Nachbarschaft an einer Clique dran?
- Wie viele Kanten kann es in $N(v)$ geben?
  - $\binom{\deg(v)}{2}$
- Wie viele Kanten sind tatsächlich in $N(v)$?



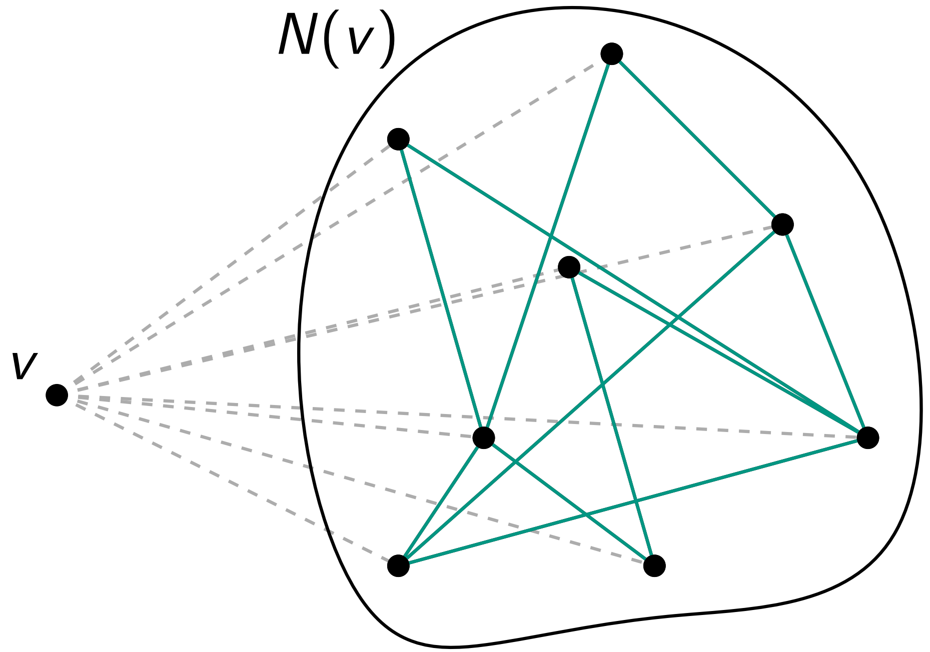
Local Clustering Coefficient

- Wie nah ist meiner Nachbarschaft an einer Clique dran?
- Wie viele Kanten kann es in $N(v)$ geben?
  - $\binom{\deg(v)}{2}$
- Wie viele Kanten sind tatsächlich in $N(v)$?
- Local Clustering von v: $\frac{\#\text{edges in } N(v)}{\binom{\deg(v)}{2}}$
- $\Rightarrow$ Durchschnitt über alle Knoten bilden



Local Clustering Coefficient

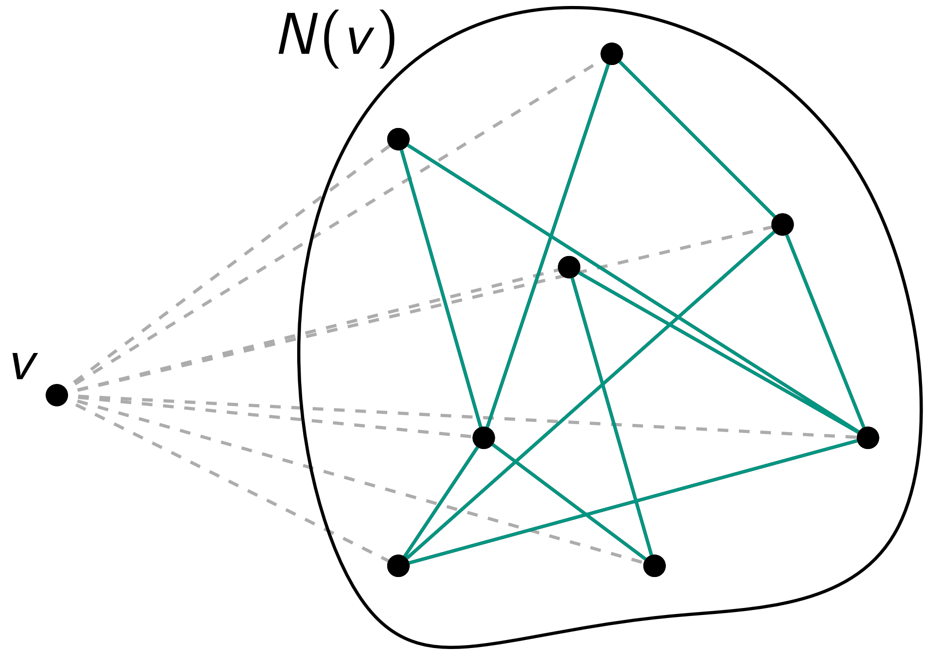
- Wie nah ist meiner Nachbarschaft an einer Clique dran?
- Wie viele Kanten kann es in $N(v)$ geben?
  - $\binom{\deg(v)}{2}$
- Wie viele Kanten sind tatsächlich in $N(v)$?
- Local Clustering von v: $\frac{\#\text{edges in } N(v)}{\binom{\deg(v)}{2}}$
- $\Rightarrow$ Durchschnitt über alle Knoten bilden



- Was ist der erwartete Local Clustering Coefficient von Erdős–Rényi Graphen?

Local Clustering Coefficient

- Wie nah ist meiner Nachbarschaft an einer Clique dran?
- Wie viele Kanten kann es in $N(v)$ geben?
  - $\binom{\deg(v)}{2}$
- Wie viele Kanten sind tatsächlich in $N(v)$?
- Local Clustering von v: $\frac{\#\text{edges in } N(v)}{\binom{\deg(v)}{2}}$
- $\Rightarrow$ Durchschnitt über alle Knoten bilden

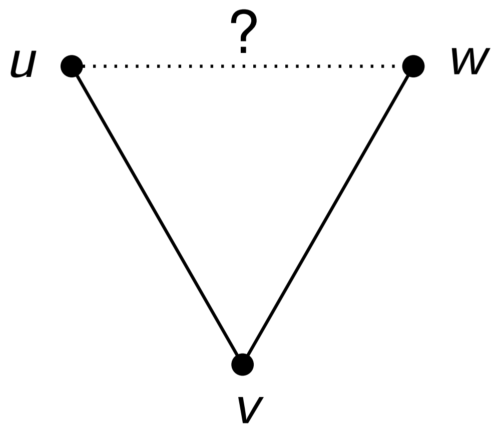


- Was ist der erwartete Local Clustering Coefficient von Erdős–Rényi Graphen?

- Jede Kante kommt with Wahrscheinlichkeit p vor $\Rightarrow \frac{p\binom{\deg(v)}{2}}{\binom{\deg(v)}{2}} = p$

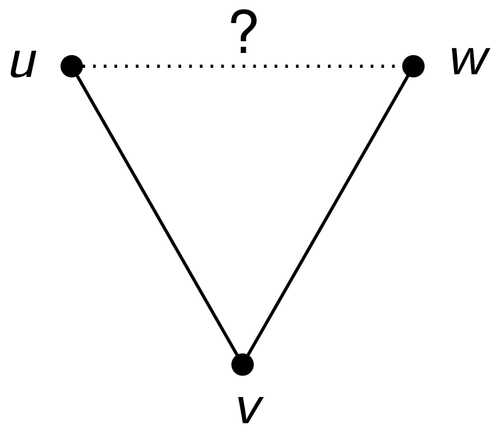
Global Clustering Coefficient

- Betrachte $\{u, v\}, \{v, w\} \in E$
- Ist dann auch $\{u, w\} \in E$?



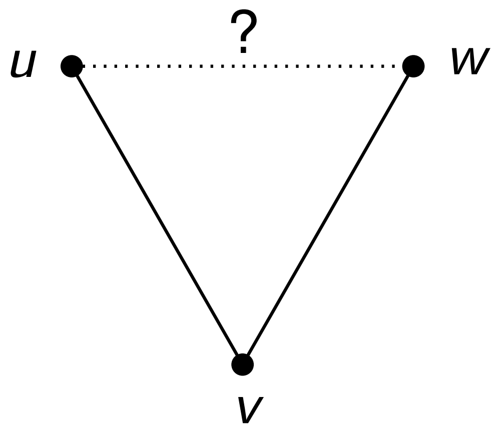
Global Clustering Coefficient

- Betrachte $\{u, v\}, \{v, w\} \in E$
- Ist dann auch $\{u, w\} \in E$?
- Tripel: drei Knoten die mit mindestens 2 Kanten verbunden sind
  - offen: $\{u, w\} \notin E$
  - geschlossen: $\{u, w\} \in E$



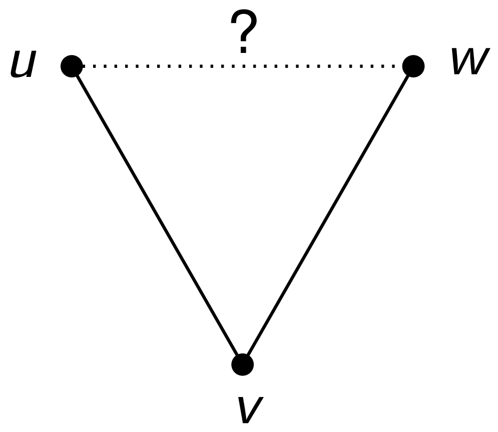
Global Clustering Coefficient

- Betrachte $\{u, v\}, \{v, w\} \in E$
- Ist dann auch $\{u, w\} \in E$?
- Tripel: drei Knoten die mit mindestens 2 Kanten verbunden sind
  - offen: $\{u, w\} \notin E$
  - geschlossen: $\{u, w\} \in E$
- Global Clustering Coefficient: $\frac{\text{\# geschlossene Tripel}}{\text{\#Tripel gesamt}}$



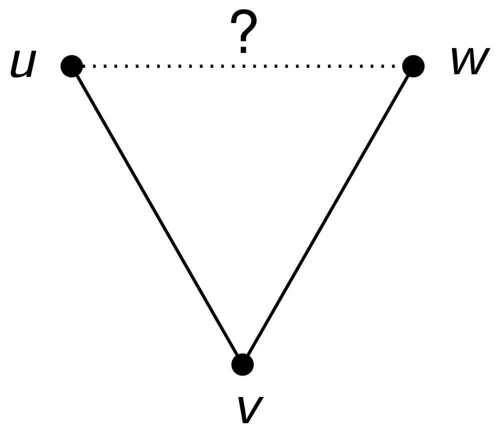
Global Clustering Coefficient

- Betrachte $\{u, v\}, \{v, w\} \in E$
- Ist dann auch $\{u, w\} \in E$?
- Tripel: drei Knoten die mit mindestens 2 Kanten verbunden sind
  - offen: $\{u, w\} \notin E$
  - geschlossen: $\{u, w\} \in E$
- Global Clustering Coefficient: $\frac{\text{\# geschlossene Tripel}}{\text{\#Tripel gesamt}}$
- Nicht über alle Tripel aus Knoten iterieren
- Lieber über Kanten und Nachbarschaften



Global Clustering Coefficient

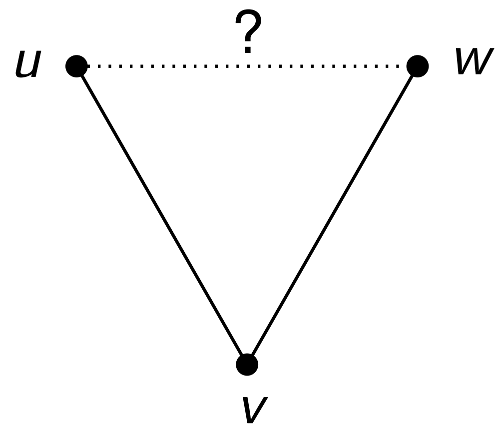
- Betrachte $\{u, v\}, \{v, w\} \in E$
- Ist dann auch $\{u, w\} \in E$?
- Tripel: drei Knoten die mit mindestens 2 Kanten verbunden sind
  - offen: $\{u, w\} \notin E$
  - geschlossen: $\{u, w\} \in E$
- Global Clustering Coefficient: $\frac{\text{\# geschlossene Tripel}}{\text{\#Tripel gesamt}}$
- Nicht über alle Tripel aus Knoten iterieren
- Lieber über Kanten und Nachbarschaften



- Was ist der erwartete Global Clustering Coefficient von Erdős–Rényi Graphen?

Global Clustering Coefficient

- Betrachte $\{u, v\}, \{v, w\} \in E$
- Ist dann auch $\{u, w\} \in E$?
- Tripel: drei Knoten die mit mindestens 2 Kanten verbunden sind
  - offen: $\{u, w\} \notin E$
  - geschlossen: $\{u, w\} \in E$
- Global Clustering Coefficient: $\frac{\text{\# geschlossene Tripel}}{\text{\#Tripel gesamt}}$
- Nicht über alle Tripel aus Knoten iterieren
- Lieber über Kanten und Nachbarschaften

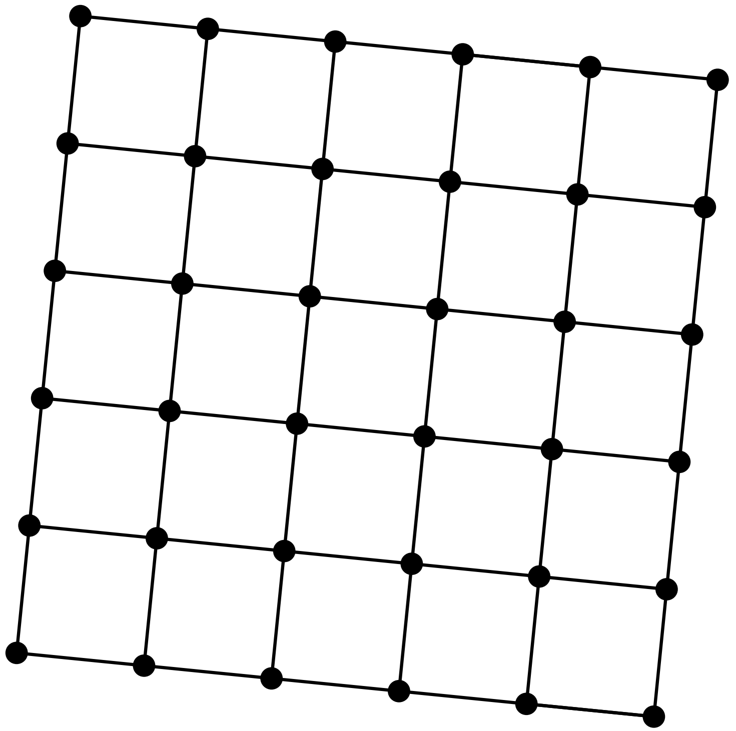


- Was ist der erwartete Global Clustering Coefficient von Erdős–Rényi Graphen?

- Auch p aber nicht so leicht zu zeigen...

Probleme mit den Clustering Coefficients

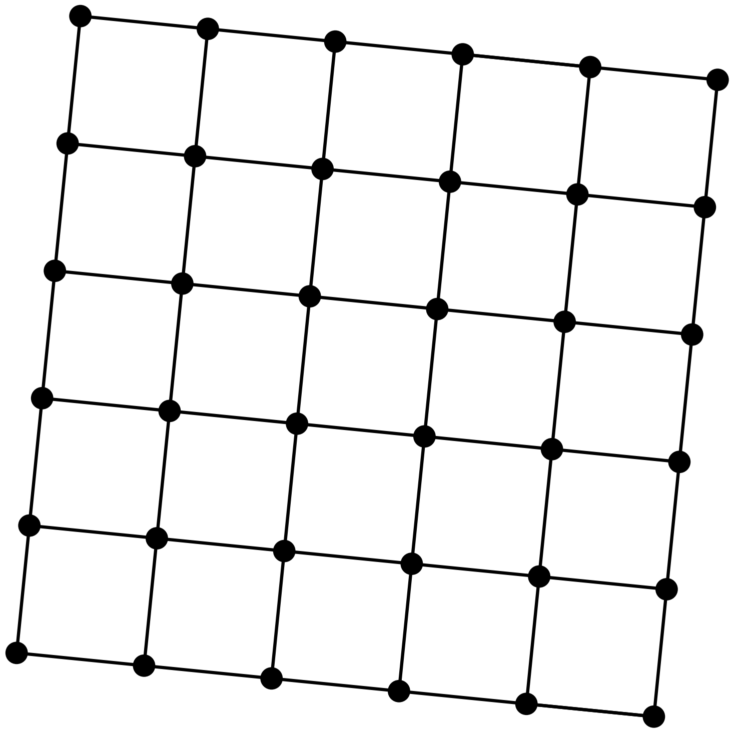
- Was ist der Local / Global Clustering Coefficient von Grid Graphen?



Probleme mit den Clustering Coefficients

- Was ist der Local / Global Clustering Coefficient von Grid Graphen?

- Beide sind 0, weil der Graph bipartit ist
- Eigentlich hat der Graph viel *Lokalität*

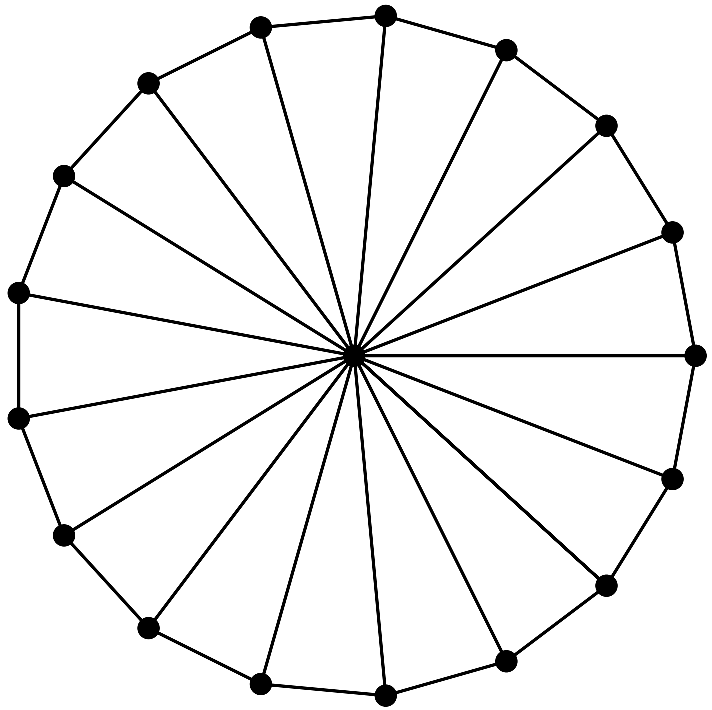


Probleme mit den Clustering Coefficients

- Was ist der Local / Global Clustering Coefficient von Grid Graphen?

- Beide sind 0, weil der Graph bipartit ist
- Eigentlich hat der Graph viel *Lokalität*

- Was ist der Local / Global Clustering Coefficient von Wheel Graphen?



Probleme mit den Clustering Coefficients

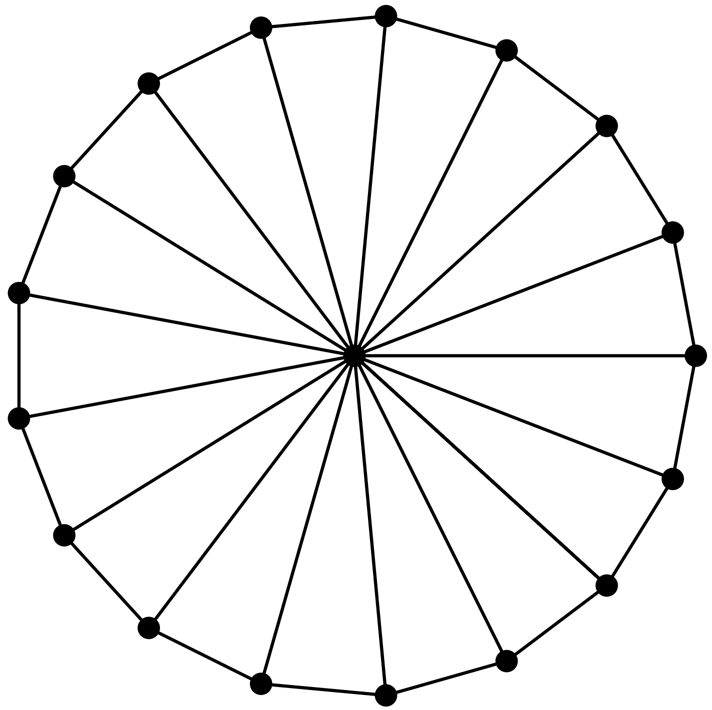
Karlsruher Institut für Technologie

- Was ist der Local / Global Clustering Coefficient von Grid Graphen?

- Beide sind 0, weil der Graph bipartit ist
- Eigentlich hat der Graph viel *Lokalität*

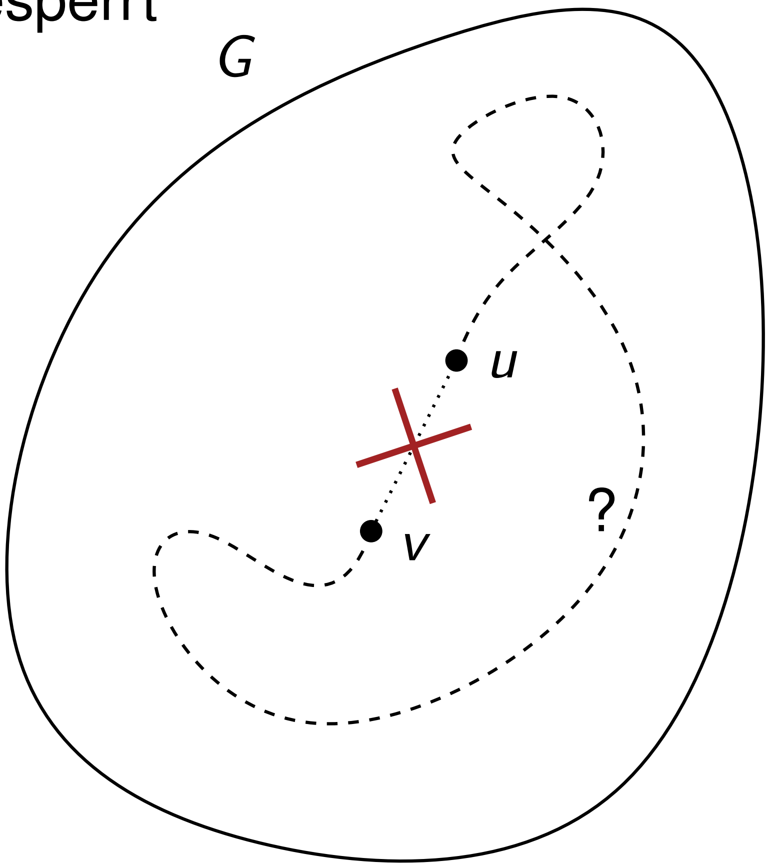
- Was ist der Local / Global Clustering Coefficient von Wheel Graphen?

- Local: geht gegen $2/3$ für $n \to \infty$
- Global: geht gegen 0 für $n \to \infty$



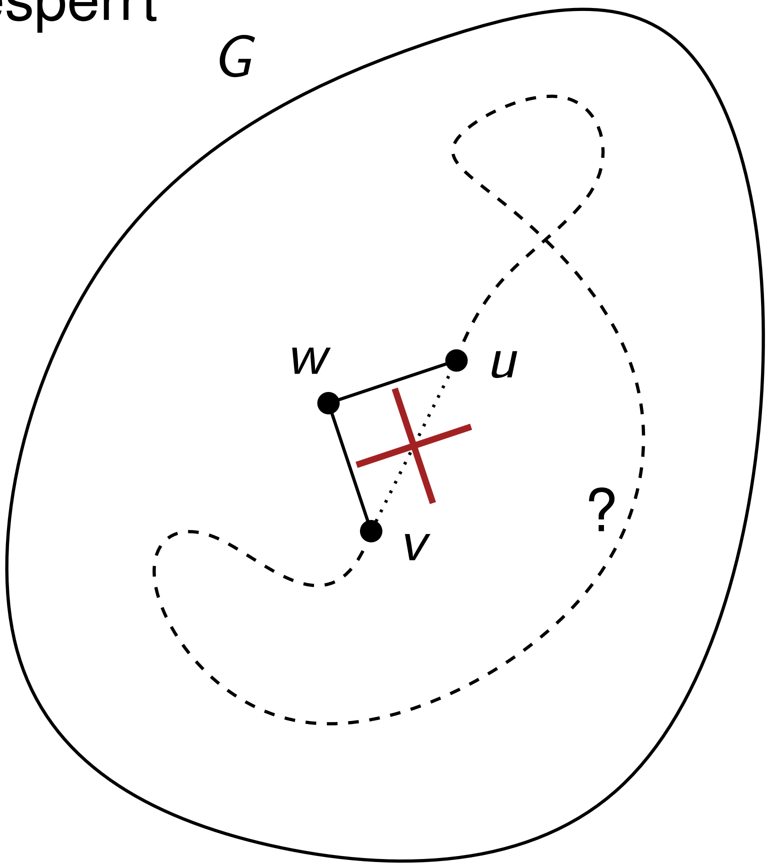
Detour Distance

- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?



Detour Distance

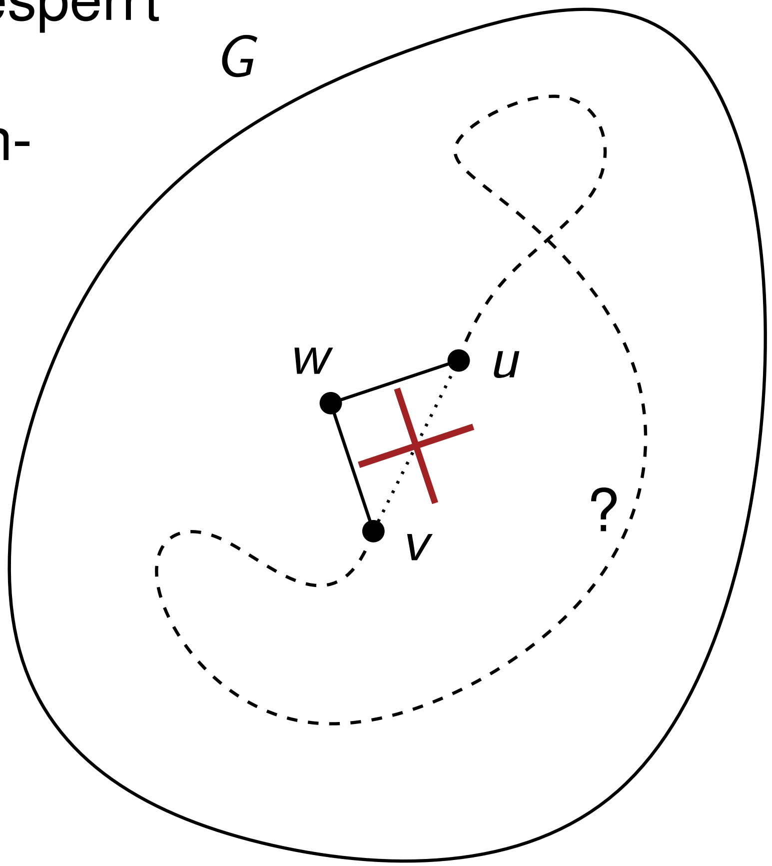
- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?



Detour Distance

- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?
- Betrachte Verhältnis zwischen Umweg und durchschnittlicher Distanz zwischen nicht Nachbarn
- Für Kante $e = \{u, v\} \in E$:

$$C(e) = 1 - \frac{d_{G \setminus e}(u, v) - 2}{\text{avg. Distance of non neighbors} - 2}$$

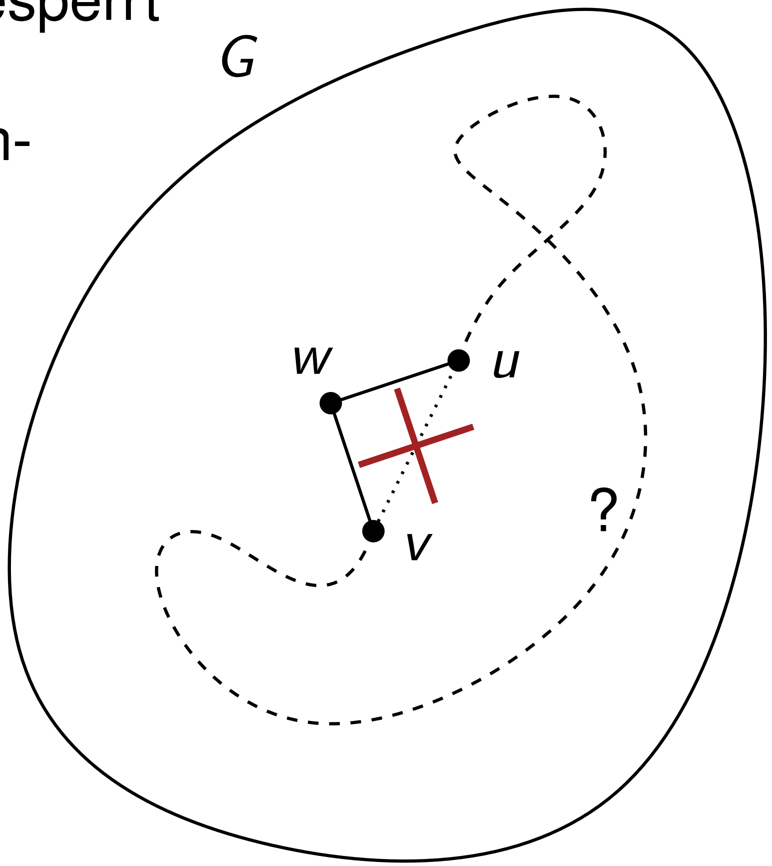


Detour Distance

- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?
- Betrachte Verhältnis zwischen Umweg und durchschnittlicher Distanz zwischen nicht Nachbarn
- Für Kante $e = \{u, v\} \in E$:

$$C(e) = 1 - \frac{d_{G \setminus e}(u, v) - 2}{\text{avg. Distance of non neighbors} - 2}$$

- $C(e) = 1 \Leftrightarrow u, v$ haben gemeinsamen Nachbarn
- Bilde Durschschnitt über alle Kanten

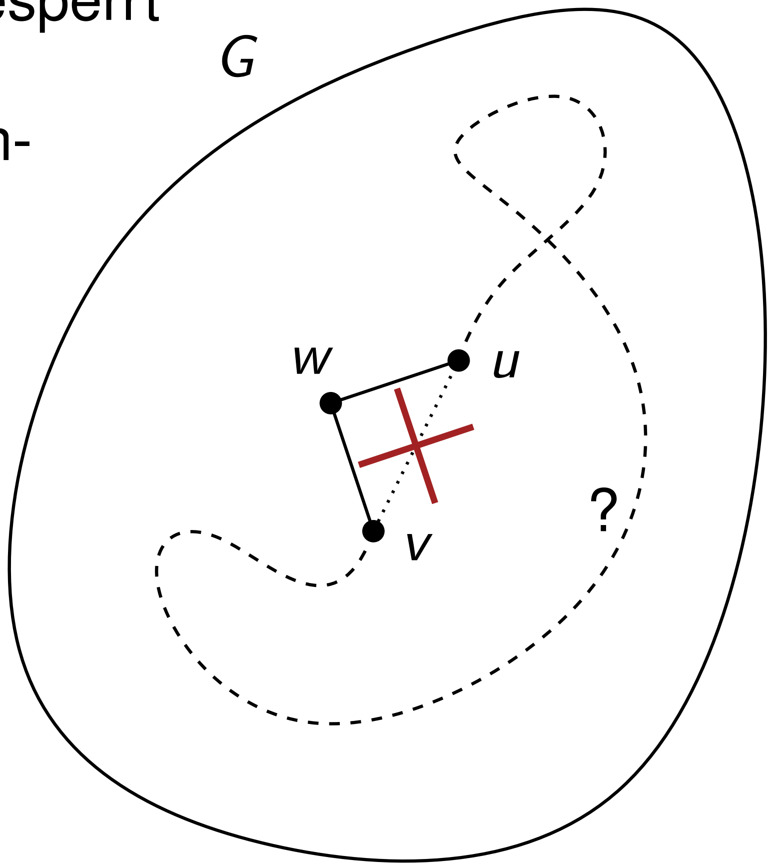


Detour Distance

- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?
- Betrachte Verhältnis zwischen Umweg und durchschnittlicher Distanz zwischen nicht Nachbarn
- Für Kante $e = \{u, v\} \in E$:

$$C(e) = 1 - \frac{d_{G \setminus e}(u, v) - 2}{\text{avg. Distance of non neighbors} - 2}$$

- $C(e) = 1 \Leftrightarrow u, v$ haben gemeinsamen Nachbarn
- Bilde Durschschnitt über alle Kanten
- Aufpassen, dass man nicht durch 0 teilt

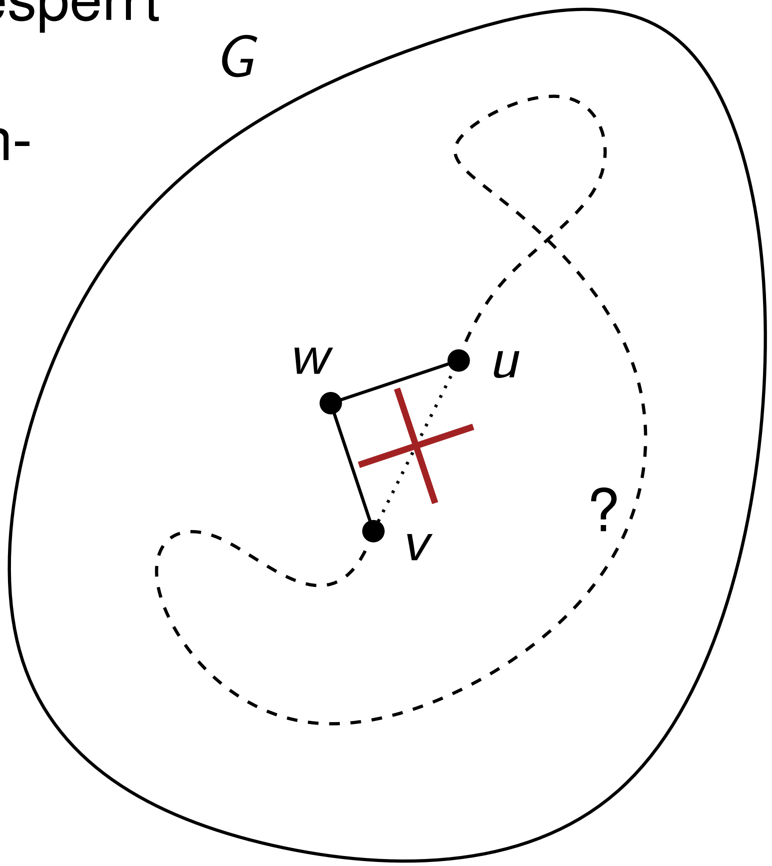


Detour Distance

- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?
- Betrachte Verhältnis zwischen Umweg und durchschnittlicher Distanz zwischen nicht Nachbarn
- Für Kante $e = \{u, v\} \in E$:

$$C(e) = 1 - \frac{d_{G \setminus e}(u, v) - 2}{\text{avg. Distance of non neighbors} - 2}$$

- $C(e) = 1 \Leftrightarrow u, v$ haben gemeinsamen Nachbarn
- Bilde Durschschnitt über alle Kanten
- Aufpassen, dass man nicht durch 0 teilt
- Aufwändig zu berechnen $\Rightarrow$ sampeln



Detour Distance

- Wie groß ist der Umweg, wenn die Straße vor mir gesperrt wird?
- Betrachte Verhältnis zwischen Umweg und durchschnittlicher Distanz zwischen nicht Nachbarn
- Für Kante $e = \{u, v\} \in E$:

$$C(e) = 1 - \frac{d_{G \setminus e}(u, v) - 2}{\text{avg. Distance of non neighbors} - 2}$$

- $C(e) = 1 \Leftrightarrow u, v$ haben gemeinsamen Nachbarn
- Bilde Durschschnitt über alle Kanten
- Aufpassen, dass man nicht durch 0 teilt
- Aufwändig zu berechnen $\Rightarrow$ sampeln

  - Was ist die Detour Distance Grid / Wheel Graphen?