Lecture 10

Internet Protokolle II

Peer-to-Peer Systems
& Random Graphs

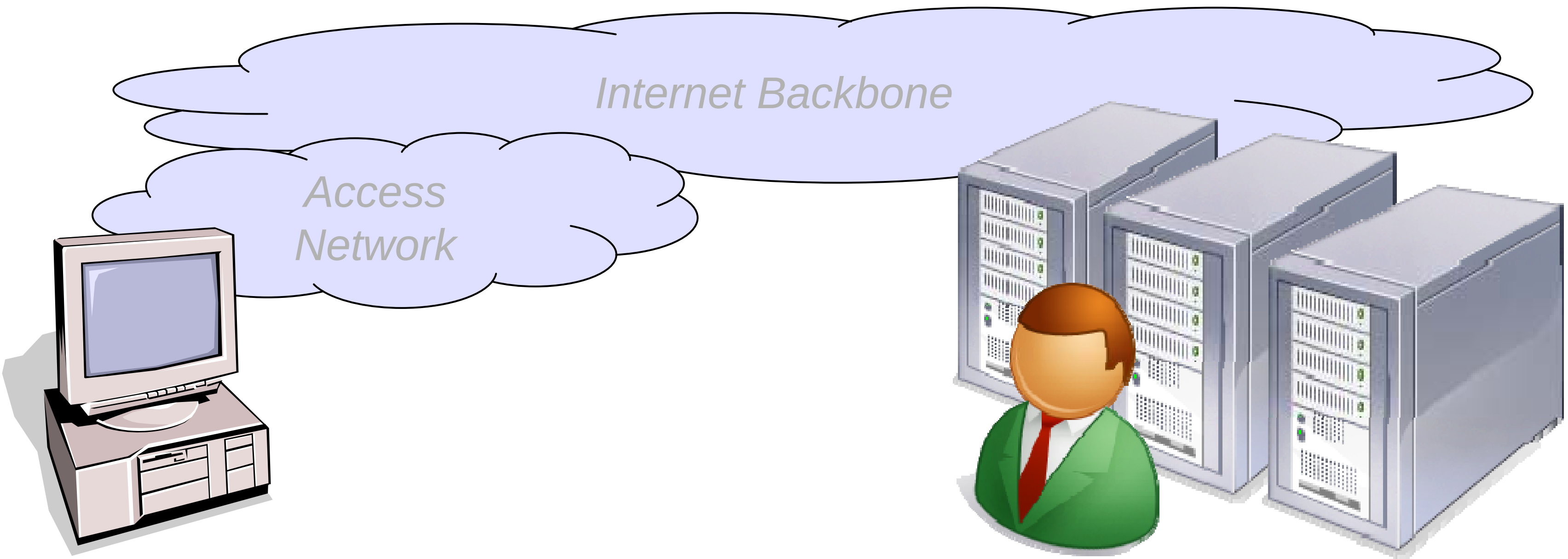
Thomas Fuhrmann

Network Architectures
Computer Science Department
Technical University Munich

Motivation (1)

De facto ist das Internet heute zentralistisch:

- *Hierarchische Netze mit administrativer Kontrolle*
- *Dienstnehmer und -geber klar getrennt*

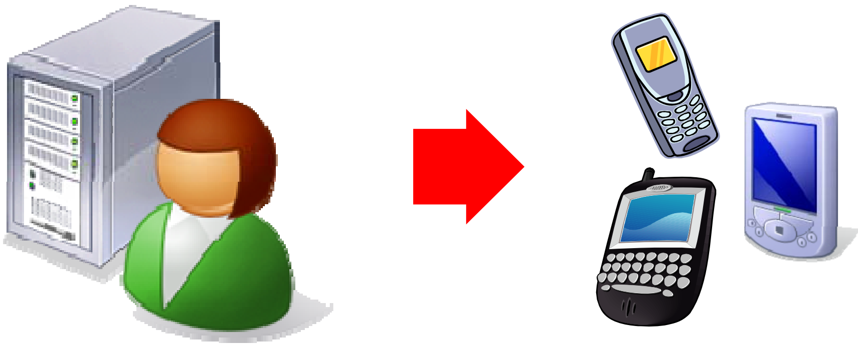


Motivation (2)

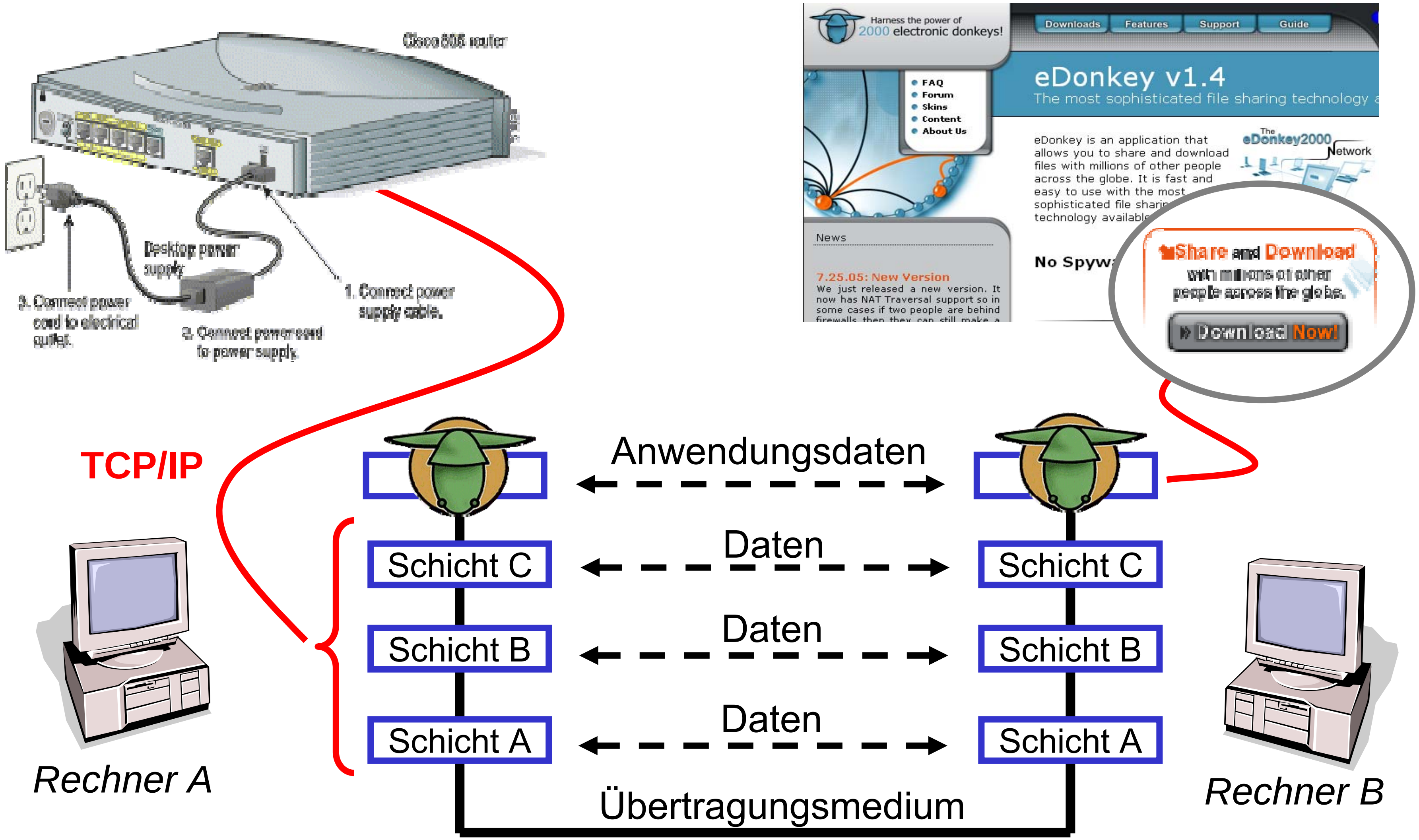
New requirements:

- *Mesh Networks, Multihoming*
- *Mobility*
- *Multimedia*
- *New Services (Disappearing computer)*

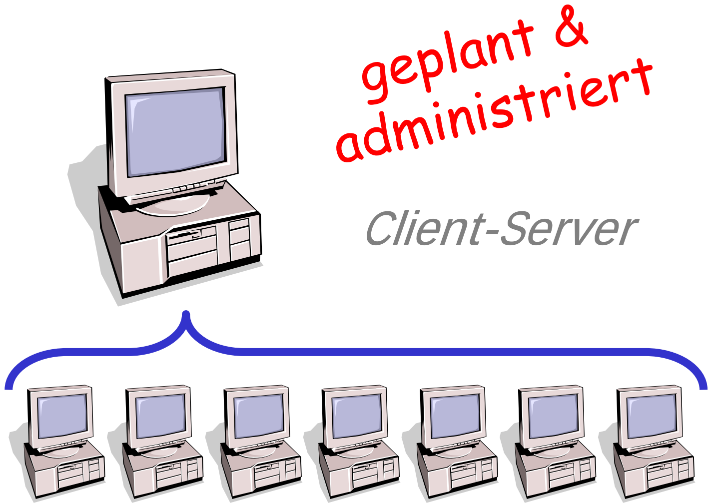
> *The Internet has become ossified when it became commercially important in the early 1990s.*

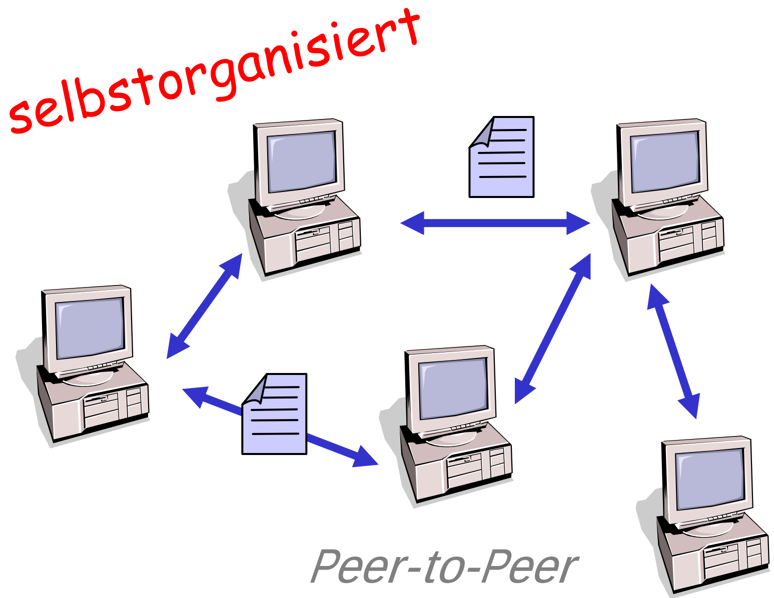


Was ist Peer-to-Peer? – Sichtweise A



Was ist Peer-to-Peer? – Sichtweise B





Peer-to-Peer …

Organisches Wachstum

- Jeder kann jede Funktion übernehmen.
- Neue Geräte bringen Ihre benötigten Ressourcen quasi selbst mit

Robust und fehlertolerant

- Alle Einheiten sind im Prinzip austauschbar

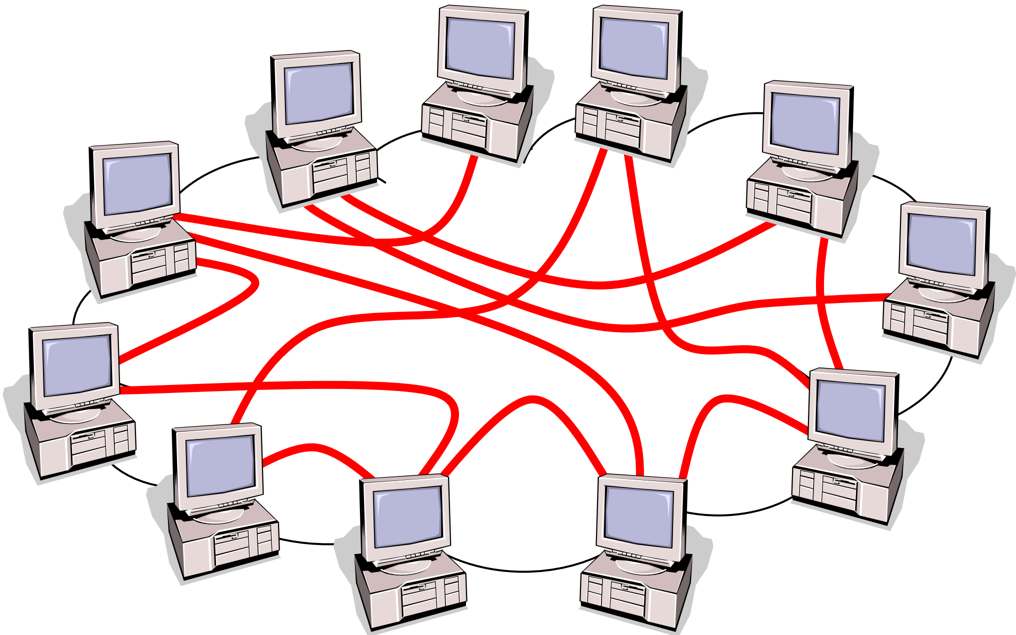
Forschungsfrage:

- Welche Protokollregeln führen zu welchem Verhalten?

Peer-to-Peer-Netze (1)

Frühe Definition von Peer-to-Peer (nach Clay Shirky):

Eine Klasse von Anwendungen, die Ressourcen am Rande des Internets nutzen und zum Grossteil unabhängig von zentralen Servern ist.



Beispiel: Rechner nutzen die Konnektivität des Internets, um gemeinsam einen Zweck zu erfüllen

Peer-to-Peer-Netze (2)

- Ressourcen können sein
  - Speicherplatz, z.B. Napster, Gnutella
  - Rechenzeit, z.B. Seti@Home
  - Bandbreite, z.B. MANETs
  - Human Resources, z.B. USENET
- Meist: P2P-System „oberhalb“ des DNS
  - Dial-up Rechner sind nicht durch IP identifizierbar
- P2P ist nicht neu
  - Internet ist als P2P-System konzipiert
  - Beispiel: Remote Login auf andere Workstations, FTP, e-Mail, … - Die Anwendungen des frühen Internets sind häufig P2P-Anwendungen
  - Beispiel: „Archie“ sucht auf benach-barten FTP-Servern nach Dateien
  - Erst die Kommerzialisierung durch die Verbreitung des WWW schafft den Eindruck einer Client-Server-Struktur
- Der Begriff „Peer-to-Peer“ ist problematisch
  - Häufig enger Fokus auf einen Anwendungstyp, nämlich File-Sharing-P2P-Netze
  - Aber P2P-Systeme nutzen manchmal auch zentrale Server
  - Umgekehrt nutzen Client-Server-Systeme immer auch Ressourcen am Netzwerk-Rand, z.B. Web-Browser, Web-Surfer, Buchhalter am SAP-R3-System, …

Peer-to-Peer-Netze (3)

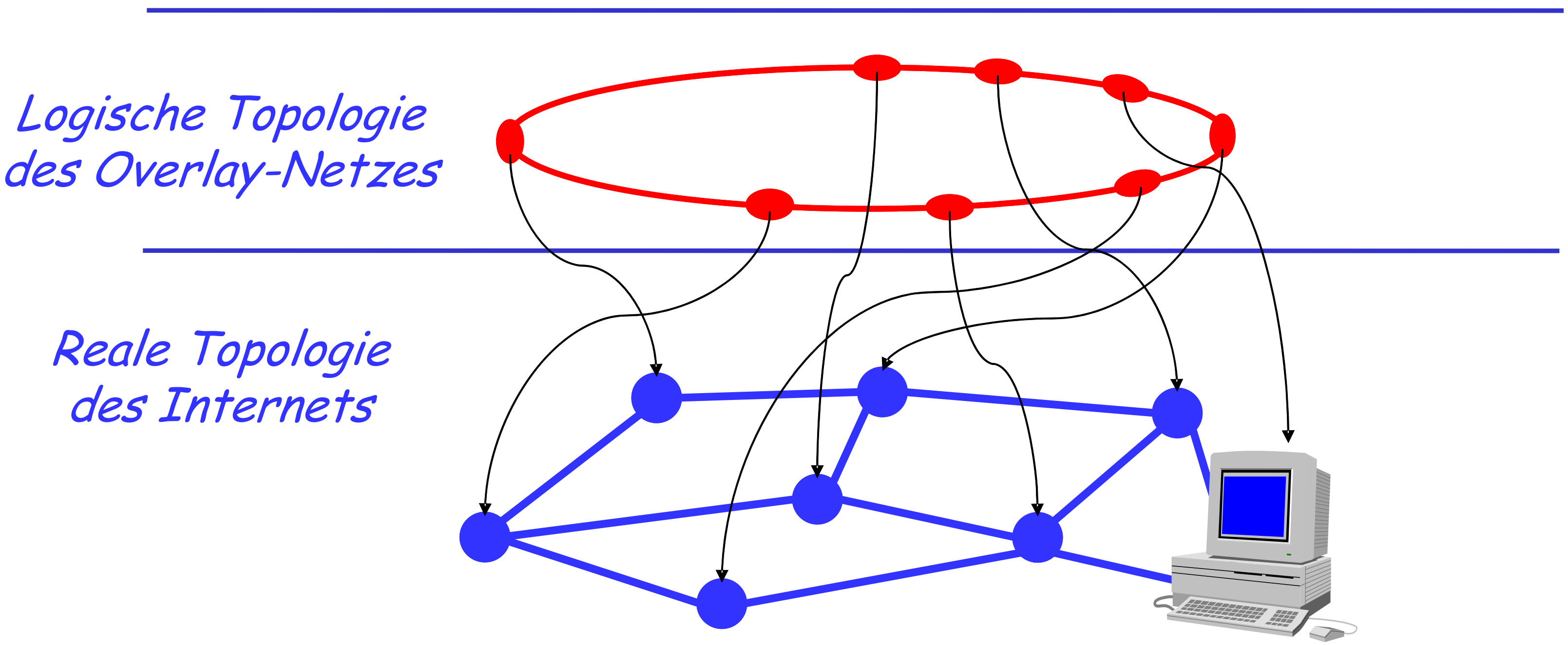
- Bessere Definition von Peer-to-Peer-Anwendungen:

Eine Klasse von Anwendungen, die bezüglich ihrer Netzwerkaktivitäten typischerweise aus funktional gleichartigen Implementierungen bestehen.

- Damit:
  – Napster, Gnutella, BitTorrent, … sind P2P-Anwendungen, wobei Index (Napster) und Tracker (BitTorrent) eigentlich als Server gesehen werden müssten.
  – HTTP, FTP, e-Mail, … sind keine P2P-Anwendungen, weil man typischerweise Client und Server unterscheiden kann.

Overlay-Netze (1)

Overlay-Netze sind „logische“ bzw. „virtuelle“ Netze, die über eine vorhandene Netzwerkinfrastruktur gelegt werden:



Overlay-Netze (2)

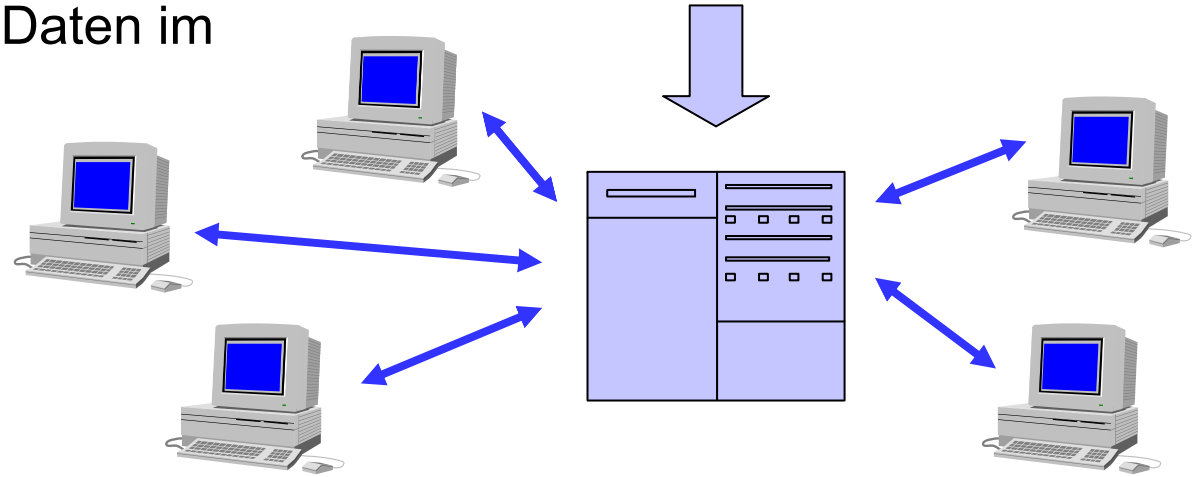
- Das Internet selbst begann als Overlay über dem Telefonnetz
  - Dort standen bereits Adressierung, Wegewahl, etc. zur Verfügung
  - Es wurde aber eine eigene Adressierung (IP-Adressen statt Telefonnummern), eigene Vermittlung (Paketvermittlung), eigene Wegewahl (RIP), etc. implementiert
- Overlay-Netze sind ein typischer Ansatz, um Funktionalität eines Netzwerks zu erweitern. Beispiele:
  - Experimente mit Multicast im MBone-Overlay
  - Experimente mit IPv6 im 6Bone
  - Sicherheit durch Virtual Private Networks
  - Ausfallsicherheit durch Resilient Overlay Networks
- Wichtig(st)er Unterschied zu physischer Infrastruktur:

Ein Overlay-Netz hat die Freiheit sich selbst eine Topologie zu geben

Die P2P Urahnen (1): SETI@home

Nutzung verteilter Ressourcen, aber nicht P2P: ***SETI@home***

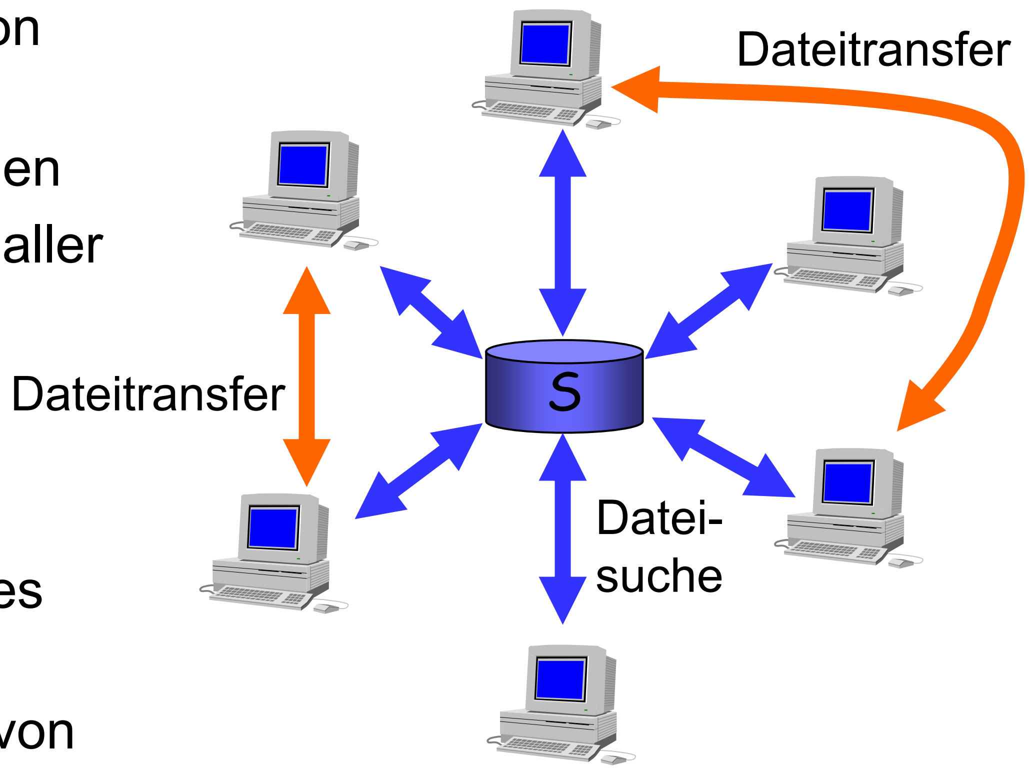
- Zentraler Server verschickt Daten eines großen Radioteleskops an Clients in alle Welt
- Clients haben ungenutzte Rechenzeit, z.B. Screensaver auf Desktop-PCs
- Clients analysieren Daten im Hinblick auf evtl. außerirdische Botschaften
- Suchergebnis wird an Server zurückgemeldet



Die P2P Urahnen (2): Napster

Nutzung verteilter Ressourcen,
aber nur bedingt P2P: ***Napster***

- Erste P2P-Killer-Applikation (1999-2001)
- Austausch von MP3-Dateien
- Zentraler Server hat Liste aller verfügbaren Dateien
- Peers fragen Server, welcher anderer Peer eine gesuchte Datei hat
- Server nennt Adresse eines solchen Peers
- Anfrager lädt Datei direkt von diesem Peer



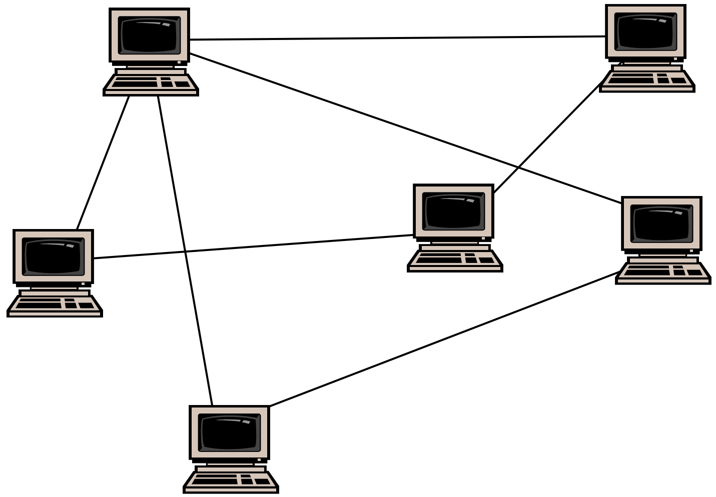
Warum ist das Spannend?

- P2P- und Overlay-Netze liefern einen bedeutenden Beitrag zum Verkehr im Internet
  - Verkehrscharakteristiken untersuchen:
    - Im klassischen Telefonnetz sprechen Menschen zu Menschen
    - WWW: Menschen sprechen mit Maschinen
    - P2P: Maschinen sprechen mit Maschinen
  - Auswirkungen?
- Aufgrund ihrer ***viralen Verbreitung*** wachsen sie extrem schnell
  - Die Nutzer geben die Programme unter einander weiter
  - Es müssen keine zentralen Server nachgerüstet werden oder sonstige physische Komponenten modifiziert werden
- Im Gegensatz zu den klassischen Infrastrukturnetzen ***erzeugen viele Overlay-Protokolle selbst die Topologie*** des Netzes
  - Völlig neuer Untersuchungsgegenstand abseits der klassischen Leistungsanalyse mittels Warteschlangen

Beispiel: Gnutella

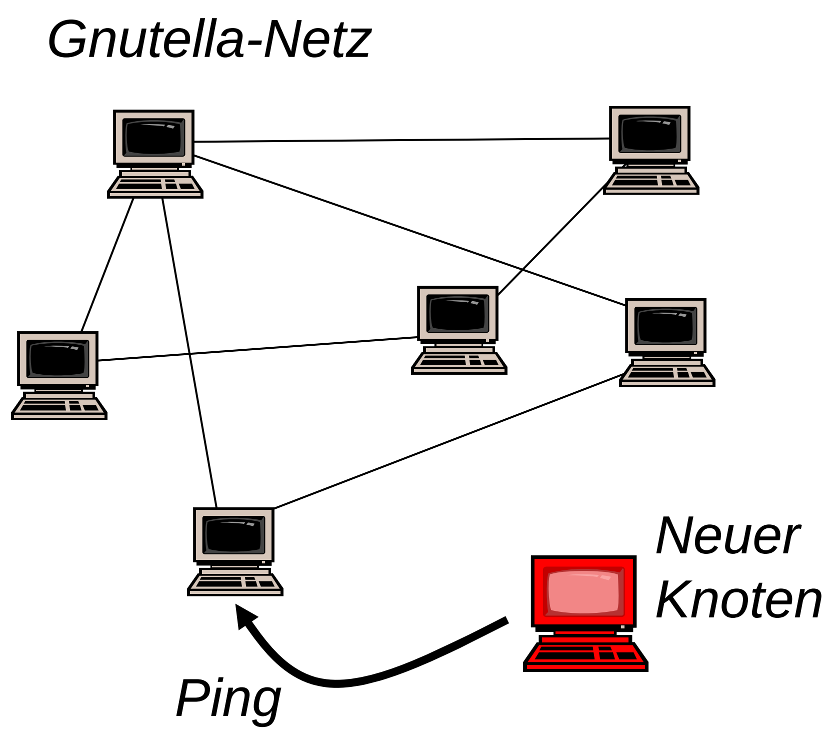
- Selbstorganisierendes Netz
- Dezentraler Dateitausch
- Objektcode im März 2000 kurzzeitig verfügbar
- Danach Vielzahl weiterer „Servants“ mittels Protocol Reverse Engineering entwickelt
- März 2001: 10 000 Knoten
- Mai 2001: 50 000 Knoten

Gnutella-Netz



Beispiel: Gnutella

- Selbstorganisierendes Netz
- Dezentraler Dateitausch
- Objektcode im März 2000 kurzzeitig verfügbar
- Danach Vielzahl weiterer „Servants“ mittels Protocol Reverse Engineering entwickelt
- März 2001: 10 000 Knoten
- Mai 2001: 50 000 Knoten

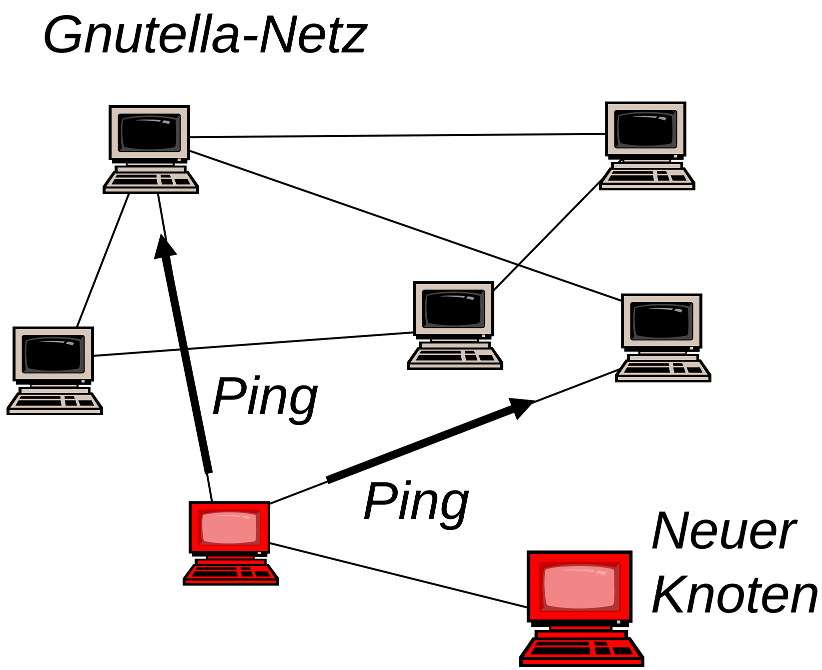


1

Neuer Knoten kennt einen weiteren Knoten und verbindet sich

Beispiel: Gnutella

- Selbstorganisierendes Netz
- Dezentraler Dateitausch
- Objektcode im März 2000 kurzzeitig verfügbar
- Danach Vielzahl weiterer „Servants“ mittels Protocol Reverse Engineering entwickelt
- März 2001: 10 000 Knoten
- Mai 2001: 50 000 Knoten

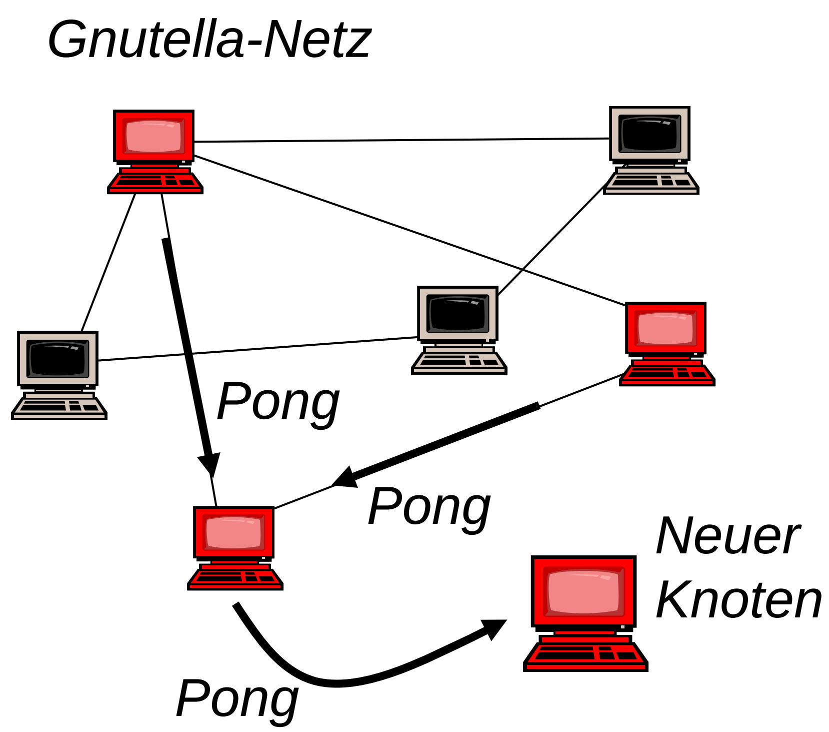


1 Neuer Knoten kennt einen weiteren Knoten und verbindet sich

2 Im Overlay werden Ping-Nachrichten geflutet

Beispiel: Gnutella

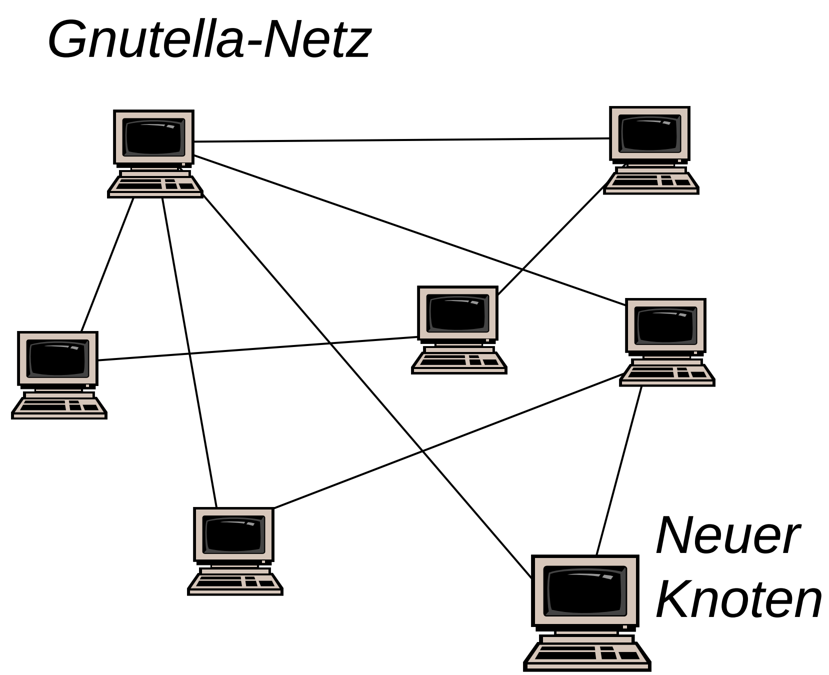
- Selbstorganisierendes Netz
- Dezentraler Dateitausch
- Objektcode im März 2000 kurzzeitig verfügbar
- Danach Vielzahl weiterer „Servants" mittels Protocol Reverse Engineering entwickelt
- März 2001: 10 000 Knoten
- Mai 2001: 50 000 Knoten



1. Neuer Knoten kennt einen weiteren Knoten und verbindet sich
2. Im Overlay werden Ping-Nachrichten geflutet
3. Antworten laufen im Overlay zum neuen Knoten zurück

Beispiel: Gnutella

- Selbstorganisierendes Netz
- Dezentraler Dateitausch
- Objektcode im März 2000 kurzzeitig verfügbar
- Danach Vielzahl weiterer „Servants“ mittels Protocol Reverse Engineering entwickelt
- März 2001: 10 000 Knoten
- Mai 2001: 50 000 Knoten



1. Neuer Knoten kennt einen weiteren Knoten und verbindet sich
2. Im Overlay werden Ping-Nachrichten geflutet
3. Antworten laufen im Overlay zum neuen Knoten zurück
4. Neuer Knoten verbindet sich zu einigen gefund-enen Knoten

Herausforderung für die Forschung

- Forschungsfragen bezüglich der Topologie
  - Wie lässt sich die Topologie von Gnutella und anderen P2P Netzen klassifizieren?
  - Welche Auswirkung hat die Topologie auf die P2P-Anwendung, z.B. Effektivität (=„Tut es das was es soll?“) und Effizienz (= „Wie viele Ressourcen verbraucht es dabei?“)?
  - Welche Auswirkungen hat die Topologie zusammen mit der Anwendung auf die darunter liegende Infrastruktur, d.h. das Internet?
- Beispiel: Suche nach Dateien

Suche in Overlay-Netzen

- „Suche“ in normalen Infrastruktur-Netzen
  - Wo finde ich die Maschine mit dieser IP-Adresse?
  - Lösung: Wegewahl
- Suche in Overlay-Netzen
  - Fluten
  - Random-Walk
  - High-Degree Seeking
  - Distributed Hash Tables
- Bewertungskriterien
  - Verkehrslast, die auf den Overlay-Kanten erzeugt wird
  - Aufwand pro Overlay-Knoten während der Suche
  - Wahrscheinlichkeit mit der ein Inhalt gefunden wird, falls er überhaupt im System vorhanden ist
  - Zeit bis ein Inhalt gefunden wird bzw. die Suche abgebrochen wird
- Weitere Kriterien können z.B. sein, ob unscharfe Suchen sind

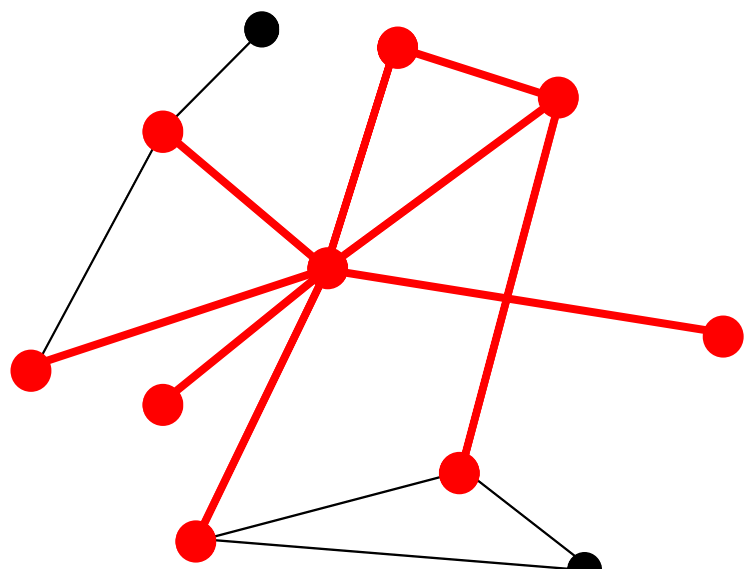
Suchen durch Fluten (1)

Fluten

- Anfragen werden an alle Nachbarn weitergeleitet (außer dem von dem sie kam)
- Duplikate werden sofort verworfen (Speichern und Wiedererkennen von Anfragen erforderlich, z.B. mittels Hash-Werten oder Unique Identifiern)

Bewertung

- Einfachstes Verfahren
- Angewandt z.B. bei Gnutella
- Inhalte werden nur gefunden, wenn sie innerhalb des Suchhorizonts liegen
- Ergebnisse werden schnell gefunden
- Sehr große Belastung des Netzes



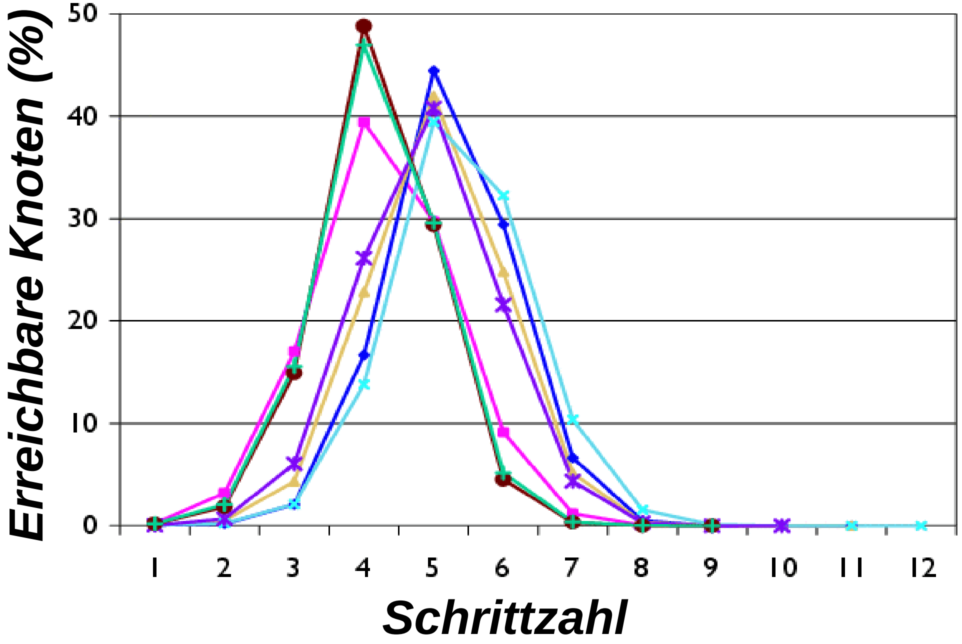
Suchen durch Fluten (2)

- Fluten erreicht im n-ten Schritt alle n-ten Nachbarn
- *Frage:* Wie sind die Knoten bezogen auf ihren Abstand verteilt?
  - Baum: Zahl der Nachbarn wächst exponentiell mit dem Abstand L, d.h. k^L, wenn k der Knotengrad ist.
  - Gitter: Zahl der Nachbarn wächst mit L^{d-1} , wenn d die Dimension des Gitters ist.
  - Small-World-Netz: Viele meiner Nachbar sind verbunden (vgl. hoher Cluster-Koeffizient), d.h. Zahl der Nachbarn wächst weniger stark.
- *Bemerkung:* Trotz ihrer großen Bedeutung für die Entwicklung effizienter und skalierbarer Overlay-Protokolle gab es zunächst kaum Antworten auf diese Fragen.

Beispiel: Verkehrslast bei Gnutella

- Erste Protokollversionen verwenden TTL=7 beim Fluten und erreichen damit ca. 95% aller Knoten!

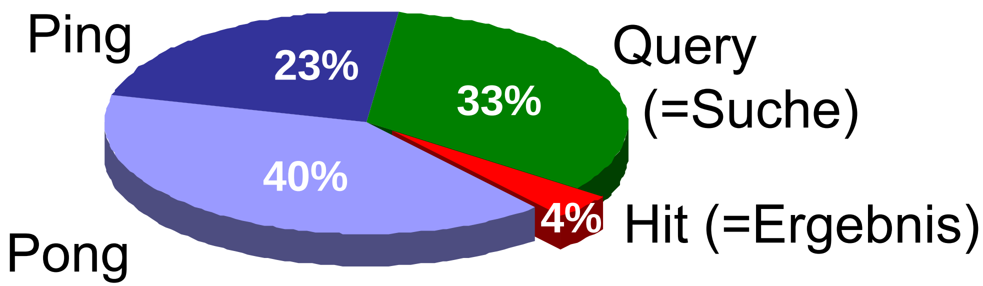
- Als Folge entsteht eine immense Netzlast, die kaum zum sinnvollem Nutzwert führt!



Quellen:

Matei Ripeanu, Adriana Iamnitchi, Ian Foster: „Mapping the Gnutella Network", IEEE Internet Computing, Jan/Feb 2002

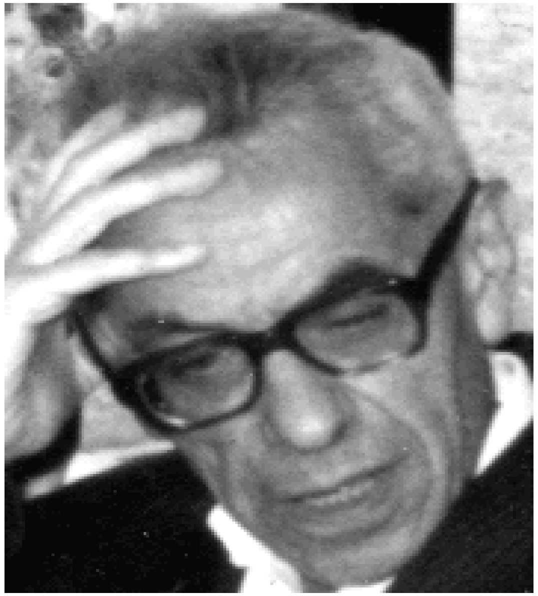
Zeinalipour-Yazti, Folias, Faloutsos, "A Quantitative Analysis of the Gnutella Network Traffic", Tech. Rep. May 2002



Einführung in die Zufallsgraphen

Paul Erdös

„Will I be able to read?“, asked Erdös before the corneal transplant on his bad eye. „Yes, that's the whole point of the Surgery.“, answered the surgeon.



Paul Erdös (1913-1996)

But when they turned the lights down for the surgery, Erdös complained, „You said I'd be able to read.“ He caused such a disturbance that they had to call in a mathematician from the local university to talk math while he had his operation.

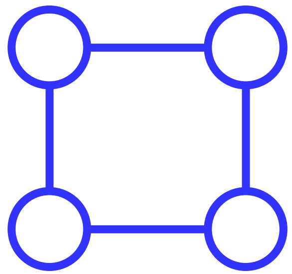
Graphen (1)

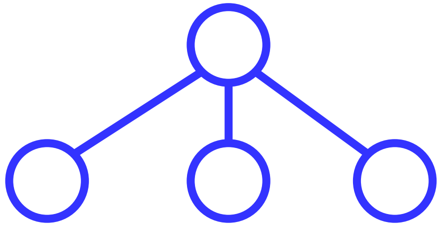
- Ein Graph kann definiert werden als Menge von Knoten und Kanten zwischen diesen Knoten

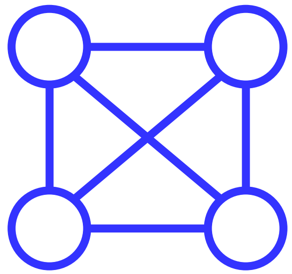
$$G = (V, E)$$

$$E \subset \{M \subset V : |M| = 2\}$$

- Diese Definition schränkt ein
  – Höchstens eine Kante zwischen zwei Knoten, d.h. keine Multigraphen
  – Keine Kante eines Knotens zu sich selbst
- Andere Definitionen könnten dies zulassen bzw. Graphen auch noch weiter verallgemeinern, beispielsweise auf
  – Gerichtete statt ungerichtete Kanten
  – Benannte statt unbenannte Kanten

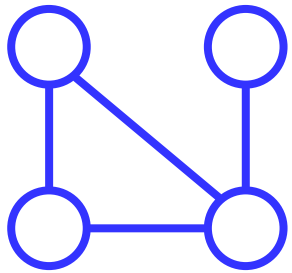


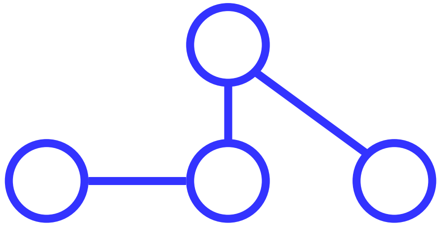


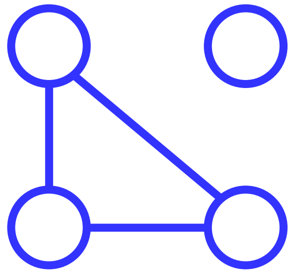


Graphen (2)

- Wie kann man Graphen unterscheiden?
- Wie viele Graphen gibt es überhaupt?
  - Mit gegebener Knotenzahl
  - Mit gegebener Kantenzahl
  - Mit gegebener Knoten- und Kantenzahl
- Beispiel:
  - Voll vermaschter Graph mit vier Knoten hat sechs Kanten
  - Also ist $2^6=64$ eine Obergrenze
  - Diese wird wegen der Ununterscheidbarkeit der Knoten nicht erreicht
  - Kombinatorik:
    Je ein Graph mit 0 bzw. 6 Kanten, je ein Graph mit 1 bzw. 5 Kanten, je zwei Graphen mit 2 bzw. 4 Kanten, drei Graphen mit 3 Kanten

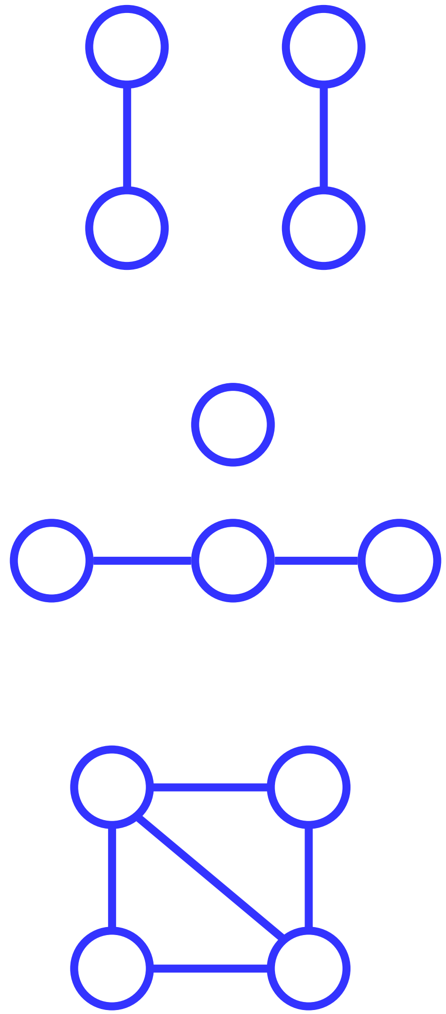






Graphen (3)

- Wie können diese Graphen weiter klassifiziert werden?
  - Sprich welche Eigenschaften haben sie?
  - … außer Knoten und Kantenzahl?
- Beispiel:
  - Zahl und Größe der Zusammenhangskomponenten
  - Durchmesser der Zusammenhangskomponenten



Zusammenhangskomponente

Definition: Zusammenhang eines Graphen

- Eine Folge von Kanten der Art (x,y_1), (y_1,y_2), …, (y_{n-1},y_n), (y_n,z) heißt ***Weg*** von x nach z.
- Ein Knoten x heißt ***erreichbar*** von y, wenn es einen Weg von y nach x gibt.
- Ein Graph heißt ***zusammenhängend***, wenn jeder Knoten von einem gegebenen Knoten aus erreichbar ist.
- Ein zusammenhängender Teilgraph $T \subset G$ heißt ***Zusammenhangskomponente***, wenn kein Knoten $x \notin T$ von $y \in T$ erreichbar ist.

Durchmesser eines Graphen

Definitionen: Abstand und Durchmesser

- Der kürzeste Weg von x nach y heißt ***Abstand*** von x und y.
- Der größte Abstand aller Knotenpaare eines zusammenhängenden Graphen heißt ***Durchmesser*** dieses Graphen.

- Der Durchmesser eines Graphen entspricht somit einer Worst-Case-Abschätzung für das Routing auf diesem Graphen.
  - Vgl. Definition von „Infinity“ in Distanzvektor-Routing-Protokollen: Übersteigt der scheinbare Abstand zweier Knoten den Durchmesser des Graphen, so weiß man, dass der Graph in (mindestens) zwei Zusammenhangskomponenten zerfallen ist.

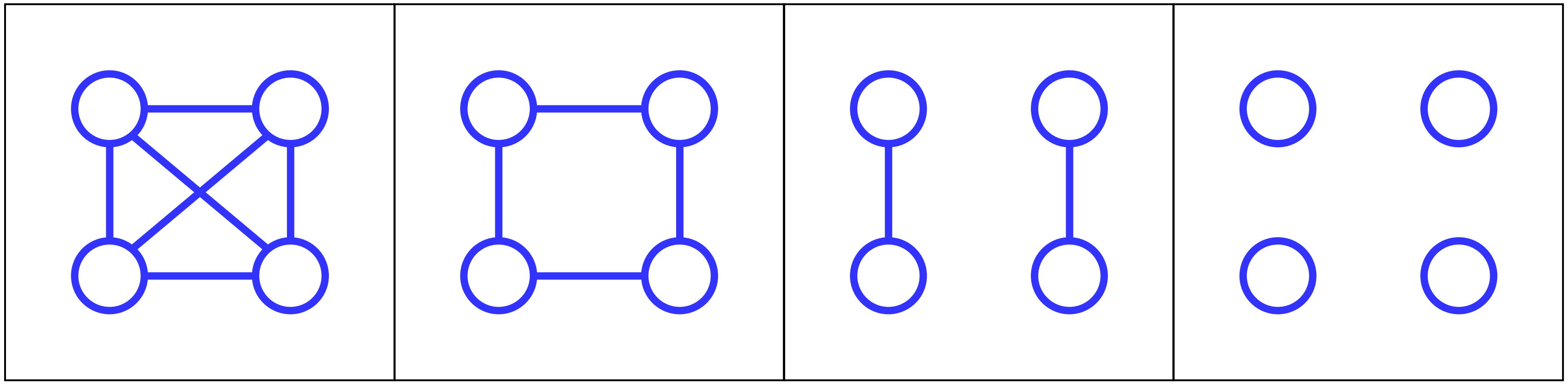
Grad eines Knotens

Definition: Grad eines Knotens

– Die Zahl von Kanten eines Knotens heißt ***Grad*** dieses Knotens.

Definition: Reguläre Graphen

– Haben alle Knoten eines Graphen den Grad r, so heißt der Graph ***r-regulär***.



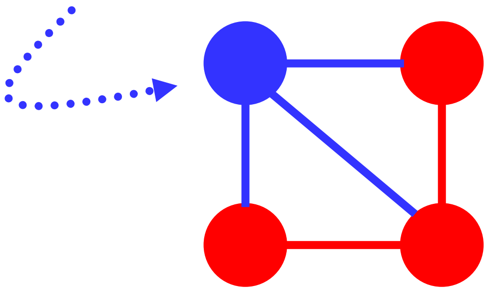
Durchmesser 1, 3-regulär	*Durchmesser 2, 2-regulär*	*Nicht zusammen-hängend, 1-regulär*	*Nicht zusammen-hängend, 0-regulär*

Cluster-Koeffizient

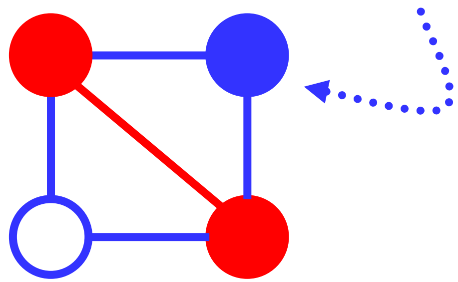
Definition: Cluster-Koeffizient

- Der Teilgraph Γ(x), der alle Knoten enthält, die von x Abstand 1 haben, heißt ***Nachbarschaft*** von x.
- Hat x den Grad n, so enthält Γ(x) n Knoten uns kann somit höchstes n(n-1)/2 Kanten haben.
- Der Quotient aus Kantenzahl in Γ(x) und n(n-1)/2 heißt ***Cluster-Koeffizient*** von x.

Cluster-Koeff. = 2/3



Cluster-Koeff. = 1/1



Mengen von Graphen (1)

- Neben diesen direkten Aussagen über einzelnen Eigenschaften eines Graphen bzw. einzelner seiner Knoten sind oft auch Häufigkeitsverteilungen von Interesse
  - Welchen Grad bzw. Cluster-Koeffizient haben die Knoten des Graphen?
  - Welchen Abstand haben Knotenpaare eines zusammenhängenden Graphen?
- Dieser Übergang von Einzelaussagen hin zu Häufigkeitsaussagen lässt sich noch ein Schritt weiter treiben:
  - Gegeben eine ***Menge von Graphen***, wie sind die Eigenschaften der Graphen in dieser Menge verteilt?
- Beispiel:
  - Welcher Anteil von Graphen mit n Knoten und m Kanten ist zusammenhängend?

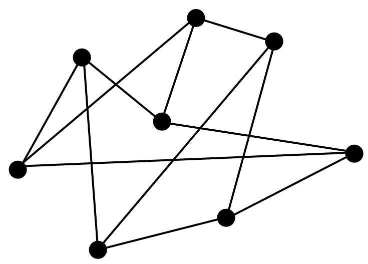
Mengen von Graphen (2)

- Hat man anhand bestimmter Kriterien eine Menge von Graphen definiert, kann man auch stochastische Fragen stellen:
  - Was ist der Erwartungswert für den Durchmesser eines Graphen?
  - Wie sind die Knotengrade in diesen Graphen verteilt?
- Mathematisch gesprochen, fasst man die Menge als ***Wahrscheinlichkeitsraum*** auf.
  - Dazu muss man ein Maß auf diesem Raum definieren, beispielsweise das Zählmaß.
- Schlussendlich interessieren Fragen der folgenden Art:
  - Mit welcher Wahrscheinlichkeit ist ein beliebiger Graph (sprich Netzwerk) zusammenhängend?
  - Wieviel Prozent der Kanten (sprich Leitungen) müssen entfernt werden, damit ein zusammenhängender Graph in zwei Zusammenhangskomponenten zerfällt?

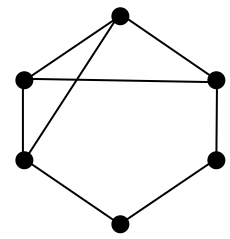
Unterscheidbarkeit

- Solche Betrachtungen werden dadurch erschwert, dass
  - nicht immer einfach nachzuvollziehen ist, ob zwei vermeintlich verschiedene Graphen nicht doch identisch sind.
  - verschiedene Konstruktionen zum gleichen Graphen führen
- Beispiel:
  - Der vollvermaschte Graph mit für Knoten hat 6 Kanten.
  - Es wäre einfach, wenn man alle 2^6=64 Kombinationen durchprobieren könnte.
  - Wenn aber Kanten und Knoten ununterscheidbar sind, führt das häufig zu identischen Graphen.
- Um dieses Problem zu umgehen, definiert man häufig Knoten und Kanten als *unterscheidbar*.

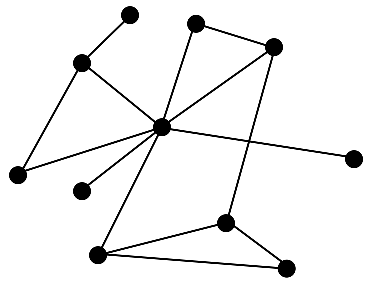
Zufallsgraphen (1)



- Ca. 1960 begannen Mathematiker mit der Untersuchung von Zufallsgraphen, allen voran Paul Erdös.
- Grundidee
  - Betrachte Wahrscheinlichkeitsraum von Graphen
  - D.h. konstruiere Graphen und weise ihnen ein Wahrscheinlichkeitsmaß zu



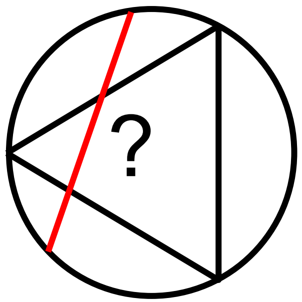
- Beispiel: Erdös-Renyi untersuchten
  - Zufallsgraphen mit konstanter Kantenwahrscheinlichkeit
  - bzw. konstanter Kantenzahl

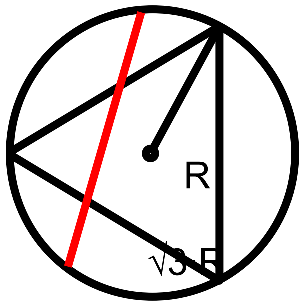


- Aber, solche Zufallsgraphen sind nur eine Möglichkeit von vielen …

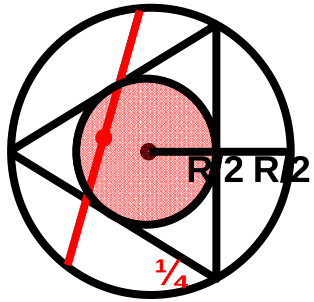
Zufallsgraphen (2)

Aber, welche Art von Zufall … ?

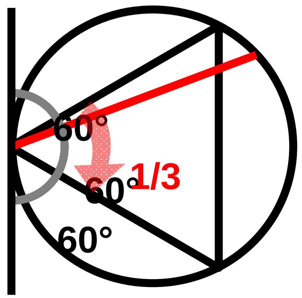




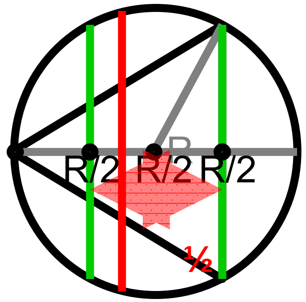
~ 0,134



0,25



0,333



0,5

⇒ Es gibt eine große Fülle unterschiedlicher Definitionen von Zufallsgraphen mit jeweils unterschiedlichen Eigenschaften!

Zufallsgraphen (3)

Frage:

- Gibt es allgemeine Grundprinzipien bei Zufallsgraphen?
- Wie kann man Zufallsgraphen klassifizieren?
- Welche Eigenschaften haben die verschiedenen Klassen von Zufallsgraphen?

Andeutung einer Antwort:

- Viele Eigenschaften von Zufallsgraphen werden durch Schwellenfunktionen beschrieben, d.h. übersteigt ein Parameter eine bestimmte Grenze, wächst die Wahrscheinlichkeit, dass eine gegebene Eigenschaft vorliegt, schlagartig auf fast 100%
- Für viele Eigenschaften von Zufallsgraphen kann man Wahrscheinlichkeitsdichten angeben, beispielsweise Knotengrade, Durchmesser, etc.

Small-World Zufallsgraphen

- Seit ihrer ersten Untersuchung vor ca. 40 Jahren sind Zufallsgraphen nicht nur mathematisch interessant.
- Vor ca. 30 Jahren entdecken Soziologen, dass persönliche Kontakte Gesellschaften sehr eng verknüpfen
  - „Ein Bekannter des Bekannten meines Bekannten ...“
  - Erreicht nach ca. 5-6 Schritten die gesamten USA
  - Ähnlich für berufliche Zusammenarbeit (Schauspieler, Publikationen, ...)
  - Prägung des Begriffs „***Small-World-Netze***“
  - Typisch: Einige wenige Menschen knüpfen sehr viele Kontakte. Weniger gut „vernetzte“ Menschen nutzen diese „Vermittler“.
- Seit Ende 1990er Jahre verstärkte Untersuchung derartiger Zusammenhänge, insbesondere in der statistischen Physik
  - Small-World-Netzen liegt (oft) eine Verteilung der Knotengrade nach einem Potenzgesetz zugrunde, d.h. einige wenige Knoten haben sehr viele Verbindungen (siehe oben)
  - Bildungsmodell für „Small-World-Netze“ (Barabasi & Albert)

Erdös-Renyi-Zufallsgraphen (1)

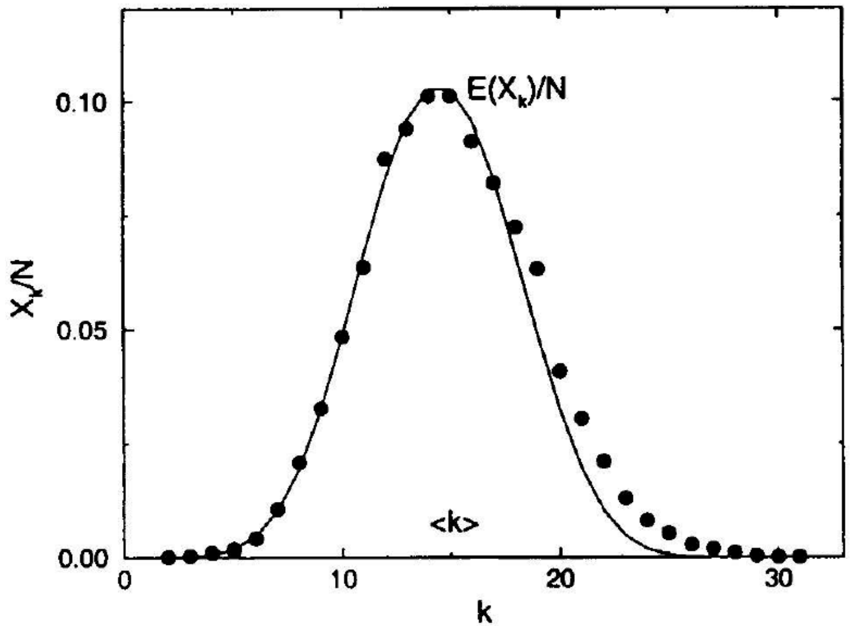
- Vollvermaschter Graph mit n Knoten hat $N = \binom{n}{2}$ Kanten.
- Somit gibt es $\binom{N}{M}$ Graphen mit n Knoten und M Kanten. (Achtung: Wir setzen hier Unterscheidbarkeit der Knoten voraus.)
- ***Erdös-Renyi Zufallsgraphen***:
  - Graphen mit fester Kantenzahl. – Jedem Graph wird die Wahrscheinlichkeit
    $$\binom{N}{M}^{-1}$$
    zugeordnet.
  - Graphen mit fester Kantenwahrscheinlichkeit. – Jedem Graph wird die Wahrscheinlichkeit
    $$p^M (1-p)^{N-M}$$
    zugeordnet.

Erdös-Renyi-Zufallsgraphen (2)

Beim Erdös-Renyi-Zufallsgraphen mit fester Kantenwahrscheinlichkeit erhalten wir eine Poisson-Verteilung für die Knotengrade:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\simeq \frac{z^k e^{-z}}{k!}$$



Simulation mit N=10.000 und p=0.15%

wobei z der Erwartungswert der Knotengrade ist. Dieses analytische Ergebnis deckt sich gut mit Simulationsergebnissen.

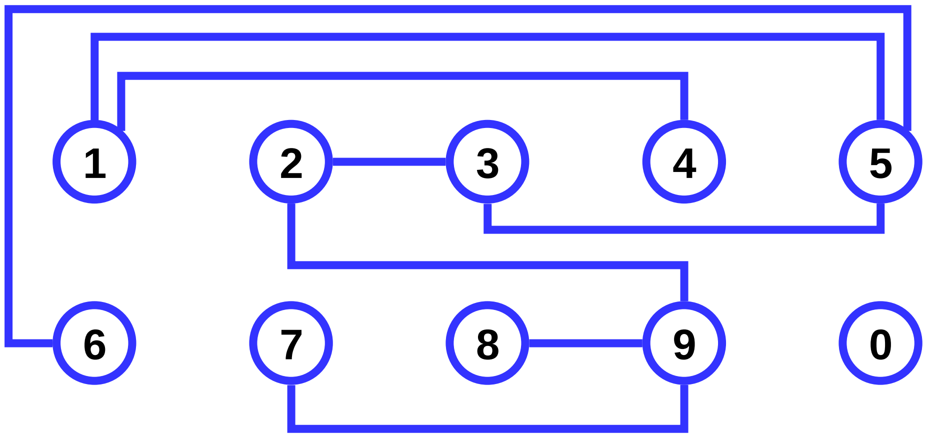
Schwellenfunktionen

- Viele Sätz über Zufallsgraphen beziehen sich auf Aussagen, die im Grenzwert $n \to \infty$ gelten.
- Man spricht in diesem Fall davon, dass ***fast alle*** Graphen (einer ggf. zuvor definierten Menge) eine bestimmte Eigenschaft haben.
- Häufig tritt eine solche Eigenschaft plötzlich auf, in dem Sinne, dass es eine ***Schwellenfunktion*** (engl. *threshold function*) gibt,
  - die eine Menge von Graphen in zwei disjunkte Teilmengen trennt,
  - bei denen die Graphen der einen fast alle diese Eigenschaft haben,
  - die der anderen hingen fast alle diese Eigenschaft nicht haben.
- Man sagt auch, für diese Eigenschaft gilt ein ***0-1-Gesetz***.

Schwellenfkt. für Zusammenhang (1)

Beispiel:

– Graph mit zehn Knoten
– Nimm Ziffernpaare einer zufälligen Ziffernfolge und verbinde entsprechende Knoten:



3, 14 15 92 65 35 89 79 32 38 46 26 43 38 32 70

Ergebnis:

– Ab der achten Kante besteht der Graph aus einer großen Zusammenhangskomponente und einem unverbundenen Knoten.

Schwellenfkt. für Zusammenhang (2)

Satz (Erdös-Renyi, 1959):

$$M(n) = \frac{n}{2} \ln n$$

ist eine Schwellenfunktion für die Kantenzahl eines regulären Zufallsgraphen mit n *Knoten (und fester Kantenzahl) und die Eigenschaft des Graphen, zusammenhängend zu sein*

Bemerkung: Anders ausgedrückt ist ***ln n*** eine Schwellenfunktion für den mittleren Knotengrad und den Zusammenhang eines Graphen.

Schwellenfkt. für Zusammenhang (3)

Satz (Kovalenko, 1979):

$$p(n) = \frac{1}{n} \ln n$$

ist eine Schwellenfunktion dafür, dass Zufallsgraphen mit fester Kantenwahrscheinlichkeit aus einer (sehr großen) Zusammenhangskomponente und einzelnen, isolierten Knoten bestehen.

Die Zahl der isolierten Knoten ist Poisson-verteilt mit Erwartungswert

$$\lambda = np^n$$

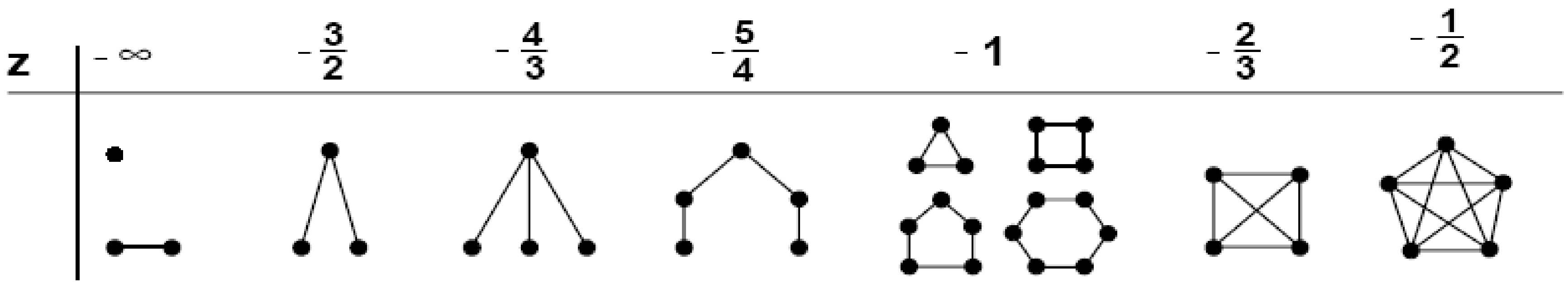
Schwellenfkt. für Zusammenhang (4)

- Beachte: In beiden Fällen (Erdös-Renyi, 1959 und Kovalenko, 1979) ist ***ln n*** eine Schwellenfunktion für den Knotengrad bezüglich des Zusammenhangs des Graphen.
  - In unserem Beispiel finden wir 1.6, aus der mathematischen Analyse erwarten wir 2.3.
  - Aber Schwellenfunktionen machen ja nur Aussagen über den Grenzwert sehr großer Graphen.
- Bemerkung: Bäume, als Spezialfall eines zusammenhängenden Graphen haben einen mittleren Knotengrad von 2-2/n, für große Bäume also ca. 2.

Weitere Schwellenfunktionen

Kantenwahrscheinlichkeit im Erdös-Renyi-Graphen:

$$p \sim N^z$$



Bäume

Isolierte Knoten und Knotenpaare

Zyklen aller Größen entstehen

Schrittweise immer größere, voll vermaschte Teilgraphen

Quelle: Reka Albert, PhD Thesis, Notre Dame University

Folgerungen für Overlay-Netze

- Für Overlay-Netzwerke folgt daraus beispielsweise
  - Bei einer Million Rechnern muss jeder Rechner mindestens 14 Verbindungen zu beliebigen* anderen Rechnern haben, damit das Netz zusammenhängend ist.
- Aber, reale Overlay-Netze funktionieren anders:
  - Die Wahrscheinlichkeit, Verbindungen mit „wichtigen“ Rechnern zu halten ist höher als für Verbindungen mit „durchschnittlichen“ Rechnern.
  - Für die meisten Rechner genügt also eine Verbindung bzw. einige wenige Verbindungen, solange es einige wenige Rechner gibt, die sehr viele Verbindungen halten können.
- Dies ist das Grundprinzip der so genannten Small-World-Netze (→nächste Vorlesung).

* gleichverteilt zufällig gewählt

Zusammenfassung

- Die Wahrscheinlichkeitsrechnung misst die „relative Größe“ von Teilmengen einer Menge:
  - Teilmengen definiert durch eine Eigenschaft, oder
  - Teilmengen definiert als Urbild einer Teilmenge eines reellen Wertebereichs (vgl. Zufallsvariable).
- Bei Zufallsgraphen wendet man die W-Rechnung auf Graphen an.
- Betrachtete Eigenschaften sind z.B. Zusammenhang, Abstand von Knoten, Durchmesser, Knotengrad, Clusterkoeffizient, etc.
- Oft sind die zugehörigen Verteilungsfunktionen Schwellenfunktionen:
  - D.h. eine Eigenschaft ändert sich (fast) schlagartig beim Ändern eines Parameters.
  - Im Umkehrschluss: Solange der kritische Wert des Parameters nicht erreicht ist, verhalten sich die Graphen trotz „Zufall“ gleich!

Questions?

Thomas Fuhrmann

Department of Informatics
Self-Organizing Systems Group
c/o I8 Network Architectures and Services
Technical University Munich, Germany

fuhrmann@net.in.tum.de