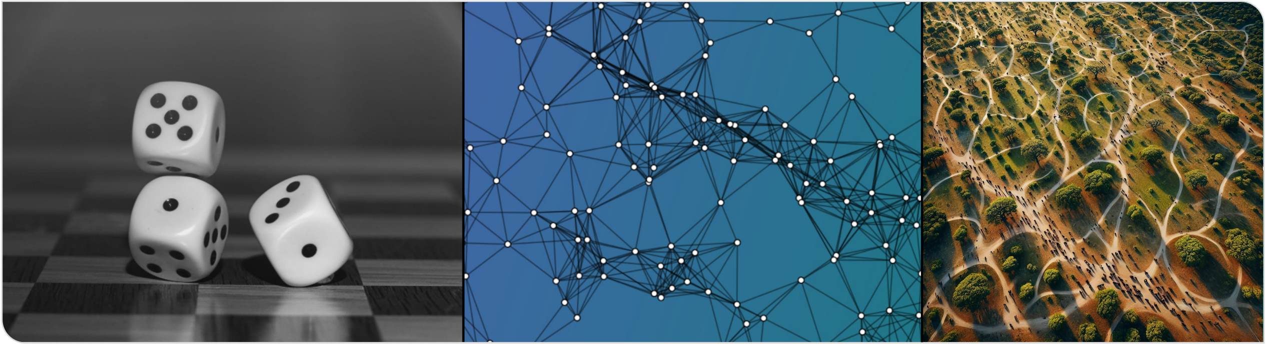
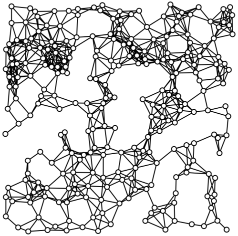
Probability and Computing – Random Graphs

Stefan Walzer | WS 2024/2025



Content

Motivation 1: Average Case Analysis



Theory-Practice Gap

Minimum Vertex Cover is APX-hard $\overset{???}{\longleftrightarrow}$ small vertex covers can often be computed efficiently in practice

$\rightsquigarrow$ relevant graph classes (e.g. social networks) are not worst-case.

Bridging the Gap

1. Define a distribution $\mathcal{G}$ on graphs.
   - $\mathcal{G}$ should be realistic, i.e. model real world instances
   - $\mathcal{G}$ should have simple mathematical structure
2. Consider randomised complexity of handling $G \sim \mathcal{G}$.

Goals

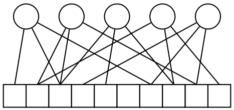
- model real world instances
- identify useful properties of these instances
- build algorithms exploiting these properties

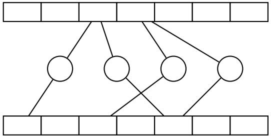
Motivation 2: Data Structure Design

Stay tuned

Random graphs occur naturally in

- cuckoo hash tables
- retrieval data structures
- perfect hash functions





Content

1. Motivation

2. Erdős-Renyi Random Graphs
   - Degree Distribution
   - Degree Statistics
   - Tree-like local structure
   - Emergence of the Giant Component

3. Random Geometric Graphs

4. Scale-Free Networks (Teaser)

The Erdős-Renyi Model and Related Distributions

Original Erdős-Renyi Model $G(n, m)$: "Uniformly random graph with n nodes and m edges"

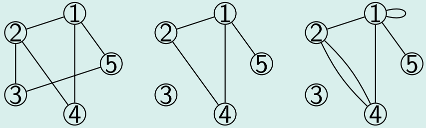
Gilbert Model $G(n, p)$: "Every edge with probability p"

Uniform Endpoint Model $G^{\text{UE}}(n, m)$: "randomly attach the $2m$ endpoints of edges"

Definition

Let $n \in \mathbb{N}$, $0 \leq m \leq \binom{n}{2}$. We use $G(n, m)$ to refer to a graph sampled uniformly from the set of all graphs with vertex set $[n]$ and m edges.

Example: $n = 5$, $m = 6$



probability	$1/\binom{\binom{n}{2}}{m}$	0	0

The Erdős-Renyi Model and Related Distributions

Original Erdős-Renyi Model $G(n, m)$: "Uniformly random graph with n nodes and m edges"

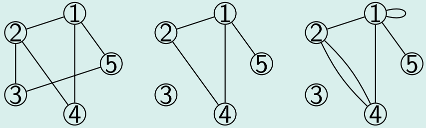
Gilbert Model $G(n, p)$: "Every edge with probability p"

Uniform Endpoint Model $G^{\mathrm{UE}}(n, m)$: "randomly attach the $2m$ endpoints of edges"

Definition

Let $n \in \mathbb{N}$ and $p \in (0, 1)$. We use $G(n, p)$ to refer to a graph with vertex set $[n]$ that contains each of the $\binom{n}{2}$ possible edges with probability p, independently from other edges.

Example: $n = 5$



probability	$p^6(1-p)^4$	$p^4(1-p)^6$	0

The Erdős-Renyi Model and Related Distributions

Karlsruhe Institute of Technology

Original Erdős-Renyi Model $G(n, m)$: "Uniformly random graph with n nodes and m edges"

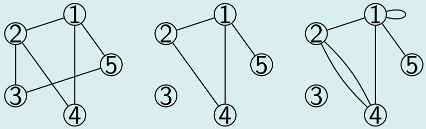
Gilbert Model $G(n, p)$: "Every edge with probability p"

Uniform Endpoint Model $G^{\mathrm{UE}}(n, m)$: "randomly attach the $2m$ endpoints of edges"

Definition

Let $n, m \in \mathbb{N}$ and $v_1, \ldots, v_{2m} \sim \mathcal{U}([n])$. We use $G^{\mathrm{UE}}(n, m)$ to refer to a multi-graph with vertex set $[n]$ and a multiset of edges that contains a copy of $\{v_{2i-1}, v_{2i}\}$ for each $i \in [m]$.

Example: $n = 5, m = 6$



probability $6! \cdot 2^6 \cdot 5^{-12}$ 0 $6! \cdot 2^4 \cdot 5^{-12}$

The Erdős-Renyi Model and Related Distributions

Karlsruhe Institute of Technology

Original Erdős-Renyi Model $G(n, m)$: "Uniformly random graph with n nodes and m edges"

Gilbert Model $G(n, p)$: "Every edge with probability p"

Uniform Endpoint Model $G^{\text{UE}}(n, m)$: "randomly attach the $2m$ endpoints of edges"

Remarks

- for $p = m/\binom{n}{2}$ the three distributions are similar in many ways
- the original Erdős-Renyi model is often inconvenient to work with
- the uniform endpoint model is non-standard (we'll need it in later chapters)

Plan for the Next Few Slides: Sparse Graphs

Karlsruhe Institute of Technology

Focus on Expected Degree $\lambda \in \mathcal{O}(1)$

- for $G(n, m)$ choose $m = \frac{\lambda n}{2}$ $\Rightarrow$ average vertex degree $\frac{2m}{n} = \lambda$
- for $G(n, p)$ choose $p = \frac{\lambda}{n-1}$ $\Rightarrow$ expected vertex degree $(n-1) \cdot p = \lambda$
- for $G^{\mathrm{UE}}(n, m)$ choose $m = \frac{\lambda n}{2}$ $\Rightarrow$ average vertex degree $\frac{2m}{n} = \lambda$ // loops contribute 2 to a vertex degree

Goals

- Build intuition for properties of Erdős-Renyi graphs.
- Get a feeling for how to work with them.
- For simplicity: Focus on the Gilbert model only.

Selected Properties of Erdős-Renyi Graphs

Karlsruhe Institute of Technology

On the next few slides we consider:

Vertex Degrees

For large n, the degree of a given vertex is approximately Poisson distributed.

Local Structure

The neighbourhood around a vertex resembles a Galton-Watson tree.

Degree Statistics

The number of vertices of each degree is highly concentrated around its expectation.

Largest Connected Component

Size of the largest component is highly predictable.

Content

Degree Distribution of Erdős-Reyni Graphs

Exercise: Degrees are approximately Poisson distributed

For each $n \in \mathbb{N}$ consider $G(n, \lambda/n)$ and the degree $X_n \sim \mathrm{Bin}(n-1, \lambda/n)$ of vertex 1. Moreover, let $X \sim \mathrm{Pois}(\lambda)$. Then

$$X_n \xrightarrow{d} X \text{ for } n \to \infty.$$

The same holds for $G(n, \lfloor \lambda n/2 \rfloor)$ and $G^{\mathrm{UE}}(n, \lambda n/2)$.

Content

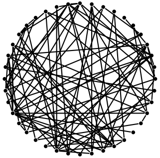
The Number N_d of Vertices of Degree N_d

Notation

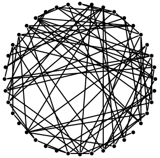
- Let $d \in \mathbb{N}$, $\lambda > 0$. We consider $G(n, \lambda/n)$. // Gilbert model
- Let $N_d := |\{v \in [n] \mid \deg(v) = d\}|$

Is N_d highly concentrated?

- Note: $(\deg(v))_{v \in [n]}$ are correlated.
- Otherwise N_d would have a binomial distribution and we could use Chernoff bounds.



d	0	1	2	3	4	5	6	7	8	9
N_d	0	2	8	6	7	7	3	2	3	1



d	0	1	2	3	4	5	6	7	8	9
N_d	1	2	5	8	9	11	3	1	0	0

Concentration of N_d

Lemma (Near Independence of Degrees)

Let $u \neq v$ be two vertices of $G(n, \lambda/n)$. Then $\Pr[\deg(u) = d, \deg(v) = d] = \Pr[\deg(u) = d]\Pr[\deg(v) = d] \pm \Theta(1/n)$.

Proof

Let $\deg'(u) = \deg(u) - [\{u, v\} \in E]$ be the degree of u when ignoring $\{u, v\}$ if present. Then

$$\Pr[\deg(u) \neq \deg'(u)] = \Pr[\{u, v\} \in E] = \lambda/n = \Theta(1/n).$$

The same holds for $\deg'(v) = \deg(v) - [\{u, v\} \in E]$. We conclude:

$$\begin{aligned}\Pr[\deg(v_1) = d, \deg(v_2) = d] &= \Pr[\deg'(v_1) = d, \deg'(v_2) = d] \pm \Theta(1/n)\\ &= \Pr[\deg'(v_1) = d]\Pr[\deg'(v_2) = d] \pm \Theta(1/n) = \Pr[\deg(v_1) = d]\Pr[\deg(v_2) = d] \pm \Theta(1/n).\end{aligned}$$

Concentration of N_d

Lemma (Near Independence of Degrees)

Let $u \neq v$ be two vertices of $G(n, \lambda/n)$. Then $\Pr[\deg(u) = d, \deg(v) = d] = \Pr[\deg(u) = d] \Pr[\deg(v) = d] \pm \Theta(1/n)$.

Theorem

$\Pr[|N_d - np_d| \geq n^{2/3}] = \mathcal{O}(n^{-1/3})$ where $p_d = \Pr[\deg(1) = d] \approx e^{-\lambda} \frac{\lambda^d}{d!}$.

Proof

$\mathbb{E}[N_d] = np_d$ $\qquad$ $\mathbb{E}[N_d^2] = n^2 p_d^2 \pm \mathcal{O}(n)$ $\qquad$ $\mathrm{Var}(N_d) = \mathcal{O}(n)$.

$$\mathbb{E}[N_d] = \mathbb{E}[\sum_{v \in [n]} [\deg(v) = d]] = n \cdot \Pr[\deg(1) = d] = n \cdot p_d.$$

Concentration of N_d

Lemma (Near Independence of Degrees)

Let $u \neq v$ be two vertices of $G(n, \lambda/n)$. Then $\Pr[\deg(u) = d, \deg(v) = d] = \Pr[\deg(u) = d] \Pr[\deg(v) = d] \pm \Theta(1/n)$.

Theorem

$\Pr[|N_d - np_d| \geq n^{2/3}] = \mathcal{O}(n^{-1/3})$ where $p_d = \Pr[\deg(1) = d] \approx e^{-\lambda} \frac{\lambda^d}{d!}$.

Proof

$\mathbb{E}[N_d] = np_d$ $\qquad$ $\mathbb{E}[N_d^2] = n^2 p_d^2 \pm \mathcal{O}(n)$ $\qquad$ $\mathrm{Var}(N_d) = \mathcal{O}(n)$.

$$\mathbb{E}[N_d^2] = \mathbb{E}\Big[\Big(\sum_{v \in [n]} [\deg(v) = d]\Big)^2\Big] = \mathbb{E}\Big[\sum_{u \in [n]} \sum_{v \in [n]} [\deg(u) = d, \deg(v) = d]\Big]$$

$$= \sum_{u \in [n]} \sum_{v \in [n]} \Pr[\deg(u) = d, \deg(v) = d] = \sum_{u \in [n]} \Pr[\deg(u) = d] + \sum_{u \in [n]} \sum_{v \neq u} \Pr[\deg(u) = d, \deg(v) = d]$$

$$= n \cdot p_d + n \cdot (n-1) \cdot (p_d^2 \pm \mathcal{O}(1/n)) = n^2 p_d^2 \pm \mathcal{O}(n).$$

Concentration of N_d

Lemma (Near Independence of Degrees)

Let $u \neq v$ be two vertices of $G(n, \lambda/n)$. Then $\Pr[\deg(u) = d, \deg(v) = d] = \Pr[\deg(u) = d] \Pr[\deg(v) = d] \pm \Theta(1/n)$.

Theorem

$\Pr[|N_d - np_d| \geq n^{2/3}] = \mathcal{O}(n^{-1/3})$ where $p_d = \Pr[\deg(1) = d] \approx e^{-\lambda} \frac{\lambda^d}{d!}$.

Proof

$\mathbb{E}[N_d] = np_d$ $\qquad$ $\mathbb{E}[N_d^2] = n^2 p_d^2 \pm \mathcal{O}(n)$ $\qquad$ $\mathrm{Var}(N_d) = \mathcal{O}(n)$.

$$\mathrm{Var}(N_d) = \mathbb{E}[N_d^2] - \mathbb{E}[N_d]^2 \leq n^2 p_d^2 + \mathcal{O}(n) - (np_d)^2 = \mathcal{O}(n).$$

Concentration of N_d

Lemma (Near Independence of Degrees)

Let $u \neq v$ be two vertices of $G(n, \lambda/n)$. Then $\Pr[\deg(u) = d, \deg(v) = d] = \Pr[\deg(u) = d] \Pr[\deg(v) = d] \pm \Theta(1/n)$.

Theorem

$\Pr[|N_d - np_d| \geq n^{2/3}] = \mathcal{O}(n^{-1/3})$ where $p_d = \Pr[\deg(1) = d] \approx e^{-\lambda} \frac{\lambda^d}{d!}$.

Proof

$\mathbb{E}[N_d] = np_d$ $\quad$ $\mathbb{E}[N_d^2] = n^2 p_d^2 \pm \mathcal{O}(n)$ $\quad$ $\mathrm{Var}(N_d) = \mathcal{O}(n)$.

$$\text{Hence: } \Pr[|N_d - np_d| \geq n^{2/3}] = \Pr[|N_d - \mathbb{E}[N_d]| \geq n^{2/3}] \overset{\text{Cheb.}}{\leq} \frac{\mathrm{Var}(N_d)}{n^{4/3}} = \mathcal{O}(n^{-1/3}).$$

Content

Erdős-Renyi Graphs have Few Cycles

Karlsruhe Institute of Technology

Theorem: There are few short cycles in Erdős-Renyi graphs

Let C_k be the number of cycles of length k in $G(n, \lambda/n)$ where $k, \lambda = \Theta(1)$. Then $\mathbb{E}[C_k] \leq \frac{\lambda^k}{2k} = \Theta(1)$.

Proof.

The number of potential cycles is $\underbrace{n(n-1) \cdot \ldots \cdot (n-k+1)}_{\text{sequences } (v_1, \ldots, v_k)} \cdot \underbrace{\frac{1}{k} \cdot \frac{1}{2}}_{\text{startpoint and direction irrelevant}}$

The probability that $(v_1, \ldots, v_k, v_1)$ is a cycle is $(\lambda/n)^k$. Hence:

$$\mathbb{E}[C_k] \leq \frac{n^k}{2k}\left(\frac{\lambda}{n}\right)^k = \frac{\lambda^k}{2k}.$$

□

The Galton-Watson Branching Process

Karlsruhe Institute of Technology

Definition

Let $\mathcal{D}$ be a distribution on $\mathbb{N}_0$ and $X_{i,j} \sim \mathcal{D}$ for $i, j \in \mathbb{N}$. Define $Z_0 = 1$ and $Z_i = \sum_{j=1}^{Z_{i-1}} X_{i,j}$ for $i \geq 1$.

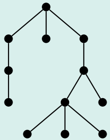
Intuition

- Start with a population of size $Z_1 = 1$.
- Each individual has a random number of decendents.
- Key question: What is the probability of extinction, i.e. for $\lim_{i\to\infty} Z_i = 0$?

Exercise: Galton-Watson Process with $\mathcal{D} = \text{Pois}(\lambda)$

If $\lambda \leq 1$ then the process goes extinct with probability 1.
If $\lambda > 1$ then the process survives with probability $s_\lambda > 0$.

Galton-Watson *Tree*



$X_{i,j}$	1	2	3	4	...
1	3	1	0	2	...
2	1	0	1	3	...
3	1	2	2	0	...
4	0	3	0	0	...
5	0	0	0	2	...
⋮	⋮	⋮	⋮	⋮	

Local Structure of Erdős-Renyi Graphs

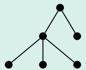
Karlsruhe Institute of Technology

Theorem: The Neighbourhood of v looks like a Galton Watson Tree

Let $R = \mathcal{O}(1)$. Let H be an ordered tree of depth R given by a sequence $c_1, \ldots, c_k$ specifiying the number of children of nodes in all layers except the last, in level order.
Let $\mathrm{GWT}(\lambda)|_R$ be the first R layers of a Pois(λ)-Galton-Watson tree.
Let $G(n, \lambda/n)|_{v,R}$ be the subgraph of $G(n, \lambda/n)$ induced by vertices with distance $\leq R$ from v.

$$\Pr[\mathrm{GWT}(\lambda)|_R = H] \overset{\text{(i)}}{=} \prod_{i=1}^{k} \Pr_{X\sim\mathrm{Pois}(\lambda)}[X = c_i] = \prod_{i=1}^{k} e^{-\lambda}\frac{\lambda^{c_i}}{c_i!} \overset{\text{(ii)}}{\approx} \Pr[G(n, \lambda/n)|_{v,R} = H].$$

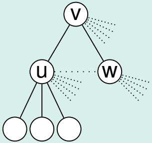
Example for H



$(c_1, c_2, c_3) = (2, 3, 0)$

Proof of (ii) by Example: The following has to "go right" for $G(n, \lambda/n)|_{v,R} = H$

random variable	**desired outcome**	**probability**
$\deg(v) \sim \mathrm{Bin}(n-1, \lambda/n) \approx \mathrm{Pois}(\lambda)$	2	$\approx e^{-\lambda}\frac{\lambda^2}{2!}$
$[\{u, w\} \in E]$	0	$1 - \frac{\lambda}{n} \approx 1$
$\deg(u) - 1 \sim \mathrm{Bin}(n-3, \lambda/n) \approx \mathrm{Pois}(\lambda)$	3	$\approx e^{-\lambda}\frac{\lambda^3}{3!}$
$\deg(w) - 1 \sim \mathrm{Bin}(n-3, \lambda/n) \approx \mathrm{Pois}(\lambda)$	0	$\approx e^{-\lambda}\frac{\lambda^0}{0!}$



Local Structure of Erdős-Renyi Graphs

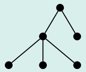
Karlsruhe Institute of Technology

Theorem: The Neighbourhood of v looks like a Galton Watson Tree

Let $R = \mathcal{O}(1)$. Let H be an ordered tree of depth R given by a sequence $c_1, \ldots, c_k$ specifiying the number of children of nodes in all layers except the last, in level order.
Let $\mathrm{GWT}(\lambda)|_R$ be the first R layers of a $\mathrm{Pois}(\lambda)$-Galton-Watson tree.
Let $G(n, \lambda/n)|_{v,R}$ be the subgraph of $G(n, \lambda/n)$ induced by vertices with distance $\leq R$ from v.

$$\Pr[\mathrm{GWT}(\lambda)|_R = H] \overset{\text{(i)}}{=} \prod_{i=1}^{k} \Pr_{X\sim\mathrm{Pois}(\lambda)}[X = c_i] = \prod_{i=1}^{k} e^{-\lambda}\frac{\lambda^{c_i}}{c_i!} \overset{\text{(ii)}}{\approx} \Pr[G(n, \lambda/n)|_{v,R} = H].$$

Example for H



$(c_1, c_2, c_3) = (2, 3, 0)$

Corollaries

- $G(n, \lambda/n)|_{v,R} \xrightarrow{d} \mathrm{GWT}(\lambda)|_R$ // convergence in distribution for $n \to \infty$
- The number N_H of "copies" of H in $G(n, \lambda/n)$ satisfies $\mathbb{E}[N_H] \approx n \cdot \prod_{i=1}^{k} e^{-\lambda}\frac{\lambda^{c_i}}{c_i!}$.
  Concentration of N_H can be proved much like we proved concentration of N_d earlier.

Content

How does $G(n, \lambda/n)$ look like for different λ?

Karlsruhe Institute of Technology

Theorem: Sudden Emergence of the Giant Component (Erdős, Renyi 1960)

Consider $G(n, \lambda/n)$. The following holds with probability approaching 1 for $n \to \infty$.

- **i** If $\lambda < 1$ then $G(n, \lambda/n)$ only has components of size $\mathcal{O}(\log n)$.
  Each component is a tree or pseudotree. pseudotree means: connected and as many edges as vertices
  $\hookrightarrow$ Intuition: GWT(λ) dies out with probability 1.
- **ii** If $\lambda > 1$ then $G(n, \lambda/n)$ has one "giant" component of size $\approx s(\lambda) \cdot n$.
  $\hookrightarrow$ Intuition: $s(\lambda)$ is the probability that GWT(λ) is infinite.
- **iii** If $\lambda = 1$ then the largest component of $G(n, \lambda/n)$ has size $\Theta(n^{2/3})$.
  $\hookrightarrow$ Intuition: ?

Content

Locality: A Property of Networks in Practice

Karlsruhe Institute of Technology

Observation: Locality in Practice

Take social networks. A friend of my friend is more likely to be my friend than a random person.

Definition: Locality[1]

$L = \Pr[\{u, w\} \in E \mid \{v, u\} \in E \wedge \{v, w\} \in E]$ where v, u, w are distinct (random) vertices.

[1] Similar numbers are sometimes called *clustering coefficient*.

Observation: No Locality in Erdős-Renyi Random Graphs

In $G(n, \lambda/n)$ we have $L = \frac{\lambda}{n} = \mathcal{O}(\frac{1}{n})$.

Next: Random *Geometric* Graphs with $L = \Omega(1)$.

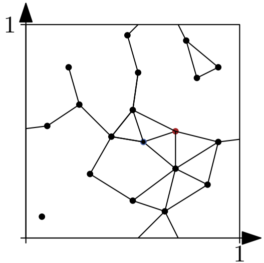
Random Geometric Graphs

Definition: Random Geometric Graph (RGG)

An RGG is obtained by distributing vertices in a metric space and connecting any two vertices with a probability depending on their distance.

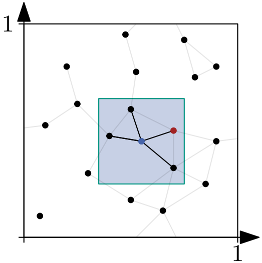
Simple Example: $G^{\mathbb{T}^2}(n, r)$

- number of vertices: n
- space: 2-dimensional torus $\mathbb{T}^2 = [0, 1)^2$
  // standard unit square is more common but less simple
- metric: L_∞ // L_2 is more common but less simple
  $\hookrightarrow \text{dist}((x_1, y_1), (x_2, y_2)) = \max(\text{dist}(x_1, x_2), \text{dist}(y_1, y_2))$.
- vertex distribution: for $v \in [n]$: $P_v \sim \mathcal{U}(\mathbb{T}^2)$
- edge “probability” is 0 or 1: $\{u, v\} \in E \Leftrightarrow \text{dist}(P_u, P_v) \le r$
  // not random when P_u and P_v are given



Degree Distribution of $G^{\mathbb{T}^2}(n, r)$

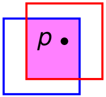
- Consider arbitrary $v \in [n]$.
- By symmetry of $\mathbb{T}^2$ each outcome of P_v behaves the same.
- $\Pr[\{u, v\} \in E] = \Pr[P_u \text{ is in the } 2r \times 2r \text{ square centered at } P_v] = 4r^2$.
- Hence $\deg(v) \sim \mathrm{Bin}(n-1, 4r^2)$ and $\mathbb{E}[\deg(v)] = 4r^2(n-1)$.



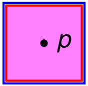
Locality in $G^{\mathbb{T}^2}(n, r)$

Observation

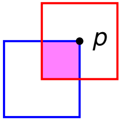
Let $p, q \sim \mathcal{U}([0,1]^2)$ and S_p the unit square around p.
Then $\Pr[q \in S_p] = \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} (1 - |x|)(1 - |y|) dx\, dy = \frac{9}{16}$.



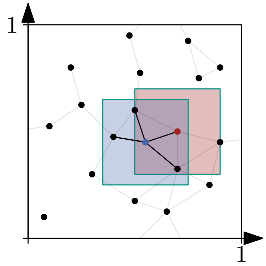
general case



best case



worst case



Corollary

By “rescaling” the observation we get $L = \Pr[\underbrace{\{u, w\} \in E}_{P_w \text{ in square around } P_u} \mid \underbrace{\{v, u\} \in E \wedge \{v, w\} \in E}_{P_u,\, P_w \text{ in square around } P_v}] = \frac{9}{16} = \Omega(1)$.

Poissonisation

Karlsruhe Institute of Technology

Poissonised Variant $G^{\mathbb{T}^2}_{\text{Pois}}(n, r)$ of $G^{\mathbb{T}^2}(n, r)$

Replace the point set with a Poisson point process on $\mathbb{T}^2$ with intensity n.

$\hookrightarrow$ i.e. region of size λ contains $\text{Pois}(\lambda n)$-many points, independent for disjoint regions

Equivalently: $G^{\mathbb{T}^2}_{\text{Pois}}(n, r) = G^{\mathbb{T}^2}(N, r)$ where $N \sim \text{Pois}(n)$.

Advanatages

- No long-distance correlations. ✓
- $\text{Pois}(4r^2)$-distributed degrees. ✓

Disadvantages

- Less natural in practice. ✗
- Number of vertices $N \sim \text{Pois}(n)$ not fixed. ✗

De-Poissonisation (an analogous result holds for de-Poissonising balls-into-bins)

Let P be a graph property. If P is very unlikely for $G^{\mathbb{T}^2}_{\text{Pois}}(n, r)$ then P is unlikely for $G^{\mathbb{T}^2}(n, r)$:

$$\Pr[G^{\mathbb{T}^2}(n, r) \in P] = \Pr[G^{\mathbb{T}^2}_{\text{Pois}}(n, r) \in P \mid N = n] \leq \frac{\Pr[G^{\mathbb{T}^2}_{\text{Pois}}(n,r) \in P]}{\Pr[N=n]} = \Theta(n^{1/2}) \Pr[G^{\mathbb{T}^2}_{\text{Pois}}(n, r) \in P].$$

Content

Scale-Free Networks

Karlsruhe Institute of Technology

Semi-Formal Definition

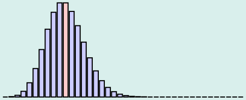
A scale-free network is a graph with a degree distribution that follows a power law (in an asymptotic sense)

Practical Consequence

There are vertices of very high degree (*hubs*).

Contrast: Erdős-Renyi

exponentially decreasing tail.

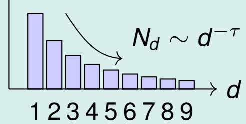


Pois(λ)

for $\lambda = 10$.

Power Laws

$N_d = \#\{v \in V \mid \deg(v) = d\}$



$\tau \leq 1$: not a distribution

$1 < \tau \leq 2$: distribution, but $\mathbb{E}[\deg(v)] = \infty$

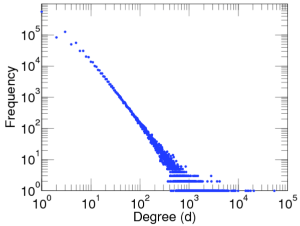
$2 < \tau \leq 3$: $\mathbb{E}[\deg(v)] < \infty$, but $\mathrm{Var}(\deg(v)) = \infty$

$3 < \tau \leq 4$: variance $< \infty$, but higher moments are ∞

$\tau \in (2, 3]$ is especially popular

"Youtube"

konect.cc/plot/degree.a.youtube-links.full.png



Scale-Free Networks

Karlsruhe Institute of Technology

Semi-Formal Definition

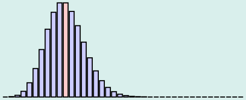
A scale-free network is a graph with a degree distribution that follows a power law (in an asymptotic sense)

Practical Consequence

There are vertices of very high degree (*hubs*).

Contrast: Erdős-Renyi

exponentially decreasing tail.



Pois(λ)

for $\lambda = 10$.

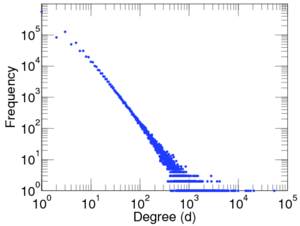
The Name "Scale-Free"

From *Barabási: "Linked: The New Science of Networks", 2002*.

In a random network [...] the vast majority of nodes have the same number of links [...]. Therefore, a random network has a characteristic scale in its node connectivity [...]. In contrast, the absence of a peak in a power-law degree distribution implies that [...] we see a continuous hierarchy of nodes, spanning from rare hubs to the numerous tiny nodes. There is no intrinsic scale in these networks. This is the reason my research group started to describe networks with power-law degree distribution as scale-free.

"Youtube"

konect.cc/plot/degree.a.youtube-links.full.png



A Scale-Free Random Geometric Graph

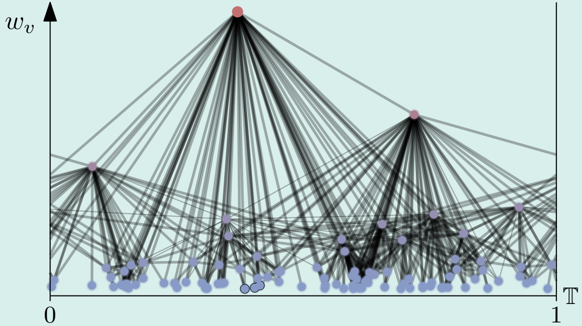
Karlsruhe Institute of Technology

Reminder: Random Geometric Graph (RGG)

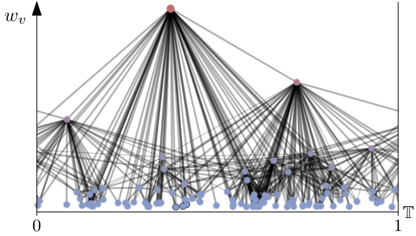
Distribute vertices in a metric space and connect any two vertices with a probability depending on their distance.

Definition: Geometric Inhomogeneous Random Graph (GIRG)

- number of vertices: n
- metric space $\mathbb{T}$ // more generally: $\mathbb{T}^d$ for $d \in \mathbb{N}$
- for each v: position $x_v \sim \mathcal{U}(\mathbb{T})$
- for each v: weight $w_v \sim \text{Par}(\tau - 1, 1)$
  the Pareto distribution is a power law distribution with exponent τ
- $\{u, v\} \in E \Leftrightarrow \text{dist}(x_u, x_v) \leq \frac{\lambda}{n} w_u w_v$
  $\Leftrightarrow \frac{n}{\lambda w_v} \leq \frac{w_u}{\text{dist}(x_u, x_v)}$.



How GIRGs are Useful



GIRGs are Scale-Free

$\mathbb{E}[\deg(v) \mid w_v] = \Theta(w_v)$ and $\deg(v)$ follows a power law if w_v does.

GIRGs are a Good Model for Real World Networks (Bläsius, Fischbeck, 2022)

- consider two graph parameters: locality and heterogeneity ($\approx \log \mathrm{Var}(\deg(v))$).
- in many contexts, a real network behaves like a GIRG with the same parameters

---

On the External Validity of Average-Case Analyses of Graph Algorithms, ESA 2022.

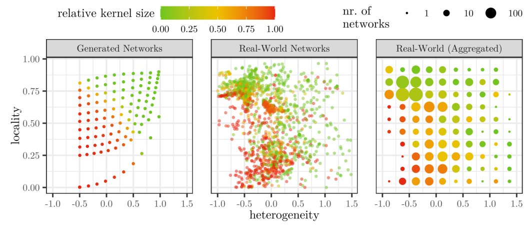


Figure 7 The relative kernel size of the vertex cover domination rule.

Hyperbolic Geometric Graphs

Poincaré Model of Hyperbolic Geometry

Illustration by M.C. Escher, Circle Limit III, 1959.

(All creatures are congruent in hyperbolic space.)

Result (Bläsius, Friedrich, Katzmann, 2021)

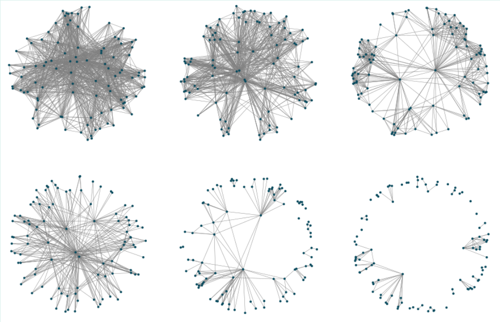
Vertex Cover can be Approximated on HGGs.

Efficiently Approximating Vertex Cover on Scale-Free Networks with Underlying Hyperbolic Geometry, ESA 2021.

Hyperbolic Random Graph (HGGs)

Sample points with bias towards the centre.
Connect points if distance is beneath a threshold.

https://commons.wikimedia.org/wiki/File:Hyperbolic_graphs.png



Can yield power law distribution for node degrees.

Preferential Attachment and the Barabasi-Albert Model

How a Graph is Grown Over Time

- There is a parameter $m \in \mathbb{N}$.
- start with any graph on $\geq m$ nodes.
- add new nodes one by one
  - new node is connected to m existing nodes
  - existing nodes are selected with probability proportional to their degree

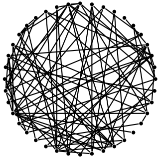
Why the Model is Interesting

- node degrees approach a power law distribution with exponent 3
- model may explain *why* scale-free networks emerge in practice

Summary

Erdős-Renyi Random Graphs

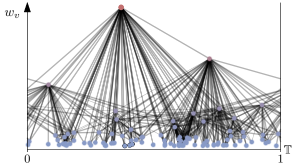
- simplest type of random graphs
- "Erdős-Renyi" refers to various related models
- arise in certain data structures (stay tuned)
- look locally like Poisson Galton-Watson Trees
- no locality or high-degree vertices



Random Graphs for Average Case Analsis

Mimic properties of real world networks:

- locality // a friend of my friend is often my friend
  - arises naturally in random geometric graphs
- "scale-freeness" $\approx$ existence of hubs
  - assign weights to vertices (in GIRGs)
  - use hyperbolic geometry
  - use preferential attachment



Anhang: Mögliche Prüfungsfragen I

- Was ist mit Theory-Practice Gap im Kontext von Graphalgorithmen gemeint?
- Wie kann die Theorie der Praxis entgegenkommen?
- Was ist das klassische Modell von Erdős und Renyi?
  - Welche Varianten des Erdős-Renyi Modells haben wir betrachtet?
  - Was gilt für die Verteilung von $\deg(v)$, wenn wir $\mathbb{E}[\deg(v)] = \lambda$ einstellen?
  - Was lässt sich über $N_d = |\{v \in [n] \mid \deg(v) = d\}|$ sagen?
  - Wir haben uns die R-Nachbarschaft $G(n, \lambda/n)|_{v,R}$ eines Knotens v angeschaut.
    - Was gilt für die Verteilung von $G(n, \lambda/n)|_{v,R}$ und warum?
    - Was ist ein Galton-Watson Baum?
    - Was lässt sich über die Aussterbewahrscheinlichkeit eines Poisson-Galton-Watson Baumes sagen?
  - Was versteht man unter „Sudden Emergence of the Giant Component“. Formuliere die Aussage formal.
  - Wir haben eine Größe L betrachtet, und Lokalität genannt. Wie ist sie definiert?
  - Welche Lokalität haben Erdős-Renyi Graphen?
- Nenne Eigenschaften, die Netzwerke in der Praxis von Erdős-Renyi Graphen unterscheiden.
  - Gib ein Beispiel für einen geometrischen Zufallsgraphen. Was ist die Lokalität in diesem Modell?

Anhang: Mögliche Prüfungsfragen II

- Inwiefern könnte ein Poissonisiertes Modell bequemer sein?
- Wann ist ein Netzwerk “Scale-Free”?
  - Gib ein Beispiel für ein Netzwerk aus der Praxis dem man diese Eigenschaft zuschreibt.
  - Beschreibe mindestens zwei Arten, wie man Netzwerke dieser Art generieren kann.