M2 Complex Systems - Complex Networks

Lecture 3 - Network models

Erdös-Rényi random graphs and configuration model

Automn 2021 - ENS Lyon

Christophe Crespelle
christophe.crespelle@ens-lyon.fr

* Thanks to Daron Acemoglu and Asu Ozdaglar for pedagogical material used for these slides.

Network models

Model = **random generation** of synthetic networks

- To simulate :
  - phenomena
  - algorithms
  - protocols
- In order to :
  - design
  - test
  - predict
  - better understand
- Example :

  Would Internet protocols still work if Internet was 10 times larger ?
  - generate a synthetic network and simulate

Network models

Model = **random generation** of synthetic networks

... having the properties of real-world networks ! ! !

Network models

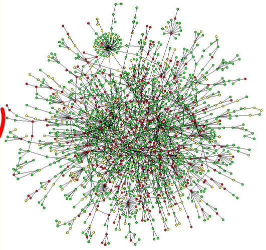
Model = **random generation** of synthetic networks
... having the properties of real-world networks ! ! !

Four classic properties of real-world complex networks :

- Low global density

$m = \#\text{edges} \ll \frac{n(n-1)}{2}$

$\frac{m}{\frac{n(n-1)}{2}} \ll 1$



10^{-6} 10^{-5}

Network models

Model = **random generation** of synthetic networks

... having the properties of real-world networks ! ! !

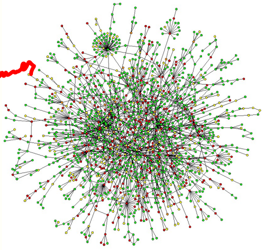
Four classic properties of real-world complex networks :

- Low global density
- Short distances

↦ randomness

average distance ~ cte

(diameter) ~ log n

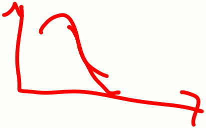


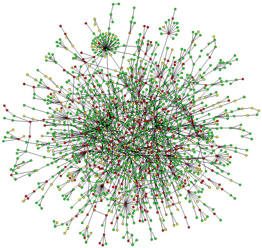
Network models

Model = **random generation** of synthetic networks
... having the properties of real-world networks ! ! !

Four classic properties of real-world complex networks :

- Low global density
- Short distances
- Heterogeneous degrees





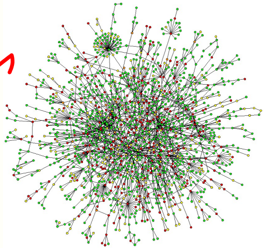
Network models

Model = **random generation** of synthetic networks

... having the properties of real-world networks ! ! !

Four classic properties of real-world complex networks :

- Low global density
- Short distances
- Heterogeneous degrees
- High local density



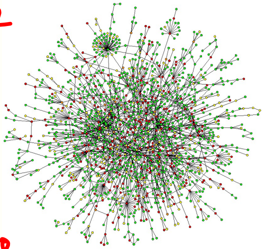
Network models

Model = **random generation** of synthetic networks
… having the properties of real-world networks ! ! !

Four classic properties of real-world complex networks :

- Low global density   0,000342
- Short distances   3,45
- Heterogeneous degrees
- High local density

Clust coef = 0,08



Goal : generate synthetic networks having these four properties (in a generic way)

Erdös-Rényi (ER) random graphs

There are two models :

- $G_{n,m}$: choose uniformly at random (u.a.r.) m edges among the n vertices

Erdös-Rényi (ER) random graphs

There are two models :

- $G_{n,m}$: choose uniformly at random (u.a.r.) m edges among the n vertices
- $G_{n,p}$: for each couple of the n vertices, put an edge with probability p

$$p \frac{n(n-1)}{2} = m$$

$$p = \frac{2m}{n(n-1)}$$

Erdös-Rényi (ER) random graphs

There are two models :

- $G_{n,m}$: choose uniformly at random (u.a.r.) m edges among the n vertices
- $G_{n,p}$: for each couple of the n vertices, put an edge with probability p

$\Rightarrow$ "essentialy" equivalent when $p = \frac{2m}{n(n-1)}$

Erdös-Rényi (ER) random graphs

1959 $G_{n,m}$

There are two models :

- $G_{n,m}$: choose uniformly at random (u.a.r.) m edges among the n vertices
- $G_{n,p}$: for each couple of the n vertices, put an edge with probability p

$\Rightarrow$ "essentialy" equivalent when $p = \frac{2m}{n(n-1)}$

Should we use $G_{n,m}$ or $G_{n,p}$? → $O(n^2)$

- For generating networks ? $G_{n,m}$ $O(m)$
- For mathematical analysis of the model ? $G_{n,p}$

$G_{n,m}$: implementation and complexity $O(m)$

- Algo : Pick m times two vertices uniformly at random

$0 \ldots n-1$ $\quad a = \text{rand}(0, n-1)$ $\quad (a, b)$

$b = \text{rand}(0, n-1)$

$(a\ b)$ risk of collision → multiple edges

$(a\ a)$ self loop

Pick until you get m distinct edges.

$O(m)$ time.

$G_{n,m}$: implementation and complexity

- Algo : Pick m times two vertices uniformly at random
  - How to deal with self-loops ?

$G_{n,m}$: implementation and complexity

- Algo : Pick m times two vertices uniformly at random
  - How to deal with self-loops ?
  - How to deal with multiple edges ?

Properties of $G_{n,p}$

Four properties to check :

- Low global density
  - p parameter of the model, controls m : $\mathbb{E}(m) = \frac{pn(n-1)}{2}$

Properties of $G_{n,p}$

Four properties to check :

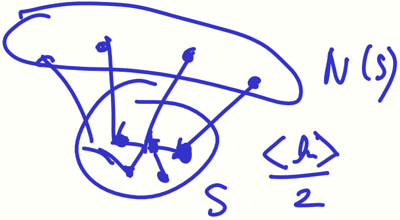
- Low global density ✓
  - p parameter of the model, controls m : $\mathbb{E}(m) = \frac{pn(n-1)}{2}$
  - law of large numbers : m is very concentrated around its mean

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances

Properties of $G_{n,p}$



Four properties to check :

- Low global density ✓
- Short distances

Expansion property

- def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances

  Expansion property

  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$?

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances
  Expansion property
  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$ $\frac{2m}{n} = \langle k \rangle$

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances ✓

  Expansion property

  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances ✓
  Expansion property
  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$
- Heterogeneous degrees

Properties of $G_{n,p}$

Four properties to check :

p(n)

- Low global density ✓
- Short distances ✓

  Expansion property

  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$
- Heterogeneous degrees
  - fix the average degree $\lambda = p(n-1)$

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances ✓

  Expansion property
  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$
- Heterogeneous degrees
  - fix the average degree $\lambda = p(n-1)$
  - $$\begin{aligned}\mathbb{P}(d^\circ = k) &= \binom{n-1}{k} p^k (1-p)^{(n-1-k)} \\ &= \frac{A_n^k}{k!} \frac{\lambda^k}{(n-1)^k} \left(1 - \frac{\lambda}{n-1}\right)^{n-1-k} \\ &= \frac{A_n^k}{(n-1)^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n-1}\right)^{n-1} \left(1 - \frac{\lambda}{n-1}\right)^{-k}\end{aligned}$$

k: fix
λ: fix
$n \to +\infty$

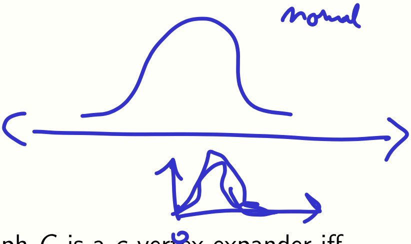
$\downarrow 1$ $\quad \downarrow e^{-\lambda}$ $\quad \downarrow 1$

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances ✓

  Expansion property
  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$



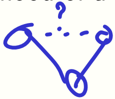
- Heterogeneous degrees ✗
  - fix the average degree $\lambda = p(n-1)$
  - $$\begin{aligned} \mathbb{P}(d^\circ = k) &= \binom{n-1}{k} p^k (1-p)^{(n-1-k)} \\ &= \frac{A_n^k}{k!} \frac{\lambda^k}{(n-1)^k} \left(1 - \frac{\lambda}{n-1}\right)^{n-1-k} \\ &= \frac{A_n^k}{(n-1)^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n-1}\right)^{n-1} \left(1 - \frac{\lambda}{n-1}\right)^{-k} \end{aligned}$$
  - then when $n \to +\infty$, $\mathbb{P}(d^\circ = k) \to \frac{\lambda^k e^{-\lambda}}{k!}$: Poisson law

Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances ✓

  Expansion property
  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$
- Heterogeneous degrees ✘
- High local density
  - probability of an edge in the neighbourhood of a vertex ?



Properties of $G_{n,p}$

Four properties to check :

- Low global density ✓
- Short distances ✓
  Expansion property
  - def. (graph theory) : a graph G is a c-vertex-expander iff $\forall S \subseteq V$ s.t. $|S| \leq \frac{|V(G)|}{2}$, we have $|N(S)| \geq c \cdot |S|$
  - expansion of $G_{n,m}$? $\sim \frac{m}{n}$
  - until $\frac{n}{2}$, exponential growth of $|B(u,d)| \sim (1+c)^d$
- Heterogeneous degrees ✗
- High local density ✗
  - probability of an edge in the neighbourhood of a vertex ?
  - same as everywhere : p (couples of vertices are independant)

Phase transitions in $G_{n,p}$

N.B. : p (eventually) depends on n

- Threshold function $t(n)$ for property A :
  - $\mathbb{P}(A) \to 0$ if $\frac{p(n)}{t(n)} \to 0$
  - $\mathbb{P}(A) \to 1$ if $\frac{p(n)}{t(n)} \to +\infty$
  - makes sense for monotonic properties (for inclusion of edge set)

$p(n) = o(t(n))$

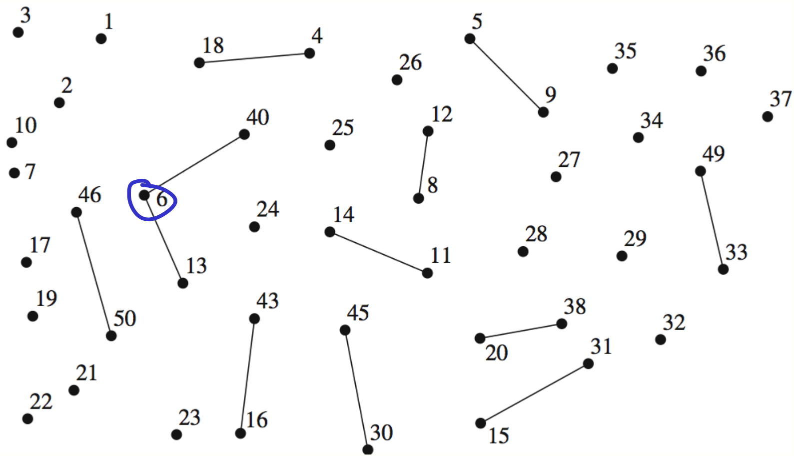
$t(n) = o(p(n))$

- such a threshold function exists $\Rightarrow$ **phase transition**

- Seminal work of Erdös and Rényi in 1959

Phase transitions in $G_{n,p}$

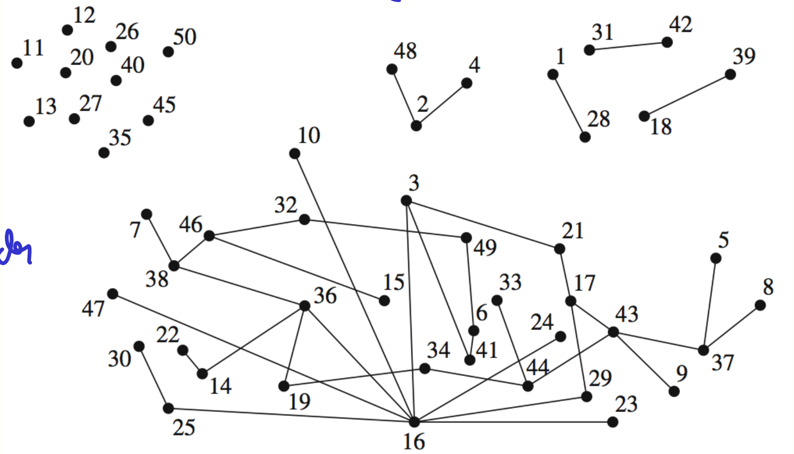
$n = 50,\ p = 0.01$



Phase transitions in $G_{n,p}$

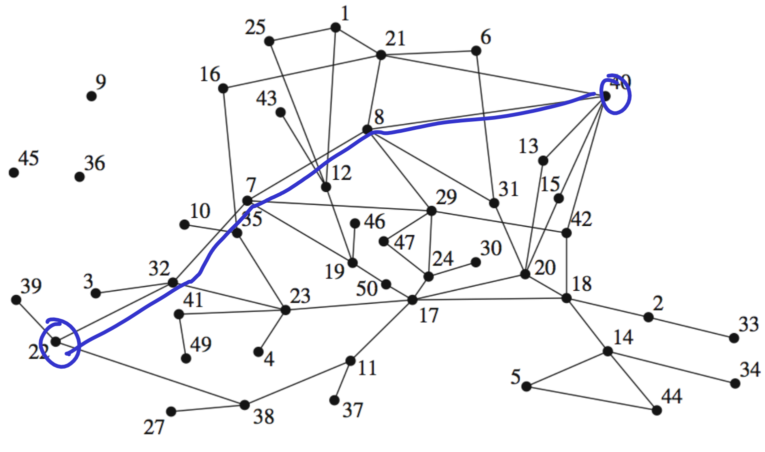
$n = 50,\ p = 0.03$

cycles



Phase transitions in $G_{n,p}$

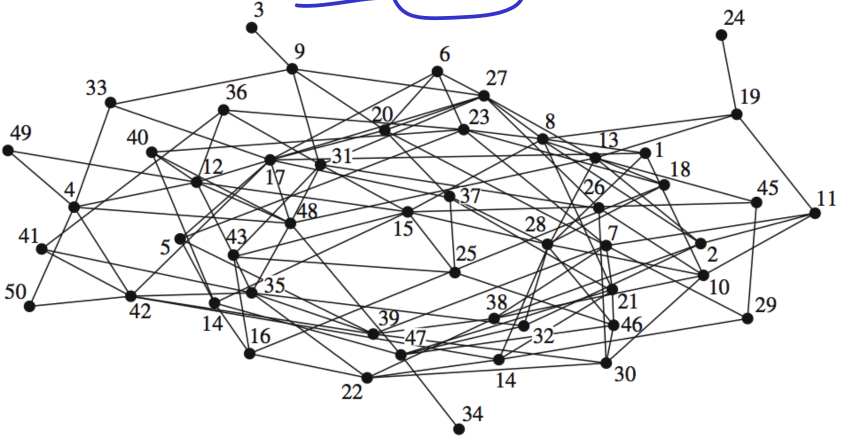
$n = 50,\ p = 0.05$



giant compont..

Phase transitions in $G_{n,p}$

$n = 50, p = 0.10$



Threshold for connectivity

- We show a threshold with function $t(n) = \frac{\log n}{n}$
- Denote $p(n) = \lambda \frac{\log n}{n}$ (mean degree $\sim \lambda \log n$)
- We show a (much) stronger statement for threshold function $\frac{\log n}{n}$:
  1. $\mathbb{P}(connectivity) \to 0$ if $\lambda < 1$
  2. $\mathbb{P}(connectivity) \to 1$ if $\lambda > 1$

Proof of (1)

$\lambda < 1$ $\quad p(n) = \lambda \frac{\log n}{n}$

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1

0

Proof of (1)

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1

- Let I_i be the Bernouilli random variable defined as
  - $I_i = 1$ if vertex i is isolated
  - $I_i = 0$ otherwise

Proof of (1)

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1

- Let I_i be the Bernouilli random variable defined as
  - $I_i = 1$ if vertex i is isolated
  - $I_i = 0$ otherwise

- probability that a vertex is isolated :
  $q = \mathbb{P}(I_i = 1) = (1 - p)^{n-1} \sim e^{-\lambda \log n} = n^{-\lambda}$

Proof of (1)

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1

- Let I_i be the Bernouilli random variable defined as
  - $I_i = 1$ if vertex i is isolated
  - $I_i = 0$ otherwise

- probability that a vertex is isolated :
  $q = \mathbb{P}(I_i = 1) = (1-p)^{n-1} \sim e^{-\lambda \log n} = n^{-\lambda}$
- Let $X = \sum_{i=1}^{n} I_i$ be the number of isolated vertices

Proof of (1)

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1

- Let I_i be the Bernouilli random variable defined as
  - $I_i = 1$ if vertex i is isolated
  - $I_i = 0$ otherwise

- probability that a vertex is isolated :
  $q = \mathbb{P}(I_i = 1) = (1-p)^{n-1} \sim e^{-\lambda \log n} = n^{-\lambda}$
- Let $X = \sum_{i=1}^{n} I_i$ be the number of isolated vertices
- We have $\mathbb{E}[X] = n.n^{-\lambda} \to +\infty$ for $\lambda < 1$

$\frac{n}{n^{\lambda}}$

Proof of (1)

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1
- Let I_i be the Bernouilli random variable defined as
  - $I_i = 1$ if vertex i is isolated
  - $I_i = 0$ otherwise
- probability that a vertex is isolated :
  $q = \mathbb{P}(I_i = 1) = (1-p)^{n-1} \sim e^{-\lambda \log n} = n^{-\lambda}$
- Let $X = \sum_{i=1}^{n} I_i$ be the number of isolated vertices
- We have $\mathbb{E}[X] = n.n^{-\lambda} \to +\infty$ for $\lambda < 1$
- enough to conclude that $\mathbb{P}(X = 0) \to 0$?

Proof of (1)

- to prove (1), it is enough to show that the probability of existence of at least one isolated node goes to 1
- Let I_i be the Bernouilli random variable defined as
  - $I_i = 1$ if vertex i is isolated
  - $I_i = 0$ otherwise
- probability that a vertex is isolated :
  $q = \mathbb{P}(I_i = 1) = (1-p)^{n-1} \sim e^{-\lambda \log n} = n^{-\lambda}$
- Let $X = \sum_{i=1}^{n} I_i$ be the number of isolated vertices
- We have $\mathbb{E}[X] = n.n^{-\lambda} \to +\infty$ for $\lambda < 1$
- enough to conclude that $\mathbb{P}(X = 0) \to 0$?

$\Rightarrow$ NO, we need a concentration property.

Proof of (1)

$X = \sum I_i$

- We have

$$\begin{aligned} var(X) &= \sum_i var(I_i) + \sum_i \sum_{j \neq i} cov(I_i, I_j) \\ &= nvar(I_1) + n(n-1)cov(I_1, I_2) \end{aligned}$$

Proof of (1)

- We have

$$\begin{aligned} var(X) &= \sum_i var(I_i) + \sum_i \sum_{j \neq i} cov(I_i, I_j) \\ &= n var(I_1) + n(n-1) cov(I_1, I_2) \end{aligned}$$

- And we also have :
  - $var(I_1) = \mathbb{E}[I_1^2] - \mathbb{E}[I_1]^2 = q - q^2$
  - $cov(I_1, I_2) = \mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2]$
  - $\mathbb{E}[I_1 I_2] = \mathbb{P}(I_1 = 1, I_2 = 1) = (1-p)^{2n-3} = \frac{q^2}{1-p}$

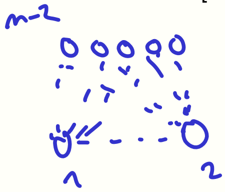
$E(X) = \frac{n}{2}$

$E(X^2) \sim n^3$

$\frac{n^2}{4}$

$q = (1-p)^{n-1}$

$i^k \quad n^{k+1}$



$$\sum_{i=1}^{n} i^2 \sim n^3$$

$$\sum_{i=1}^{n} i \sim n^2$$

$2(n-2) + 1$

Proof of (1)

- We have
$$\begin{aligned} var(X) &= \sum_i var(I_i) + \sum_i \sum_{j \neq i} cov(I_i, I_j) \\ &= nvar(I_1) + n(n-1)cov(I_1, I_2) \end{aligned}$$
- And we also have :
  - $var(I_1) = \mathbb{E}[I_1^2] - \mathbb{E}[I_1]^2 = q - q^2$
  - $cov(I_1, I_2) = \mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2]$
  - $\mathbb{E}[I_1 I_2] = \mathbb{P}(I_1 = 1, I_2 = 1) = (1-p)^{2n-3} = \frac{q^2}{1-p}$
- We then obtain $var(X) = nq(1-q) + n(n-1)\frac{q^2 p}{1-p}$

Proof of (1)

$p = \lambda \frac{\log n}{n}$

- We have

$$\begin{aligned} var(X) &= \sum_i var(I_i) + \sum_i \sum_{j \neq i} cov(I_i, I_j) \\ &= nvar(I_1) + n(n-1)cov(I_1, I_2) \end{aligned}$$

- And we also have :
  - $var(I_1) = \mathbb{E}[I_1^2] - \mathbb{E}[I_1]^2 = q - q^2$
  - $cov(I_1, I_2) = \mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2]$
  - $\mathbb{E}[I_1 I_2] = \mathbb{P}(I_1 = 1, I_2 = 1) = (1-p)^{2n-3} = \frac{q^2}{1-p}$

$q = \left(1 - \lambda \frac{\log n}{n}\right)^{n-1} \leq e^{-\lambda \log n}$

- We then obtain $var(X) = nq(1-q) + n(n-1)\frac{q^2 p}{1-p}$
- when $n \to +\infty$, then $q \to 0$ and $p \to 0$

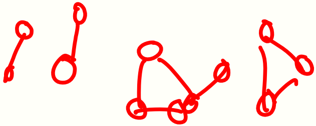
Proof of (1)

- We have
$$\begin{aligned} var(X) &= \sum_i var(I_i) + \sum_i \sum_{j \neq i} cov(I_i, I_j) \\ &= n var(I_1) + n(n-1) cov(I_1, I_2) \end{aligned}$$
- And we also have :
  - $var(I_1) = \mathbb{E}[I_1^2] - \mathbb{E}[I_1]^2 = q - q^2$
  - $cov(I_1, I_2) = \mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2]$
  - $\mathbb{E}[I_1 I_2] = \mathbb{P}(I_1 = 1, I_2 = 1) = (1-p)^{2n-3} = \frac{q^2}{1-p}$
- We then obtain $var(X) = nq(1-q) + n(n-1)\frac{q^2 p}{1-p}$
- when $n \to +\infty$, then $q \to 0$ and $p \to 0$
- this gives
$$\begin{aligned} var(X) &\sim nq + n^2 q^2 p \\ &= n n^{-\lambda} + \lambda n \log n n^{-2\lambda} \\ &\sim n n^{-\lambda} = \mathbb{E}[X] \end{aligned}$$

Proof of (1)

- so we have $var(X) \sim \mathbb{E}[X]$
- and because $var(X) \geq (0 - \mathbb{E}[X])^2 \mathbb{P}(X = 0)$
- we obtain $\mathbb{P}(X = 0) \leq \frac{1}{\mathbb{E}[X]} \to 0$
- it follows that $\mathbb{P}(X > 0) \to 1$ when $n \to +\infty$
- and consequently $\mathbb{P}(disconnected) \to 1$ when $n \to +\infty$

Proof of (2)



- we now fix $\lambda > 1$
- let's check that $\mathbb{E}[X] = nn^{-\lambda} \to 0$ when $n \to +\infty$ ✓

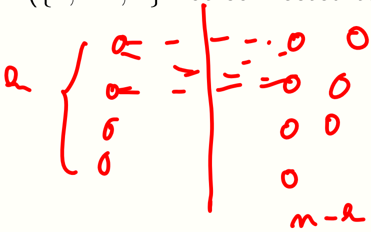
Proof of (2)

- we now fix $\lambda > 1$
- let's check that $\mathbb{E}[X] = nn^{-\lambda} \to 0$ when $n \to +\infty$ ✓

- observe that G is disconnected $\iff \exists k$ vertices without edges to the other vertices, for some $k \leq n/2$

Proof of (2)

- we now fix $\lambda > 1$
- let's check that $\mathbb{E}[X] = nn^{-\lambda} \to 0$ when $n \to +\infty$ ✓
- observe that G is disconnected $\iff \exists k$ vertices without edges to the other vertices, for some $k \le n/2$
- we have
  $\mathbb{P}(\{1, \ldots, k\}$ not connected to the rest$) = (1-p)^{k(n-k)}$



Proof of (2)

- we now fix $\lambda > 1$
- let's check that $\mathbb{E}[X] = nn^{-\lambda} \to 0$ when $n \to +\infty$ ✓

- observe that G is disconnected $\iff \exists k$ vertices without edges to the other vertices, for some $k \leq n/2$
- we have
  $\mathbb{P}(\{1, \ldots, k\}$ not connected to the rest$) = (1-p)^{k(n-k)}$
- and so
  $\mathbb{P}(\exists k$ vertices not connected to the rest$) \leq \binom{n}{k}(1-p)^{k(n-k)}$

Proof of (2)

- we now fix $\lambda > 1$
- let's check that $\mathbb{E}[X] = nn^{-\lambda} \to 0$ when $n \to +\infty$ ✓

- observe that G is disconnected $\Longleftrightarrow$ $\exists k$ vertices without edges to the other vertices, for some $k \leq n/2$
- we have
  $\mathbb{P}(\{1, \ldots, k\}$ not connected to the rest$) = (1-p)^{k(n-k)}$
- and so
  $\mathbb{P}(\exists k$ vertices not connected to the rest$) \leq \binom{n}{k}(1-p)^{k(n-k)}$
- and finally $\mathbb{P}(G$ is disconnected$) \leq \sum_{k=1}^{n/2} \binom{n}{k}(1-p)^{k(n-k)}$

Proof of (2)

- we now fix $\lambda > 1$
- let's check that $\mathbb{E}[X] = nn^{-\lambda} \to 0$ when $n \to +\infty$ ✓

- observe that G is disconnected $\iff \exists k$ vertices without edges to the other vertices, for some $k \leq n/2$
- we have $\mathbb{P}(\{1, \ldots, k\} \text{ not connected to the rest}) = (1-p)^{k(n-k)}$
- and so $\mathbb{P}(\exists k \text{ vertices not connected to the rest}) \leq \binom{n}{k}(1-p)^{k(n-k)}$
- and finally $\mathbb{P}(G \text{ is disconnected}) \leq \sum_{k=1}^{n/2} \binom{n}{k}(1-p)^{k(n-k)}$
- using this expression, one can show that $\mathbb{P}(G \text{ is disconnected}) \to 0$ when $n \to +\infty$

Threshold for giant component

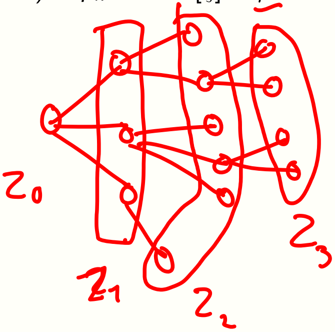
- Giant = constant fraction of the vertices
- We show a threshold with function $t(n) = \frac{1}{n}$
- Denote $p(n) = \frac{\lambda}{n}$ (mean degree $\sim \lambda$)
- We again show a strong statement for threshold function $\frac{1}{n}$:
  1. if $\lambda < 1$, $\forall a \in \mathbb{R}_+^*, \mathbb{P}(maxsize(CC) \geq a \log n) \to 0$
  2. if $\lambda > 1$, $\exists b \in \mathbb{R}_+^*, \mathbb{P}(maxsize(CC) \geq b.n) \to 1$

Proof of (1) – preliminaries

- Galton-Watson branching process
  - start with a single individual

Proof of (1) – preliminaries

- Galton-Watson branching process
  - start with a single individual
  - each individual generates a number of children according to a non-negative random variable ξ with distribution p_k
    $\mathbb{P}(\xi = k) = p_k \qquad \mathbb{E}[\xi] = \mu \qquad var(\xi) \neq 0$



Proof of (1) – preliminaries

- Galton-Watson branching process
  - start with a single individual
  - each individual generates a number of children according to a non-negative random variable ξ with distribution p_k
    $\mathbb{P}(\xi = k) = p_k \qquad \mathbb{E}[\xi] = \mu \qquad var(\xi) \neq 0$

- Let Z_k be the number of individuals in the k^{th} generation
  we have $Z_0 = 1,\ Z_1 = \xi,\ Z_2 = \sum_{i=1}^{Z_1} \xi^{(i)}$

Proof of (1) – preliminaries

- Galton-Watson branching process
  - start with a single individual
  - each individual generates a number of children according to a non-negative random variable ξ with distribution p_k
    $\mathbb{P}(\xi = k) = p_k \qquad \mathbb{E}[\xi] = \mu \qquad var(\xi) \neq 0$
- Let Z_k be the number of individuals in the k^{th} generation
  we have $Z_0 = 1$, $Z_1 = \xi$, $Z_2 = \sum_{i=1}^{Z_1} \xi^{(i)}$
- and consequently
  - $\mathbb{E}[Z_1] = \mu$

Proof of (1) – preliminaries

- Galton-Watson branching process
  - start with a single individual
  - each individual generates a number of children according to a non-negative random variable ξ with distribution p_k
    $\mathbb{P}(\xi = k) = p_k \qquad \mathbb{E}[\xi] = \mu \qquad var(\xi) \neq 0$

- Let Z_k be the number of individuals in the k^{th} generation
  we have $Z_0 = 1,\ Z_1 = \xi,\ Z_2 = \sum_{i=1}^{Z_1} \xi^{(i)}$

- and consequently
  - $\mathbb{E}[Z_1] = \mu$
  - $\mathbb{E}[Z_2] = \mathbb{E}[\mathbb{E}[Z_2|Z_1]] = \mathbb{E}[\mu Z_1] = \mu^2$

Proof of (1) – preliminaries

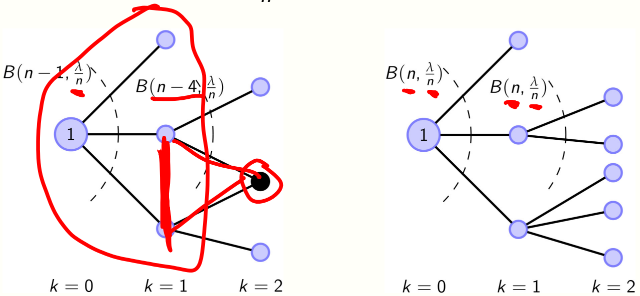
- Galton-Watson branching process
  - start with a single individual
  - each individual generates a number of children according to a non-negative random variable ξ with distribution p_k
    $\mathbb{P}(\xi = k) = p_k \qquad \mathbb{E}[\xi] = \mu \qquad var(\xi) \neq 0$

- Let Z_k be the number of individuals in the k^{th} generation
  we have $Z_0 = 1,\ Z_1 = \xi,\ Z_2 = \sum_{i=1}^{Z_1} \xi^{(i)}$

- and consequently
  - $\mathbb{E}[Z_1] = \mu$
  - $\mathbb{E}[Z_2] = \mathbb{E}[\mathbb{E}[Z_2|Z_1]] = \mathbb{E}[\mu Z_1] = \mu^2$
  - and by recursion, for $k \geq 1$, we obtain
    $\mathbb{E}[Z_k] = \mathbb{E}[\mathbb{E}[Z_k|Z_{k-1}]] = \mathbb{E}[\mu Z_{k-1}] = \mu.\mu^{k-1} = \mu^k$

"Proof" of (1)

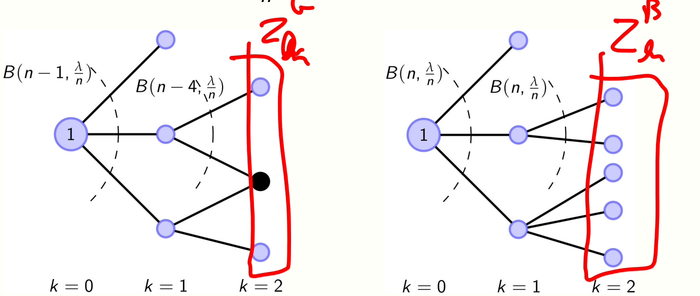
- Let $B(n, \frac{\lambda}{n})$ denote the binomial random variable with n trials and success probability $\frac{\lambda}{n}$



(a) ER graph process (b) branching process approx.

"Proof" of (1)

- Let $B(n, \frac{\lambda}{n})$ denote the binomial random variable with n trials and success probability $\frac{\lambda}{n}$

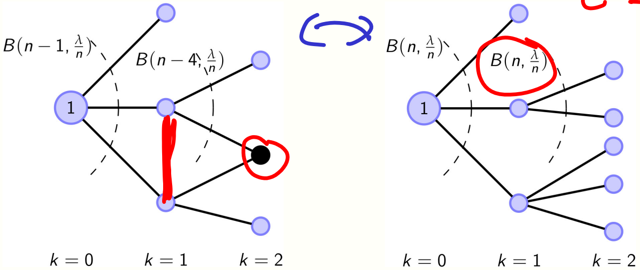


(a) ER graph process (b) branching process approx.

- Z_k^G and Z_k^B the number of individuals in generation k for the graph process and the branching process approximation

"Proof" of (1)

- Let $B(n, \frac{\lambda}{n})$ denote the binomial random variable with n trials and success probability $\frac{\lambda}{n}$



(a) ER graph process (b) branching process approx.

- Z_k^G and Z_k^B the number of individuals in generation k for the graph process and the branching process approximation
- we have $Z_k^G \leq Z_k^B$, for all k

"Proof" of (1)

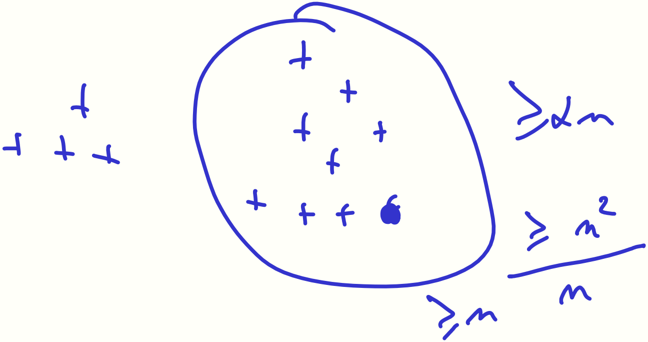
- fix $\lambda < 1$

"Proof" of (1)

- fix $\lambda < 1$
- Let S_i be the number of nodes in the connected component of vertex i

"Proof" of (1)

- fix $\lambda < 1$
- Let S_i be the number of nodes in the connected component of vertex i
- we have $\mathbb{E}[S_i] = \sum_k \mathbb{E}[Z_k^G] \leq \sum_k \mathbb{E}[Z_k^B] = \sum_k \lambda^k = \frac{1}{1-\lambda}$



"Proof" of (1)

- fix $\lambda < 1$
- Let S_i be the number of nodes in the connected component of vertex i
- we have $\mathbb{E}[S_i] = \sum_k \mathbb{E}[Z_k^G] \leq \sum_k \mathbb{E}[Z_k^B] = \sum_k \lambda^k = \frac{1}{1-\lambda}$
- so if $\lambda < 1$, the expected size of the components of vertex i is constant $\implies$ no giant component

"Proof" of (1)

$\lambda < 1 \longleftrightarrow \lambda > 1$, $\frac{\lambda}{n}$

- fix $\lambda < 1$
- Let S_i be the number of nodes in the connected component of vertex i
- we have $\mathbb{E}[S_i] = \sum_k \mathbb{E}[Z_k^G] \leq \sum_k \mathbb{E}[Z_k^B] = \sum_k \lambda^k = \frac{1}{1-\lambda}$
- so if $\lambda < 1$, the expected size of the components of vertex i is constant $\Longrightarrow$ no giant component
- one can show (not shown here) that the size of the bigger component does not exceed $\log n$:

$$\forall a > 0, \mathbb{P}(max_{1 \leq i \leq n} |S_i| \geq a \log n) \to 0 \text{ as } n \to +\infty$$

Proof of (2)

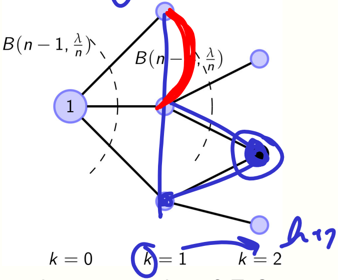
- fix $\lambda > 1$

Proof of (2)

- fix $\lambda > 1$
- We want to compute $\mathbb{E}[S_i]$ and show that it is large
  $\Longrightarrow$ we can no longer ignore conflicts

Proof of (2)

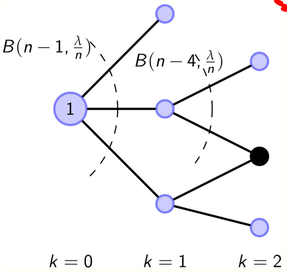
- fix $\lambda > 1$
- We want to compute $\mathbb{E}[S_i]$ and show that it is large $\implies$ we can no longer ignore conflicts
- We claim that $Z_k^G \approx Z_k^B$ as long as $\lambda^k \leq cte.\sqrt{n}$
  - $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] = n\frac{\lambda^2}{n^2}\mathbb{E}[Z_k^2]$



  - what about edges between nodes of Z_k ?

Proof of (2)

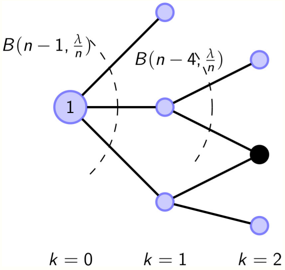
- fix $\lambda > 1$
- We want to compute $\mathbb{E}[S_i]$ and show that it is large
  $\Longrightarrow$ we can no longer ignore conflicts
- We claim that $Z_k^G \approx Z_k^B$ as long as $\lambda^k \leq cte.\sqrt{n}$
  - $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] = n\frac{\lambda^2}{n^2}\mathbb{E}[Z_k^2]$



    - what about edges between nodes of Z_k ?
  - we assume that as long as conflicts are negligible, Z_k is a Poisson variable, that is $var(Z_k) = \lambda^k$

Proof of (2)

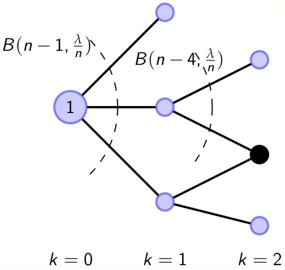
- fix $\lambda > 1$
- We want to compute $\mathbb{E}[S_i]$ and show that it is large
  $\Longrightarrow$ we can no longer ignore conflicts
- We claim that $Z_k^G \approx Z_k^B$ as long as $\lambda^k \leq cte.\sqrt{n}$
  - $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] = n\frac{\lambda^2}{n^2}\mathbb{E}[Z_k^2]$



    - what about edges between nodes of Z_k ?
  - we assume that as long as conflicts are negligible, Z_k is a Poisson variable, that is $var(Z_k) = \lambda^k$
  - we obtain $\mathbb{E}[Z_k^2] = var(Z_k) + \mathbb{E}[Z_k]^2 = \lambda^k + \lambda^{2k} \sim \lambda^{2k}$

Proof of (2)

- fix $\lambda > 1$
- We want to compute $\mathbb{E}[S_i]$ and show that it is large
  $\Longrightarrow$ we can no longer ignore conflicts
- We claim that $Z_k^G \approx Z_k^B$ as long as $\lambda^k \leq cte.\sqrt{n}$
  - $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] = n\frac{\lambda^2}{n^2}\mathbb{E}[Z_k^2]$



    - what about edges between nodes of Z_k ?
  - we assume that as long as conflicts are negligible, Z_k is a Poisson variable, that is $var(Z_k) = \lambda^k$
  - we obtain $\mathbb{E}[Z_k^2] = var(Z_k) + \mathbb{E}[Z_k]^2 = \lambda^k + \lambda^{2k} \sim \lambda^{2k}$

$\Longrightarrow$ $\mathbb{E}[\#\text{conflicts}]$ becomes $\Omega(1)$ only when $\lambda^k \approx \sqrt{n}$

Proof of (2)

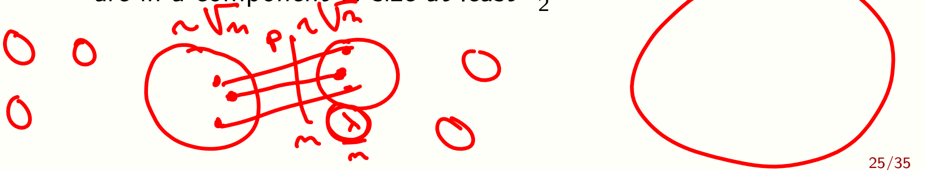
- $\mathbb{E}[S_i] = \sum_k \mathbb{E}[Z_k^G] \geq \sum_{k \leq \log_\lambda(\sqrt{n})} \mathbb{E}[Z_k^B] = \sum_{k \leq \log_\lambda(\sqrt{n})} \lambda^k$
  $\geq \frac{1-\lambda^{\log_\lambda(\sqrt{n})}}{1-\lambda} \geq \sqrt{n}$

Proof of (2)

- $\mathbb{E}[S_i] = \sum_k \mathbb{E}[Z_k^G] \geq \sum_{k \leq \log_\lambda(\sqrt{n})} \mathbb{E}[Z_k^B] = \sum_{k \leq \log_\lambda(\sqrt{n})} \lambda^k$
  $\geq \frac{1-\lambda^{\log_\lambda(\sqrt{n})}}{1-\lambda} \geq \sqrt{n}$
- Let us assume again that $|Z_k|$ follows a Poisson law of parameter λ^k
  - we then have $\mathbb{P}(||Z_k| - \lambda^k| \geq x) \leq 2e^{-\frac{x^2}{2(\lambda^k+x)}}$
  - which gives for $x = \sqrt{\lambda^k}$, $\mathbb{P}(||Z_k| - \lambda^k| \geq \sqrt{\lambda^k}) \leq 2e^{-\frac{1}{3}}$

Proof of (2)

- $\mathbb{E}[S_i] = \sum_k \mathbb{E}[Z_k^G] \geq \sum_{k \leq \log_\lambda(\sqrt{n})} \mathbb{E}[Z_k^B] = \sum_{k \leq \log_\lambda(\sqrt{n})} \lambda^k$
  $\geq \frac{1-\lambda^{\log_\lambda(\sqrt{n})}}{1-\lambda} \geq \sqrt{n}$
- Let us assume again that $|Z_k|$ follows a Poisson law of parameter λ^k
  - we then have $\mathbb{P}(||Z_k| - \lambda^k| \geq x) \leq 2e^{-\frac{x^2}{2(\lambda^k + x)}}$
  - which gives for $x = \sqrt{\lambda^k}$, $\mathbb{P}(||Z_k| - \lambda^k| \geq \sqrt{\lambda^k}) \leq 2e^{-\frac{1}{3}}$
- for large n, we obtain $\mathbb{P}(|S_i| \geq \frac{\sqrt{n}}{2}) \geq cte$
  $\Longrightarrow$ there is a constant fraction of the nodes (say $\alpha.n$) that are in a component of size at least $\frac{\sqrt{n}}{2}$



Proof of (2)

- Assume there is more than one component of size $\frac{\sqrt{n}}{2}$, we will show that the probability this happens $\to 0$ when $n \to +\infty$

Proof of (2)

- Assume there is more than one component of size $\frac{\sqrt{n}}{2}$, we will show that the probability this happens $\to 0$ when $n \to +\infty$
- Let C_1 be the smallest of these components and let A be the union of all these components

Proof of (2)

- Assume there is more than one component of size $\frac{\sqrt{n}}{2}$, we will show that the probability this happens $\to 0$ when $n \to +\infty$
- Let C_1 be the smallest of these components and let A be the union of all these components
  - we denote $|C_1| = k \geq \frac{\sqrt{n}}{2}$

Proof of (2)

- Assume there is more than one component of size $\frac{\sqrt{n}}{2}$, we will show that the probability this happens $\to 0$ when $n \to +\infty$
- Let C_1 be the smallest of these components and let A be the union of all these components
  - we denote $|C_1| = k \geq \frac{\sqrt{n}}{2}$
  - $\mathbb{P}(C_1 \text{ not connected to } A \setminus C_1) = (1-p)^{k(|A|-k)} \leq (1-\frac{\lambda}{n})^{\frac{\alpha n \sqrt{n}}{4}} \leq e^{-\frac{\lambda}{n} \cdot \frac{\alpha n \sqrt{n}}{4}} = e^{-\frac{\lambda \alpha \sqrt{n}}{4}} \to 0$ when $n \to +\infty$

Proof of (2)

- Assume there is more than one component of size $\frac{\sqrt{n}}{2}$, we will show that the probability this happens $\to 0$ when $n \to +\infty$
- Let C_1 be the smallest of these components and let A be the union of all these components
  - we denote $|C_1| = k \geq \frac{\sqrt{n}}{2}$
  - $\mathbb{P}(C_1 \text{ not connected to } A \setminus C_1) = (1-p)^{k(|A|-k)} \leq (1 - \frac{\lambda}{n})^{\frac{\alpha n \sqrt{n}}{4}} \leq e^{-\frac{\lambda}{n} \cdot \frac{\alpha n \sqrt{n}}{4}} = e^{-\frac{\lambda \alpha \sqrt{n}}{4}} \to 0$ when $n \to +\infty$
- this means that the probability that the vertices of A are grouped in a single connected component $\to 1$ when $n \to +\infty$

Proof of (2)

- Assume there is more than one component of size $\frac{\sqrt{n}}{2}$, we will show that the probability this happens $\to 0$ when $n \to +\infty$
- Let C_1 be the smallest of these components and let A be the union of all these components
  - we denote $|C_1| = k \geq \frac{\sqrt{n}}{2}$
  - $\mathbb{P}(C_1 \text{ not connected to } A \setminus C_1) = (1-p)^{k(|A|-k)} \leq (1-\frac{\lambda}{n})^{\frac{\alpha n \sqrt{n}}{4}} \leq e^{-\frac{\lambda}{n}\cdot\frac{\alpha n \sqrt{n}}{4}} = e^{-\frac{\lambda \alpha \sqrt{n}}{4}} \to 0$ when $n \to +\infty$
- this means that the probability that the vertices of A are grouped in a single connected component $\to 1$ when $n \to +\infty$
- since $|A| \geq \alpha.n$, this constitutes a giant component

Size of the giant component

- Let $G = G_{n-1,p}$ pe an ER graph with $p(n) = \frac{\lambda}{n}$ with $\lambda > 1$

Size of the giant component

- Let $G = G_{n-1,p}$ pe an ER graph with $p(n) = \frac{\lambda}{n}$ with $\lambda > 1$
- Add a n^{th} vertex to G and connect it to the rest of the vertices with probability $p(n)$ and denote G' the resulting graph

Size of the giant component

- Let $G = G_{n-1,p}$ pe an ER graph with $p(n) = \frac{\lambda}{n}$ with $\lambda > 1$
- Add a n^{th} vertex to G and connect it to the rest of the vertices with probability $p(n)$ and denote G' the resulting graph
- We denote ρ the fraction of vertices that are not in the giant component and we assume that, for large n, ρ is the same in G and G'

Size of the giant component

- Let $G = G_{n-1,p}$ pe an ER graph with $p(n) = \frac{\lambda}{n}$ with $\lambda > 1$
- Add a n^{th} vertex to G and connect it to the rest of the vertices with probability $p(n)$ and denote G' the resulting graph
- We denote ρ the fraction of vertices that are not in the giant component and we assume that, for large n, ρ is the same in G and G'
- vertex n is not in the giant component iff none of its neighbours are
  - This gives $\rho = \sum_{d \geq 0} P_d \rho^d = \Phi(\rho)$

Size of the giant component

- Let $G = G_{n-1,p}$ pe an ER graph with $p(n) = \frac{\lambda}{n}$ with $\lambda > 1$
- Add a n^{th} vertex to G and connect it to the rest of the vertices with probability $p(n)$ and denote G' the resulting graph
- We denote ρ the fraction of vertices that are not in the giant component and we assume that, for large n, ρ is the same in G and G'
- vertex n is not in the giant component iff none of its neighbours are
  - This gives $\rho = \sum_{d \geq 0} P_d \rho^d = \Phi(\rho)$
- The analysis of function Φ shows that it has a unique fixed point $\rho^* \in]0, 1[$

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n}. \log^{2k} n = \frac{\log^{2(k+1)} n}{n}$

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n}.\log^{2k} n = \frac{\log^{2(k+1)} n}{n}$
  - What about edges between vertices of Z_k ?

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n}.\log^{2k} n = \frac{\log^{2(k+1)} n}{n}$
  - What about edges between vertices of Z_k ?
- Conflicts are negligible until $\frac{\log^{2(k+1)} n}{n} = 1$, that is $k = \frac{\log n}{2 \log \log n} - 1$

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n}.\log^{2k} n = \frac{\log^{2(k+1)} n}{n}$
  - What about edges between vertices of Z_k ?
- Conflicts are negligible until $\frac{\log^{2(k+1)} n}{n} = 1$, that is $k = \frac{\log n}{2\log\log n} - 1$
- Then $|S_i| \approx (\log n)^{\frac{\log n}{2\log\log n} - 1} = \frac{\sqrt{(\log n)^{\frac{\log n}{\log\log n}}}}{\log n} = \frac{\sqrt{n}}{\log n}$

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n}. \log^{2k} n = \frac{\log^{2(k+1)} n}{n}$
  - What about edges between vertices of Z_k ?
- Conflicts are negligible until $\frac{\log^{2(k+1)} n}{n} = 1$, that is $k = \frac{\log n}{2 \log \log n} - 1$
- Then $|S_i| \approx (\log n)^{\frac{\log n}{2 \log \log n} - 1} = \frac{\sqrt{(\log n)^{\frac{\log n}{\log \log n}}}}{\log n} = \frac{\sqrt{n}}{\log n}$
- One can cover the vertex set by approx. $\sqrt{n} \log n$ balls of size approx. $\frac{\sqrt{n}}{\log n}$

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n} . \log^{2k} n = \frac{\log^{2(k+1)} n}{n}$
  - What about edges between vertices of Z_k ?
- Conflicts are negligible until $\frac{\log^{2(k+1)} n}{n} = 1$, that is $k = \frac{\log n}{2 \log \log n} - 1$
- Then $|S_i| \approx (\log n)^{\frac{\log n}{2 \log \log n} - 1} = \frac{\sqrt{(\log n)^{\frac{\log n}{\log \log n}}}}{\log n} = \frac{\sqrt{n}}{\log n}$
- One can cover the vertex set by approx. $\sqrt{n} \log n$ balls of size approx. $\frac{\sqrt{n}}{\log n}$
- the probability for two such balls not to be connected by an edge is $(1-p)^{\frac{n}{\log^2 n}} \leq e^{-\frac{1}{\log n}}$

Mean distance at the connectivity threshold

(very) roughly speaking

- $\mathbb{E}[\#\text{conflicts at stage } k] \leq np^2\mathbb{E}[Z_k^2] \approx \frac{\log^2 n}{n}. \log^{2k} n = \frac{\log^{2(k+1)} n}{n}$
  - What about edges between vertices of Z_k ?
- Conflicts are negligible until $\frac{\log^{2(k+1)} n}{n} = 1$, that is $k = \frac{\log n}{2 \log \log n} - 1$
- Then $|S_i| \approx (\log n)^{\frac{\log n}{2 \log \log n} - 1} = \frac{\sqrt{(\log n)^{\frac{\log n}{\log \log n}}}}{\log n} = \frac{\sqrt{n}}{\log n}$
- One can cover the vertex set by approx. $\sqrt{n} \log n$ balls of size approx. $\frac{\sqrt{n}}{\log n}$
- the probability for two such balls not to be connected by an edge is $(1-p)^{\frac{n}{\log^2 n}} \leq e^{-\frac{1}{\log n}}$
- so the proba for them to be connected is at least $1 - e^{-\frac{1}{\log n}} \sim \frac{1}{\log n}$

Mean distance at the connectivity threshold

- let us write $N = \sqrt{n} \log n$, and call $\tilde{G}$ the graph on the N balls that cover G

Mean distance at the connectivity threshold

- let us write $N = \sqrt{n} \log n$, and call $\tilde{G}$ the graph on the N balls that cover G
- we have $\log N \sim \frac{\log n}{2}$ and $\tilde{G}$ contains an ER graph on N vertices with $\tilde{p} = \frac{1}{2 \log N}$

Mean distance at the connectivity threshold

- let us write $N = \sqrt{n} \log n$, and call $\tilde{G}$ the graph on the N balls that cover G
- we have $\log N \sim \frac{\log n}{2}$ and $\tilde{G}$ contains an ER graph on N vertices with $\tilde{p} = \frac{1}{2 \log N}$
- In $\tilde{G}$ the probability for two given nodes to be at distance more than 2 is at most $(1 - \frac{1}{2 \log N})^{N-1} \leq e^{-\frac{N-1}{2 \log N}} \to 0$ when $N \to +\infty$.

Mean distance at the connectivity threshold

- let us write $N = \sqrt{n}\log n$, and call $\tilde{G}$ the graph on the N balls that cover G
- we have $\log N \sim \frac{\log n}{2}$ and $\tilde{G}$ contains an ER graph on N vertices with $\tilde{p} = \frac{1}{2\log N}$
- In $\tilde{G}$ the probability for two given nodes to be at distance more than 2 is at most $(1 - \frac{1}{2\log N})^{N-1} \leq e^{-\frac{N-1}{2\log N}} \to 0$ when $N \to +\infty$.
- Therefore, between any two vertices of G there exists with probability tending to 1 when $n \to +\infty$ a path of length $(\frac{\log n}{2\log\log n} - 1) + 1 + 2(\frac{\log n}{2\log\log n} - 1) + 1 + (\frac{\log n}{2\log\log n} - 1) \leq \frac{2\log n}{\log\log n}$

Configuration model – Molloy & Reed 1995

Input : an arbitrary degree distribution

Output : a random graph with this degree distribution

Configuration model – Molloy & Reed 1995

Input : an arbitrary degree distribution

Output : a random graph with this degree distribution

Generation process :

1. Assign a fixed number of semi-links to each node (according to the input degree distribution)
2. Pair the semi-links uniformly at random
3. Remove self-loops and multiple edges

Configuration model – Molloy & Reed 1995

Input : an arbitrary degree distribution

Output : a random graph with this degree distribution

Generation process :

1. Assign a fixed number of semi-links to each node (according to the input degree distribution)
2. Pair the semi-links uniformly at random
3. Remove self-loops and multiple edges

What degree distribution should we take as parameter ?

- The degree distribution of some real-world network
- A mathematically defined one, powerlaw $\mathbb{P}(k) \sim k^{-\alpha}$.

Configuration model : implementation and complexity

- Put the semi-links in a table of size $2m$
- Pick m times two of them uniformly at random

Properties of the configuration model

Four properties to check :

- Low global density ✓
  - the degree distribution is the parameter of the model and controls m : $m = \frac{\sum_{0 \leq k \leq n-1} k.N_k}{2}$

Properties of the configuration model

Four properties to check :

- Low global density ✓
- Short distances ✓

  Expansion property :

  - Degree of the extremity of one edge :
    $\mathbb{P}(d^\circ(ext) = k') = \frac{k'\mathbb{P}(k')}{<k>}$
  - Probability that following one edge leads to k new vertices :
    $q(k) = \mathbb{P}(d^\circ(ext) = k + 1)$
  - Expected number of new vertices following one edge :
    $\sum_k kq(k) = \frac{<k^2> - <k>}{<k>}$

Properties of the configuration model

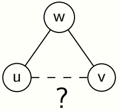
Four properties to check :

- Low global density ✓
- Short distances ✓
- Heterogeneous degrees ✓
  - the degree distribution is the parameter of the model

Properties of the configuration model

Four properties to check :

- Low global density ✓
- Short distances ✓
- Heterogeneous degrees ✓
- High local density

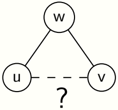


- ▶ Probability to have a link between u and k with $d^\circ(u) = k$ and $d^\circ(v) = k'$: $\mathbb{P}(uv|kk') = \frac{kk'}{<k>N}$

Properties of the configuration model

Four properties to check :

- Low global density ✓
- Short distances ✓
- Heterogeneous degrees ✓
- High local density



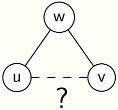
- Probability to have a link between u and k with $d^\circ(u) = k$ and $d^\circ(v) = k'$: $\mathbb{P}(uv|kk') = \frac{kk'}{<k>N}$
- Probability to have a link between u and v :

$$\begin{aligned} \mathbb{P}(triangle) &= \sum_{k\geq 1}\sum_{k'\geq 1}\frac{kk'}{<k>N}q(k)q(k') \\ &= \frac{1}{<k>N}\sum_{k\geq 1}kq(k)\sum_{k'\geq 1}k'q(k') \\ &= \frac{1}{N}\frac{(<k^2>-<k>)^2}{<k>^3} \end{aligned}$$

Properties of the configuration model

Four properties to check :

- Low global density ✓
- Short distances ✓
- Heterogeneous degrees ✓
- High local density ✗



- Probability to have a link between u and k with $d^\circ(u) = k$ and $d^\circ(v) = k'$: $\mathbb{P}(uv|kk') = \frac{kk'}{<k>N}$
- Probability to have a link between u and v :

$$
\begin{aligned}
\mathbb{P}(triangle) &= \sum_{k\geq 1}\sum_{k'\geq 1}\frac{kk'}{<k>N}q(k)q(k') \\
&= \frac{1}{<k>N}\sum_{k\geq 1}kq(k)\sum_{k'\geq 1}k'q(k') \\
&= \frac{1}{N}\frac{(<k^2>-<k>)^2}{<k>^3} \\
&\rightarrow 0 \text{ when } N \rightarrow +\infty
\end{aligned}
$$