Lecture 5 : Giant component of ER Graph

Recall: ER Graph $G(n,p)$, $np=\lambda$ ⇓ average degree; $n\uparrow$ $p\downarrow$ λ는 상수로 유지

Thm 1. Subcritical Regime $\lambda<1$

(예상) 큰 component가 나오지 않을 것이다.

If $\lambda<1$, for some constant $a=a(\lambda)$

$$\boxed{\Pr\left(|C_1^{(n)}| \le a\log n\right)} \xrightarrow{n\to\infty} 1$$

결론을 증명, 증명 하려고 하는 대상명제

↳ sense, 연습

where C_1 is the set of nodes in the largest component. C_2, C_3, C_4

→ 확률적 사실(n)이 확률 1로 사실이다 / 일어난다. if n 이 충분히 크다면

→ $|C_1| \le a\log n$ w.h.p. (with high probability)

$|C_1|\le a\log n$ with probability $1-\frac{1}{n}$

Thm2. $\lambda>1$, ——→ ~~~~

Thm3. $\lambda=1$, ——— ~~~~

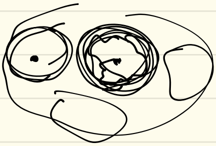
증명방법의 시작: GW Branching process, one-by-one Exploration

High-level overview of 증명

$\Pr(|C_1|>k) = 1-\Pr(|C_1|\le k)$

⇓ $|C_1|$의 분포(의 bound)

⇓ population of GW branching process with offspring distribution of Binomial$(n-1,p)$

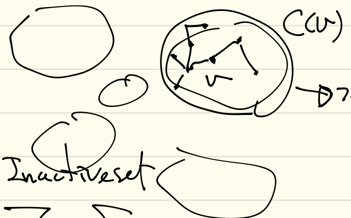


→ Intuition)

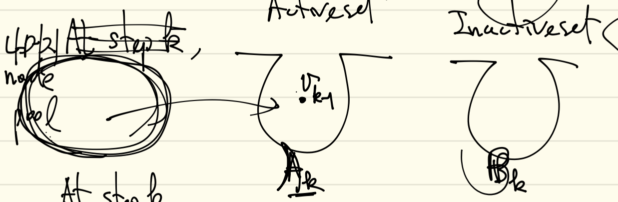
Bin$(n-1,p)$

One-by-One Exploration in ER graph

For an ~~arbitrary~~ node $v \in \{1, \ldots, n\}$, construct the connected component denoted by $C(v)$.



→ 가장 큰 덩어리는 보장도 없음



→ $A_{\not k} = 0$
→ connected component size

At step k,

(i) choose any node, say v_{k-1}, at A_k
(ii) $v_{k-1} \longrightarrow B_k$로 옮기고
(iii) all adjacent nodes to v_{k-1} ⟹ activated set A_k
남아있는 pool $\{1, 2, \ldots, n\} \setminus \{A_{k-1} \cup B_{k-1}\}$

초기조건 $A_0 = \{v\}$, $B_0 = \phi$.

Let D_k be the set of the selected nodes in the pool. $|D_k| = \xi_k$

$$\begin{cases} A_k = A_{k-1} \cup D_k \setminus \{v_{k-1}\} \\ B_k = B_{k-1} \cup \{v_{k-1}\} \end{cases}$$

$\underline{\underline{A}}$ $\underline{\underline{A}}$

~~Recur~~ Recursion:

$$\boxed{\begin{aligned} A_0 &= 1 \\ A_k &= A_{k-1} - 1 + \xi_k, \quad k > 0 \end{aligned}}$$

(Questions) GW-branching 에서의 recursion 이랑 같아?

• GW BP의 0-B-0 과 ER 0-B-0 의 차이점

GW

$A_0 = 1$

$A_k = A_{k-1} + 1 + \xi_k$

ξ_k offspring distribution 을 따르는 random variable

ER-Graph

$A_0 = 1$

$A_k = A_{k-1} + 1 + \xi_k$

ξ_k 의 분포? ① $Bin(n-1, p)$ (X)

공통점: total population size or connected component size 를 분석하기 위해 random walk 와 비슷한 방법을 이용

ξ_k 의 분포 | $\xi_1, \xi_2, \dots, \xi_{k-1}$ "conditioned on"

$\xi_k \sim Bin(n - (k-1) - A_{k-1}, p)$

전체 B_{k-1} inactive 정점 active 정점

$Pr(|C| \ge k) \Leftarrow Pr(|C(v)| \ge k)$

$n - (A_{k-1} + k - 1)$

Lemma 2.2 $A_k + k - 1 \sim Bin(n-1, 1-(1-p)^k)$

(증명) 숙제

Lemma 2.3 $Pr(|C(v)| > k) \le \exp(-\beta k)$, where $\beta = -\log\lambda - 1 + \lambda$

$Pr(X > k)$

⇓

bounding exponential 아래 감소

Proof of thm 1) $Pr(|C_1| > k) = Pr\left(\max_{1 \le i \le n} |C(i)| > k\right)$

$\le \sum_{i=1}^{n} Pr(|C_i| > k) = n \, Pr(|C(v)| > k)$

가장 큰 것 ≤ n × 어떤 것

max ≤ Σ sums

(Proof of Lemma 2.3)

$(1+\delta)\cdot\lambda k = k$, $1+\delta = \frac{1}{\lambda}$, $\boxed{\delta = \frac{1}{\lambda} - 1}$

$\mu = E(X) = npk = \lambda k$

$npk = \lambda k$

$$\Pr(|C(\omega)| > k) \leq \Pr(A_k > 0)$$

$Bin(n-1, 1-(1-p)^k)$

$$= \Pr(Bin(n-1, 1-(1-p)^k) \geq k) \text{ (from Lemma 2.2)}$$

$\leq Bin(n, kp)$

$\boxed{1-(1-p)^k} \leq kp$

$\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$

$$\leq \Pr(\underbrace{Bin(n, pk)}_{X} \geq k) \Rightarrow \Pr(X \leq (1+\delta)\mu)$$

$$\leq \Pr\left(Bin(n, pk) \geq \left(1 + \frac{1}{\lambda} - 1\right)\cdot \lambda k\right)$$

Chernoff Bound $\}$

$\Pr(X \leq (1+\delta)\mu) \leq e^{-\mu h(\delta)}$

$h(\delta) = (1+\delta)\log(1+\delta) - \delta$

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

$$= e^{-k(-\log\lambda - 1 + \lambda)}$$

(Proof of thm 1)

$$\Pr(|C_1| > k) = \Pr\left(\max_{i=1..n} |C(i)| > k\right) \leq \sum_{i=1}^{n} P(|C_i| > k)$$

$\Theta(n)$ $\Theta(\log n)$ $\Theta(\log\log n)$

$$= n \cdot P(|C(v)| > k)$$

$$\leq n \cdot e^{-k(\underbrace{-\log\lambda - 1 + \lambda}_{\beta})} \longrightarrow 0으로 가는가?$$

Question k 이 여러 값들 집어 넣어야

Choose $k = \frac{(1+\delta)\log n}{\beta}$, $\delta > 0$

$$n \cdot e^{-k\beta} = n \cdot e^{-\beta\frac{(1+\delta)\log n}{\beta}} = n\left(e^{-\log n}\right)^{(1+\delta)} = n \cdot n^{-(1+\delta)}$$

$$= n^{-\delta} \xrightarrow{n\to\infty} 0$$

$k = \Theta(n)$ $k = \log\log(n)$ $\log n \longrightarrow$

$a = \frac{1+\delta}{\beta}$

k는 잘 choose 해야 우리가 주장할 거리

전체 n $|C_1| = \Theta(n) \Rightarrow$ giant component

Thm2 (Supercritical Regime) $\lambda > 1$

$\Rightarrow$ $\boxed{\lambda > 1 \longrightarrow \text{"giant component"}}$

Assume $\lambda > 1$. Denote by $P_{ext}(\lambda)$ the extinction 확률 of a GW process with offspring distribution ~~with~~ of $\text{Poisson}(\lambda)$, i.e. $P_{ext}(\lambda)$ is the solution of

$$x = \exp(-\lambda(1-x)).$$

Then, for some constant $a' > 0$, and for all $\delta > 0$, the folling holds:

$a'(\lambda)$

(이해는 가지만 내가 할 수 있나 (?) 자력으로 가능? why?)

$$\lim_{n\to\infty} \Pr\left(\left| \frac{|C_1|}{n} - (1 - P_{ext}(\lambda)) \right| \le \delta \text{ and } |C_2| \le a' \log(n) \right) = 1$$

$|C_1| \approx (1 - P_{ext}(\lambda)) \cdot n$ and 두번째 큰 것은 $\log n$ order 보다 작거나 같다.

왜 이런 방법 이런 방법 나온 이런식을 표현

(Question) Thm1 가정. $\Pr(|C_1| \le a \log n) \to 1$, $|C_1| \le a \log n$ w.h.p.

Why? for any $\delta > 0$ $\Rightarrow$ 이런식으로 기술하는가? ////

직관력 책, 논문 Thm ......

Proof x, x, x, x $\Rightarrow$ line by line

$\boxed{\text{1step}}$ 구멍 이해

꺼림칙하다

"저자와 천재다"

"나같은 거리가 있다"

2step 과연 이렇게 증명할 수 있는가? Yes!

"이 세상에 천재는 없다"

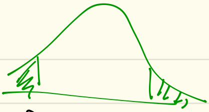
"저자도 이해가 안 되었다"

"정리를 천재처럼 했다"

상상 $\Longleftarrow$ 저자가 증명을 그렇게 하려고 했던 생각들과 똑같은 process

천재 $\rightarrow$ 범인으로 끌어내려야 한다.

$\boxed{\text{Lemma 2.4}}$ Chernoff Bound for Poisson trial.



$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu h(\delta)}, \quad h(\delta) = (1+\delta)\log(1+\delta) - \delta$$

$$\Pr(X \leq (1-\delta)\mu) \leq e^{-\mu h(-\delta)},$$

쉽게 증명.

$\boxed{\text{Lemma 2.5}}$ $\boxed{\lambda > 1}$. for all $\varepsilon > 0$, we can find a positive integer $k_0 > 0$, s.t. for ~~t~~ n large enough. "for sufficiently large n"

$$\left| \Pr\left(|C(v)| \leq k_0\right) - P_{ext}(\lambda) \right| \leq \varepsilon$$

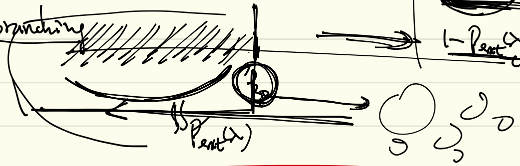
v를 골라서 v와 연결된 component

GW process, $\lambda > 1$

$P_{ext}(\lambda) < 1$.

(의미) ER graph 비슷 GW branching

$1 - P_{ext}(\lambda)$ component 크기,



(해석) 만약 $|C(v)| \leq k_0$

"small component"

v가 임의로 component를 고르면 small component를 고를 확률 $\approx P_{ext}(\lambda) < 1$.

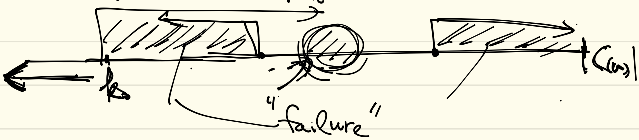
$\boxed{\text{Lemma 2.6}}$ $\lambda > 1$. For all $\varepsilon, \delta > 0$, we can find a positive integer $k_0 > 0$, s.t. for large enough,

$$\Pr\left(\underbrace{|C(v)| > k_0}_{\text{small이 아닌 component}} \text{ and } \left| \frac{|C(v)|}{\cancel{n}} - (1 - P_{ext}(\lambda)) \cancel{n} \right| > \delta \cancel{n} \right) \leq \varepsilon$$

이런일이 거의 벌어지지 않을거라.

$|C(v)| > (1 - P_{ext}(\lambda) + \delta) \cdot n$ or

$|C(v)| < (1 - P_{ext}(\lambda) - \delta) \cdot n$



ER

$C(v) \to$ (1) small $\approx P_{ext}(x) - \varepsilon \to$

(2) small(x), failure(o) $\varepsilon \to$

(3) small(x), Success(o) $\approx \dfrac{1-P_{ext}(x)-2\varepsilon}{1-P_{ext}(x)+2\varepsilon}$

Proof of Theorem 2

search

크게 잘 choose

k번까지

Consider the following algorithm:

ER

$1, \to C(1) \xrightarrow{\text{재시}} \text{small}(o) \to C(2) \to \text{small}(o) \to C(3) \to \text{small}(o)$

$\to$ failure(o) success(x)

$\to$ failure(x) success(o)

What we want to show: by choosing k smartly, such that 거의 1의 확률로

~~Success~~ Success를 많이 한다. ⇒ giant component

(Question) $\Pr\left(\text{발견할 사건}^{(n)}\right) \to 1$ Thm 1.

증명하기

$n \to \infty$ $\to 0$

⇒ "이 algorithm을 design 하고" "논리 연결" "이 방법 밖에 없나?" "이렇게 하면 무엇이 좋나?"

(i) From Lemma 2.5, the ~~prob that~~ prob of finding a small component $\leq P_{ext}(x)+\varepsilon$

(ii) ~~Lem~~ From Lemma 2.6, the prob of small(x) and failure(o) $\leq \varepsilon$

⇒ At each step, the prob that small(x), success(o) $\geq 1-P_{ext}(x)-2\varepsilon$

$$\Pr(\text{Success in at most } k \text{ step}) \geq \sum_{i=1}^{k} \underbrace{\left(P_{ext}(x)+\varepsilon\right)^{i-1} \cdot \left(1-P_{ext}(x)-2\varepsilon\right)}_{i\text{번째 step에서 success가 될 확률}}$$

$$= (1-P_{ext}(\lambda)-2\varepsilon)\ \frac{1-(P_{ext}(\lambda)-\varepsilon)^{k}}{1-P_{ext}(\lambda)+\varepsilon}$$

작게

k를 증가시키면

1에 가깝게 할 수 있음

at least $1-O(\varepsilon)$

k를 충분히 크게 잡아서

Geometry

$\to \boxed{|C_v| \leq d\log n} \to$ 책 "Homework" (hint) Def "Conjugate parameter"

Lemma 2.5, Lemma 2.6

길게 증명.

Lemma 2.5 ($\lambda > 1$). for all $\varepsilon > 0$, we can find a positive integer $k_0 > 0$, s.t. for n large enough "for sufficiently large n"

$$\left| \Pr(|C(v)| \leq k_0) - P_{ext}(\lambda) \right| \leq \varepsilon$$

v를 골라서 v와 연결된 component

($\Rightarrow$ small component를 만날 확률 $\approx P_{ext}(\lambda)$) $\to$ GW process with Poisson(λ)

(Recall) $A_0 = 1$

$A_k = A_{k-1} + \xi_k - 1$, $k > 0$

$\xi_1, \xi_2, \ldots \xi_k$ 의존

$|C(v)|$ 는 $(\xi_1, \xi_2, \ldots, \xi_k)$ 에 의해서 결정

Recall conditioned on $\xi_1, \xi_2, \ldots \xi_{i-1}$

$\xi_i \sim \mathrm{Bin}(n-1-\xi_1-\xi_2-\cdots-\xi_{i-1}, p)$

$\binom{n-1-\sum_{j=1}^{i-1} x_j}{x_i}$

$$\Pr(\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_k = x_k) =$$

$$= \prod_{i=1}^{k} \binom{n-1-\sum_{j=1}^{i-1} x_j}{x_i} p^{x_i} (1-p)^{n-1-\sum_{j=1}^{i-1} x_j - x_i}$$

상수

$x_1, x_2, \ldots, x_k$, n이 충분히 크면

아이하다 $= O(1)$

아이하다 $O(1)$

책: rigorous ($\times$)

$$= (1+o(1)) \prod_{i=1}^{k} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} = (1+o(1)) \prod_{i=1}^{k} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \quad \lambda = np$$

$\to$ order가 보다 작은 $\frac{1}{n}, \frac{1}{n^2} \ldots$

for sufficiently large n

$\boxed{\lambda > 1} \longrightarrow \boxed{P_{ext}(\lambda)} < 1.$

$1 - P_{ext}(\lambda)$

$$\Pr\left(|C(v)| < k_0\right) = \Pr\left(\text{for some } k \le k_0,\ \xi_1 + \cdots + \xi_k \le k-1\right)$$

ε차이

[illegible]

$$= (1+o(1)) \cdot \Pr\left(\text{population size of } \cancel{\text{GW}}\ \text{process with Poisson}(\lambda) \text{ offspring dist} \le k_0\right)$$

$P_{ext}(\lambda)(1+o(1))$ (by choosing k_0 large enough)

$$\boxed{\left|\Pr\left(|C(v)| < k_0\right) - P_{ext}(\lambda)\right| \le \varepsilon}$$

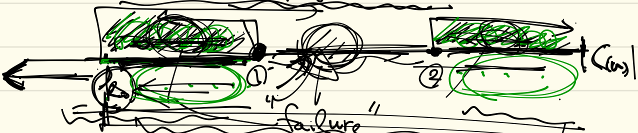
$\boxed{\text{Lemma 2.6}}$ $\lambda > 1$. For all $\varepsilon, \delta > 0$, we can find a positive integer $k_0 > 0$ s.t. for large enough,

$$\Pr\Big(\underbrace{|C(v)| > k_0}_{\text{small이어서 component}} \text{ and } \left|\,|C(v)| - (1-P_{ext}(\lambda))n\,\right| > \delta n\Big) \le \varepsilon,$$

이런 일이 거의 벌어지지 않는다

① $|C(v)| > (1 - P_{ext}(\lambda) + \delta) \cdot n$ or

② $|C(v)| < (1 - P_{ext}(\lambda) - \delta) \cdot n$



$= \Pr(\text{small}(\times), \text{failure}(\circ)) \le \varepsilon$

$$\boxed{\Pr\left(|C(v)| = k\right)} \le \Pr(A_k = 0) = \Pr(X_k = k-1), \text{ where } X_k \triangleq A_k + k - 1$$

Recall (Lemma 2.2) $\underbrace{A_k + k - 1}_{X_k} \sim \text{Bin}\left(n-1,\ 1-(1-p)^k\right) \le \text{Bin}(n, kp)$

k k, κ, κ'
Kappa
(page 27)

$\boxed{\text{Lemma 2.7}}$ If $k \le (1-P_{ext}(\lambda)-\delta)n$ or $k \ge (1-P_{ext}(\lambda)+\delta)n$,

$$\Pr(|C(v)|=k) \le e^{-k\kappa'}$$ ~~for~~ some positive constant $\kappa' > 0$.

(증명) 꼭 책이 있는건 아님?

$\boxed{\text{Homework}}$

~~Let~~ Proof of Lemma 2.6:

$$-\Pr\left(|C(v)| \ge k_0,\ \big||C(v)| - (1-P_{ext}(\lambda))n\big| \ge n\delta\right)$$

$$\le \sum_{k_0}^{(1-P_{ext}(\lambda)-\delta)n} \Pr(|C(v)|=k) + \sum_{k=(1-P_{ext}(\lambda)+\delta)n} \Pr(|C(v)|=k) \quad \text{—— (a)}$$

From Lemma 2.7

$$(a) \le \sum_{k \ge k_0} e^{-k\kappa'} = \frac{e^{-\kappa' k_0}}{1-e^{-\kappa'}} \le \varepsilon$$

by choosing k_0 large enough

(증명) → 여러분은 version → 기발하게 보여서는 안된다. 가능하면 상식적으로

수학적 증명 → 정확, 좋은 결론 $\boxed{}$ $\Pr(\text{small}) \approx P_{ext}(\lambda)$

증명

$\boxed{\text{Lemma 2.7 Lemma 2.6}}$

Subcritical / Supercritical / (critical ($\lambda = 1$)
(x) ↳ Martingale)