Lecture 7: Diameter of ER graph

For two vertices u, v

- $d_G(u,v)$ is the minimal ~~path~~ length (in hops) of a ~~path~~ path connecting the pair (u,v)
- Diameter of G is defined:

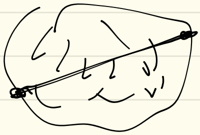
$$D(G) \triangleq \sup_{\substack{\text{all node} \\ \text{pair}(u,v)}} d_G(u,v)$$

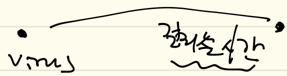
Why interesting? ① 상품 전송 (goods transportation) : upper bound on the time for goods to travel (be transported)

② diffusion, epidemics

If G is connected, → everybody infected

$\Longleftrightarrow$ $D(G)$ is finite



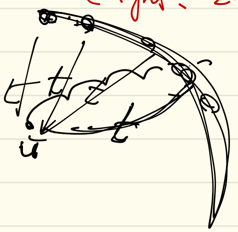


(i) General graph   Diameter $\Leftarrow$ G 특성 $\left\{ \begin{array}{l} \#\text{ of nodes } (n) \to \text{기본정보} \\ \text{maximal degree } (\Delta) \to \text{local 정보} \end{array} \right.$

Let maximal degree $=\Delta$.

$$n \le 1 + \Delta \frac{(\Delta-1)^D - 1}{\Delta - 2} \Longleftrightarrow D \geq \frac{\log\left(n\left[1-\frac{2}{\Delta}\right]+\frac{2}{\Delta}\right)}{\log(\Delta-1)}$$

$(\Delta > 2)$ G (tight? 글쎄요)

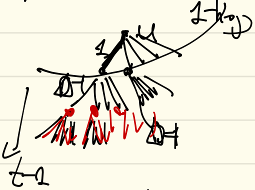


Pf) Let $\Gamma_t(u) := \{ v : d_G(u,v) = t \}$

$d_t(u) = |\Gamma_t(u)|$

By definition of Δ, for any node u $d_1(u) \leq \Delta$ and for all 자연수 t

귀납적 $d_t(u) \leq \Delta \cdot (\Delta-1)^{t-1}$



If diameter $\leq D$,

$n = 1 + d_1(u) + d_2(u) + \cdots + d_D(u) \leq 1 + \Delta\left(1 + (\Delta-1) + (\Delta-1)^2 + \cdots + (\Delta-1)^{D-1}\right)$

tight하지 않을것 같다. $\Rightarrow$ 정리 $\square$

(ii) ER graph 인 $G(n,p)$

Thm Let $\delta = (n-1)p$ (average node degree). $= np - p$

For n large enough, and for δ, s.t $\log n \ll \delta \ll \sqrt{n}$,

letting $D' = \left\lceil \frac{\log n}{2\log\delta} \right\rceil$, the following holds:

$\lim_{n\to\infty} \Pr\left(D(G(n,p)) \in \{2D'-3, 2D'-2, 2D'-1, 2D', 2D'+1\} \right) = 1$

해석) i) why $\log n \ll \delta \ll \sqrt{n}$??????

$\lambda = np = \log n + c \longrightarrow$ connect 관계 $e^{-e^{-c}}$ $c\to\infty$ 일때 1

확실하게 "connected"인 ER graph를 보겠다.

$\sqrt{n} \rightarrow$ ~~almost complete graph~~

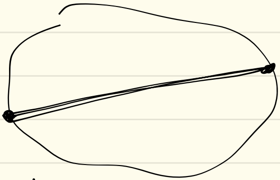
(ii) since $x-1 < \lceil x \rceil \leq x$, the diameter takes the value between $\lambda = \delta$

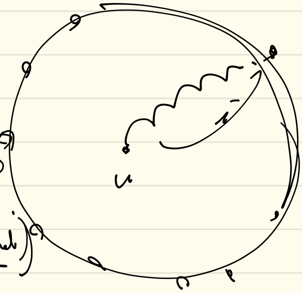
$2\left\lceil \frac{\log n}{2\log\lambda} \right\rceil - 4$ and $\left\lceil \frac{\log n}{\log\lambda} \right\rceil + 1$ $\quad \frac{\log n}{\log\lambda} \xrightarrow{n\to\infty} \infty$ 사용 $\log n$

$\frac{\log n}{\log(\log n)} \ll D$

정확히 구하기 $\frac{\log n}{\log d}$ ($\log n \ll d \le \sqrt{n}$) 해석

Diameter

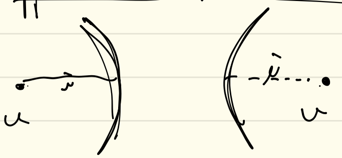




(step 1) $|\Gamma_i(u)|$: u로서 distance가 i인 node들의 수

random $d_i(u)$ 에 대응하면 (lower and upperbound)

(step 2) upperbound : "이것보다 클 수는 없다" : 임의의 두 노드 u, v, consider $\Gamma_i(u)$, $\Gamma_i(v)$.



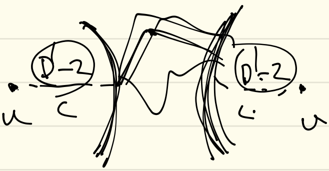
Prove that the prob that $\exists$ no link between $\Gamma_{D'}(u)$, $\Gamma_{D'}(v)$

→ Diameter가 at most 2D'+1

$\le 2D'+1$

(step 3) lower bound : "이것보다 작을 수는 없다" $\Longleftrightarrow$ $\boxed{\text{이것보다 작거나 같을 수 없다}}$ u, v

Consider a size $C = D' - 2$, and consider $\Gamma_c(u)$, $\Gamma_c(v)$



$\Gamma_c(u) \cap \Gamma_c(v) = \emptyset$ with high probability

→ diameter는 at least $2D'-3$

$d_i(u)$와 $|\Gamma_i(u)|$의 확률적인 특성 (Lemma 4.4) $(\delta = (n-1)\cdot p)$

$\boxed{\text{Lemma 4.4}}$ Given some $\varepsilon > 0$, define the following quantities

$$d_j^{\pm} = \begin{cases} (1\pm\varepsilon)^j \delta^j & \text{if } j = 1, 2 \\ (1\pm\varepsilon)^j \left(1 \pm \frac{\varepsilon}{\delta}\right)^{j-2} \delta^j & \text{if } j = 3, \dots, D' \end{cases}$$

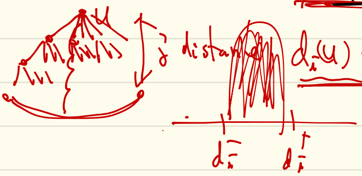
$\delta \to \infty$

Let us also define, for all $u \in \{1, 2, \dots, n\}$ and for all $i = 1, 2, \dots, D'$, the event $\mathcal{E}_i(u)$ by:

$$\mathcal{E}_i(u) := \{ d_i^- \le d_i(u) \le d_i^+ \}.$$

u $d_i(u) = |\Gamma_i(u)|$

Assume $\log n \ll \delta \ll \sqrt{n}$. Then for any fixed $K > 0$, for n large enough $\Pr(\mathcal{E}_i(u)) \ge 1 - D' n^{-K}$, $u \in \{1, \dots, n\}$, $i = 1, \dots, D'$.



$d_i(u) \approx \delta^i$ (concentration)

$\bar{\mathcal{E}}_i(u) = \mathcal{E}_i(u)$의 complement

(Proof) $\Pr(\mathcal{E}_i(u)) \ge \Pr(\underbrace{\mathcal{E}_1(u), \mathcal{E}_2(u), \dots,}_{A}\ \underbrace{\mathcal{E}_i(u)}_{B})$ → lowerbound

$\ge \Pr(\mathcal{E}_1(u), \mathcal{E}_2(u), \dots, \mathcal{E}_{i-1}(u)) - \Pr(\bar{\mathcal{E}}_i(u) \mid \mathcal{E}_1(u), \dots, \mathcal{E}_{i-1}(u))$

$\boxed{\Pr(A,B) \ge \Pr(A) - \Pr(B^c \mid A)}$

$\Pr(A) = p(A,B) + p(A,B^c)$

$= p(A,B) + p(B^c \mid A)\cdot p(A)$

$\le p(A,B) + p(B^c \mid A)$

P(A)의 확률이 거의 1이기 때문에

$\ge 1 - \sum_{j=2}^{i} \Pr(\bar{\mathcal{E}}_j(u) \mid \mathcal{E}_1(u), \dots, \mathcal{E}_{j-1}(u))$

→ 증명하라 $\boxed{\text{Homework 1}}$

$\mathcal{E}_0(u)$

$$\Pr(\Sigma_1(u), \Sigma_2(u), \Sigma_3(u)) - \Pr(\Sigma_3(u) \mid \Sigma_1(u), \Sigma_2(u))$$

$$\geq 1 - (\Pr(C^c \mid B) + \Pr(A^c \mid B, C)) \times (1 - (P(B^c) + \Pr(C^c \mid B) + \Pr(A^c \mid B, C)))$$

$$\Pr(B, C, A) \geq 1 - \Pr(C^c \mid B) = 1 - \frac{P(C^c, B)}{\Pr(B)}$$

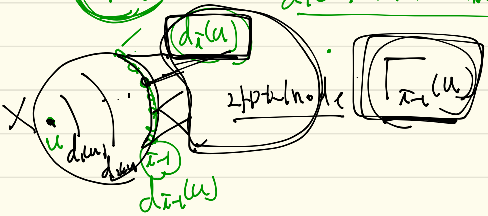
clarify

$i=2$: $\Pr(B, C) - \Pr(C^c \mid B) \leq 1 - \Pr(C^c \mid B)$

$\Pr(\overline{\Sigma}_i(u) \mid \Sigma_1(u), \ldots, \Sigma_{i-1}(u)) \Rightarrow$ handle

$d_i(u)$

$d_1(u) \ldots d_{i-1}(u)$



Note that, conditioned on $d_1(u), d_2(u), \ldots, d_{i-1}(u)$,

$d_i(u)$ admits a binomial distribution, i.e,

$$\mathcal{L}(d_i(u) \mid d_1(u), \ldots, d_{i-1}(u)) = \mathrm{Bin}(n - 1 - d_1(u) - d_2(u) - \cdots - d_{i-1}(u),\ 1 - (1-p)^{d_{i-1}(u)})$$

$$\boxed{\Pr(d_i^- \leq \mathrm{Bin}(\cdots, \cdots) \leq d_i^+)}$$

$$\cancel{= \Pr(}\ \leq \cancel{\Pr(d_i^- \leq \mathrm{Bin}(\cdots, \cdots) \geq d_i^-) + \Pr(\mathrm{Bin}(\cdots, \cdots) \leq d_i^+)}$$

$$\Pr(d_i^- \geq \mathrm{Bin}(\cdots, \cdots) \text{ or } \Pr(d_i^+ \leq \mathrm{Bin}(\cdots, \cdots))$$

$$\Pr\left(\mathrm{Bin}(\cdots,\cdots) \le d_i^-\right) + \Pr\left(\mathrm{Bin}(\cdots,\cdots) \ge d_i^+\right)$$

⇒ chernoff bound ⇒ chernoff bound

$d_{i-1}^+ \le d_i(u) \le d_{i-1}^+$

계산하기

$$\le \Pr\left(\mathrm{Bin}\left(n, 1-(1-p)^{d_{i-1}^+}\right) \ge d_i^+\right) + \Pr$$

$d_i^- \le d_i(u) \le d_i^+$

$$\Pr\left(\mathrm{Bin}\left(n-1-d_1^+-d_2^+-\cdots-d_{i-1}^+, 1-(1-p)^{d_{i-1}^-}\right) \le d_i^-\right)$$

Chernoff bound
$$\Pr(X \ge (1+\varepsilon)\mu) \le e^{-\mu h(\varepsilon)}, \quad h(x) = (1+x)\log(1+x) - x$$
$$\Pr(X \le (1-\varepsilon)\mu) \le e^{-\mu h(-\varepsilon)},$$

μ,

① $n\cdot\left(1-(1-p)^{d_{i-1}^+}\right) = \boxed{}$

② $\left(n-1-d_1^+-d_2^+-\cdots-d_{i-1}^+\right)\left(1-(1-p)^{d_{i-1}^-}\right) = \boxed{}$

계산

① $n\cdot\left(1-(1-p)^{d_{i-1}^+}\right)$ $\delta=(1+1)p$

$= \boxed{(1+o(1))\, n p \cdot d_{i-1}^+} = (1+o(1))\,\delta d_{i-1}^+$

(formal derivation)

$(1-p)^x \simeq px$ (when p is small compared to x)

$1-p \simeq e^{-p}$

$1-(1-p)^x \simeq 1-e^{-px} \simeq px$

$1-(1-p)^{d_{i-1}^+} \approx pd_{i-1}^+$

$d_{i-1}^+ \ll \sqrt{n}$

$\log n \ll \delta \ll \sqrt{n}$, $\delta = np$

$p \gg \frac{\log n}{n}$

$\frac{\log n}{n} \ll p \ll \frac{1}{\sqrt{n}}$

order가 달라지

② $\left(n-\cancel{(1+d_1^++d_2^++\cdots+d_{i-1}^+)}\right) \times \left(1-(1-p)^{d_{i-1}^-}\right) \simeq \boxed{n\cdot pd_{i-1}^- \times (1+o(1))} = (1+o(1))\,\delta d_{i-1}^-$

$\ll n$

$\to 0$

무시

P49 마지막부분 Homework $\boxed{n\cdot\left(1-(1-p)^{d_{i-1}^+}\right) = (1+o(1))\, n\cdot pd_{i-1}^+}$

(page 50 생략 ~~확률~~)

p의 조건 $\log n \ll (n-1)\cdot p \ll \sqrt{n}$

For all $j < D'$, we have:

$$d_j^- \le d_j^+ \le d_{D'-1}^+ = (1+\varepsilon)^2 \left(\frac{\delta+\varepsilon}{\delta}\right)^{\frac{\log n}{2\log\delta}-2} \cdot \delta^{\frac{\log n}{2\log\delta}}$$

d_j^+이 j에 대해 increasing

$$= (1+\varepsilon)^2 \frac{\delta^2}{(\delta+\varepsilon)^2} \left(\frac{\delta+\varepsilon}{\delta}\right)^{\frac{\log n}{2\log\delta}} \cdot \delta^{\frac{\log n}{2\log\delta}}$$

$\boxed{\log \ll \delta \le \sqrt{n}}$

$$= \left(\frac{\delta(1+\varepsilon)}{\delta+\varepsilon}\right)^2 (\delta+\varepsilon)^{\frac{\log n}{2\log\delta}}$$

$\delta \xrightarrow{n\to\infty} (n-1)p$

$$\frac{\delta(1+\varepsilon)}{\delta+\varepsilon} \to 1$$

$$= \left(\frac{\delta(1+\varepsilon)}{\delta+\varepsilon}\right)^2 \exp\left(\frac{\log n}{2} \cdot \frac{\log(\delta+\varepsilon)}{\log\delta}\right)$$

$$= \left(\frac{\delta(1+\varepsilon)}{\delta+\varepsilon}\right)^2 \left(\exp(\log n/2)\right)^{\frac{\log(\delta+\varepsilon)}{\log\delta}}, \quad \frac{\log(\delta+\varepsilon)}{\log\delta} \to 1$$

$$= 1 \cdot \sqrt{n}$$

그러므로 $\boxed{d_{D'-1}^+ = O(\sqrt{n})}$

$$1 + d_1^+ + d_2^+ + \cdots d_i^+ \le \sqrt{n} \cdot D' = \sqrt{n}\,\frac{\log n}{2\log\delta} \ll n \quad (\text{OK})$$

$\log \ll \delta < \sqrt{n}$

$$\Pr\left(\bar{\mathcal{E}}_i(u) \mid \mathcal{E}_1(u), \ldots, \mathcal{E}_{i-1}(u)\right) \le \exp\left(-(1+o(1))\,\delta d_{i-1}^- h(-\varepsilon_i)\right) + \exp\left(-(1+o(1))\,\delta d_{i-1}^+ h(\varepsilon_i)\right),$$

$$\varepsilon_i = \frac{d_i^+}{\delta d_{i-1}^+} - 1 = 1 - \frac{d_i^-}{\delta d_{i-1}^-}$$

$$\boxed{P(\mathcal{E}_i(u)) \ge 1 - D' n^{-k}}$$ for any constant k

Homework3 → 까지 ~~[illegible]~~ 마무리하는것.

Lecture 7 (part 3)

Page 51 (upperbound) : 2D'+1 보다는 작거나 같다

$$\Pr(D(G(n,p)) > 2D'+1) = \Pr\left(\max_{\text{all }(u,v)\text{ pair}} d_G(u,v) > 2D'+1\right)$$

$$= \sum_{u \neq v} \Pr(d_G(u,v) > 2D'+1) \quad \cdot n^{-k'}$$

$$\le \boxed{n^2 \cdot \Pr(d_G(u,v) > 2D'+1)} \xrightarrow{n\to\infty} 0 \quad (*)$$

For two arbitrary nodes u, v

$$\Pr(d_G(u,v) > 2D'+1) \longrightarrow 0$$

Let the event A be $A := \{d_G(u,v) > 2D'+1\}$

Let the event B be $B := \{\mathcal{E}_{D'}(u) \cap \mathcal{E}_{D'}(v)\}$

$d_{D'}(u) \in [d_{D'}^-, d_{D'}^+]$

$$= \Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c)$$

$$= \Pr(A|B)\cdot \cancel{p(B)} + \cancel{p(A|B^c)}\cdot p(B^c)$$

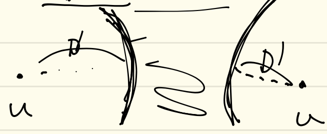
$$\le \underbrace{\Pr(A|B)}_{①} + \underbrace{p(B^c)}_{②}$$

$d_D^- \le d_D(u) \le d_D^+$

① $$\Pr(A|B) = \Pr\left(d_G(u,v) > 2D'+1 \,\middle|\, \mathcal{E}_{D'}(u) \cap \mathcal{E}_{D'}(v)\right)$$

$$\le (1-p)^{d_{D'}(u)d_{D'}(v)} \le (1-p)^{(d_D^-)^2}$$

$d_{D'}(u)$, $d_{D'}(v)$



$|\Gamma_{D'}(u)|$ $|\Gamma_{D'}(v)|$

$d_{D'}(u)$ $d_{D'}(v)$

$$(1-p)^{(d_D^-)^2} \longleftarrow d_D^- = (1-\varepsilon)^2\left(1-\frac{\varepsilon}{\delta}\right)^{D'-2}\delta^{D'}$$

$\delta = n-1p$

$$= \ldots$$

$\log(1-x) \le -x$

$\boxed{1-x \le e^{-x}}$

$$\le \exp\left(-\eta p n^{1-\varepsilon/\delta \log \delta (1+o(1))}\right) \quad \text{for some constant } \eta > 0$$

$$\le \exp(-n\delta) \qquad \text{(homework)}$$

② $P(B^c) = \Pr\left(\overline{\mathcal{E}_{D'}(u)} \cup \overline{\mathcal{E}_{D'}(v)}\right) \leq \Pr\left(\overline{\mathcal{E}_{D'}(u)}\right) + \Pr\left(\overline{\mathcal{E}_{D'}(v)}\right)$

$\leq D' n^{-k}$ for any constant k

Lemma 4.4

From ①, ②'s upper bounds,

$\Pr\left(d_G(u,v) > 2D'+1\right) \leq \exp(-\eta\delta) + D' n^{-k}$, for any k (for some $\eta > 0$)

$\boxed{\log n \leq \delta \ll \sqrt{n}}$

$\exp(-\eta \log n) = n^{-\eta}$

$\leq n^{-k'}$ for any $k' > 0$ for large n

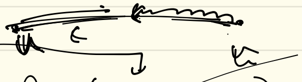
We apply this to $(*)$

$\leq n^2 \cdot n^{-k'} = n^{2-k'}$ by choosing $k' > 2$, we're done ∎

<Lowerbound) Diameter가 적어도 $2D'-3$ 이다.

For any two nodes u, v, let $c = D'-2$ $2c = 2D'-4$

$$\Pr(d_G(u,v) \le 2c) \longrightarrow 0$$



$$= \Pr\left(d_G(u,v) \le 2c \,\middle|\, \bigcap_{i=1}^{c} (\mathcal{E}_i(u) \cap \mathcal{E}_i(v))\right) \cdot \Pr(\quad)$$

$$+ \Pr\left(d_G(u,v) \le 2c \,\middle|\, \bigcup_{i=1}^{c} (\bar{\mathcal{E}}_i(u) \cup \bar{\mathcal{E}}_i(v))\right) \cdot \Pr(\ldots)$$

$$\le \underbrace{\Pr\left(d_G(u,v) \le 2c \,\middle|\, \bigcap_{i=1}^{c} (\mathcal{E}_i(u) \cap \mathcal{E}_i(v))\right)}_{(i)} + \Pr\left(\bigcup_{i=1}^{c} (\bar{\mathcal{E}}_i(u) \cup \bar{\mathcal{E}}_i(v))\right)$$

$$\le \sum_{i=1}^{c} \left(\Pr(\bar{\mathcal{E}}_i(u)) + \Pr(\bar{\mathcal{E}}_i(v))\right) \to 0$$

(ii) ($c = D'-2$ (고정되었다!))

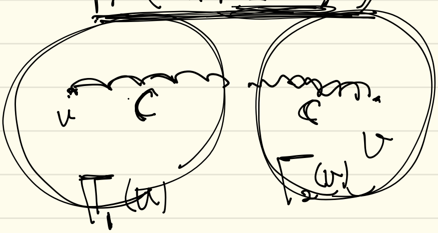
$$D' = \left\lceil \frac{\log n}{2\log \delta} \right\rceil$$

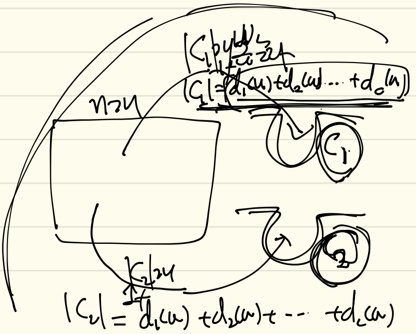
(i)

$$\Pr(d_G(u,v) \le 2c \mid d_1(u), d_2(u), d_3(u), \ldots, d_c(u), d_1(v), d_2(v), \ldots, d_c(v))$$

u와 v의 neighborhood size가 given되었을 때.

$$\le \Pr(C_1 \cap C_2 \ne \emptyset)$$





$c = D' - 2$ page 49's definition d_r^+

$|C_1| \le 1 + d_1^+(u) + d_2^+(u) + \cdots + d_c^+(u)$.

$|C_2| \le 1 + d_1(u) + d_2(u) + \cdots + d_c(u)$

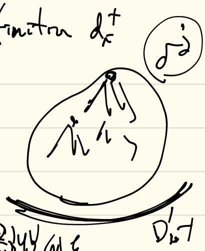
$\ll$ order $d_{D'-1}^+(u)$

$= o\left(d_{D'-1}^+(u)\right)$

$= o(\sqrt{m})$

[illegible]

$d_{D'-1}^+(u) = O(\sqrt{m})$

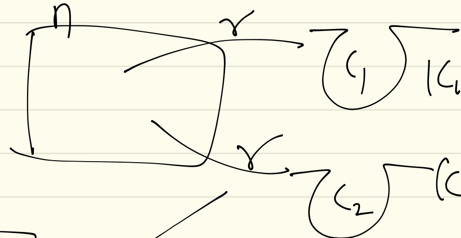


$\Pr(C_1 \cap C_2 \ne \phi)$,

$C_1 = o(\sqrt{m})$

$C_2 = o(\sqrt{m})$

$\Rightarrow$ 3 (ii)



$|C_1| = o(\sqrt{m})$

$|C_2| = o(\sqrt{m})$

Lemma 4.3

$$\Pr(C_1 \cap C_2 \ne \phi) = O\left(\frac{r^2}{n}\right)$$

$r = o(\sqrt{m})$

$\Rightarrow$ Homework

$$\text{(i)} \le O\left(\frac{\left(1 + d_1^+ + d_2^+ + \cdots d_c^+\right)^2}{n}\right) \xrightarrow{n \to \infty} 0$$

$o(\sqrt{m})$, $o(n)$

(ii)

$$\le \sum_{i=1}^{c} \left(\Pr(\bar{\Sigma}_i(u)) + \Pr(\Sigma_i(u)) \right) = O(D')\, n^{-k}$$

for any k

$\xrightarrow{n \to \infty} 0$

(i) $c = D' + 2$ (조심하자!)

$D' = \dfrac{\log n}{2 \log \delta}$ 라서 n^{-k}으로 decaying하는 것보다는 늦다