Lecture 8 (chapter 6) Small-world phenomenon

ER graph: 규칙적인, 기본적, giant component, diameter, connectivity,
structured graph: — small world 현상
—
—
—

⇒ 편지 Destination (이름, 주소, ...)
D씨

세상 좁다: Milgram's experiment

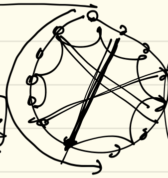
- 직접(x), known contact, local information
  - 높은 확률로 6-hop 정도 전달이 완료

심리학자 "세상 좁다"

Social network graph: 복잡하고 크지 않다.
웹page network, citation network: Erdos number chapter 6.2

Q1: 어떠한 특성, structure, 어떻게 발생할까? (shortcut) chapter 6_3
→ 수학적,

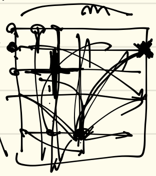
Q2: how much information
only local information → small world realize (?)
(navigability)



Diameter가 몇개의 shortcut에 의해서 줄어들었다.

6.2 Strogatz and Watts (SW) SW(n,p)

Model: Given some integer $m \geq 0$, consider $m \times m$ grid



Let $n = m^2$, $m = \sqrt{n}$. Large n regime

small world emerge (?)

- Connect nodes with two types

Type 1: local edges: just grid neighbors

Type 2: "shortcut" edges: 확률 p로 uniformly random하게 선택한 node 는 연결

(질문) ER graph는 이용하지 않았나?

Without "shortcut" edges, $D(SW(n,p)) = 2\sqrt{n}$

(Thm) Let p be fixed. Then for some constant A (depending on p)

$\boxed{D(SW(n,p)) \leq A\log(n)}$ with high probability. a.a.s

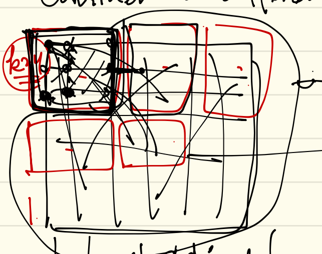
(meaning) shortcut 때문에 $D : \sqrt{n} \longrightarrow \log n$

↳ 아하! small world 현상이 이론적으로 일어날 수 있구나. (1998. 7. p) 15에 8면, 2000년 발표 정리

(증명) ER graph's diameter =

(Technique) Cluster graph approach, 주어진 SW(n,p) 에서 만들어서

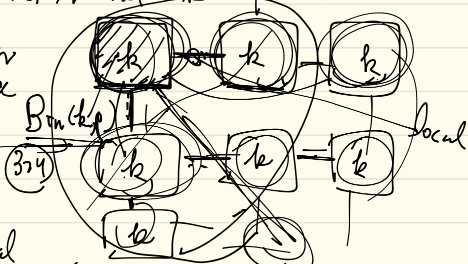
Consider a different network, called "cluster network" G' s.t



cluster의 연결을 define

5개가 k이고 (cluster square)

shortcut 반영



Neighboring cluster: local edges (of original graph SW(n,p))

~~The~~ 이

Each cluster (square) generates a random # of shortcut edges, whose distribution ~ Bin(k,p), 연결시킬 counter part : uniformly at random (out of all clusters)

$k = O(1)$

$SW(n,p) \longrightarrow G'$ induced clustergraph

(*) $\boxed{D(SW(n,p))} \le 2\sqrt{k}\left(2 + D(G')\right)$ $\quad D(G')$ orderwise

Homework 1. (*) 증명하기

Homework 2. 위 이런 방법이 useful한가?

Lemma 6.2. $\Gamma_1(u)$: a group of nodes (→ clusters) containing u of size $c\log n$, such that all nodes in $\Gamma_1(u)$ are connected by only "grid edges"

$\Gamma_2(u)$: reachable from nodes in $\Gamma_1(u)$ via shortcuts generated from $\Gamma_1(u)$

$\vdots$

Similarly, we define $\Gamma_i(u)$.

N' $\quad G'$ = set of nodes

def $\quad d_1(u) = |\Gamma_1(u)|,\ d_2(u) = |\Gamma_2(u)| \quad \cdots \quad d_i(u) = |\Gamma_i(u)|$.

Let $\varepsilon > 0$ be fixed, such that ① $kp(1-\varepsilon) > 1$, and

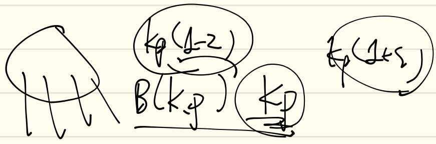
p는 사전에 주어지는 값이다

$$\frac{\log(kp(1+\varepsilon))}{\log(kp(1-\varepsilon))} < 2 \iff ② \ kp(1+\varepsilon) < kp(1-\varepsilon)^2$$

The constant $C > 0$ can be chosen s.t for all $u \in N'$, with $1 - O(n^{-2})$ the following holds:

$$kp(1-\varepsilon) \le \frac{d_i(u)}{d_{i-1}(u)} \le kp(1+\varepsilon), \quad i = 2, \dots, D,$$

where $\boxed{D = \left\lceil \frac{\log n}{2\log(kp(1+\varepsilon))} \right\rceil + 1}$



(의미) 지난번에: $\boxed{d_i^- \le d_i(u) \le d_i^+}$

$\Rightarrow$ Lemma 증명 방법: Martingale, → Azuma-Hoeffding inequality

(Proof of Theorem 6.1) $p(A) = p(A \cap B) + p(A \cap B^c)$

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

$$\le p(A|B) + p(B^c)$$

Let $\mathcal{E}_i(u)$ be (the event):

for any two nodes,

$$\mathcal{E}_i(u) = \left\{ (1-\varepsilon)kp \le \frac{d_i(u)}{d_{i-1}(u)} \le (1+\varepsilon)kp \right\}$$

$$\Pr\big(\underbrace{d_G(u,v)}_{2} > 2D + 2c\log n\big)$$

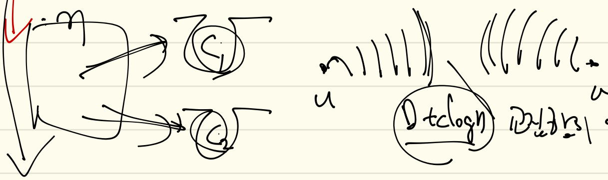
↑ 엄격 log n 써야

$$B: \bigcap_{j=2}^{D} \mathcal{E}_j(u) \cap \bigcap_{j=2}^{D} \mathcal{E}_j(v)$$

$$\le \boxed{\Pr\big(d_G(u,v) > 2D + 2c\log n \mid \mathcal{E}_2(u), \mathcal{E}_3(u), \ldots, \mathcal{E}_D(u), \mathcal{E}_2(v), \ldots, \mathcal{E}_D(v)\big)}$$

$$+ \boxed{\Pr\Big(\overline{\bigcap_{i=2}^{D} \mathcal{E}_i(u)}\Big) + \Pr\Big(\overline{\bigcap_{i=1}^{D} \mathcal{E}_j(u)}\Big)}$$

$\le o(n^{-2})$ from Lemma 6.2.



D + c log n 교집합이 없을 확률

pascal → [illegible] → $d_D(u)$

$(1-p)^{2(d_D)^2}$

$\le \pi$

↳ the probability that two sets of sizes $(c\log n)((1+\varepsilon)kp)^D$ picked uniformly at random from n nodes empty intersection

tx

Using $C\log n\, kp(1-\varepsilon_1)^D \geq C\log n \cdot \sqrt{n}$ $D = \left\lceil \frac{\log n}{2\log kp(1-\varepsilon_1)} \right\rceil + 1$

$P(C_1 \cap C_2) = \phi$ (Lemma 4.5) $C(1+o(1))\exp[(n-2r)\log(1+r/(n-2r)) + n\log(1-\frac{r}{n})]$

$\Pi \leq C(1+o(1))\exp(-C^2(\log n)^2(1+o(1)))$

$\Pr(d_G(u,v) > \underline{\quad\quad}) \leq \boxed{o(n^{-2}) +}$

$\Pr(D(G_W) \geq A\log n) \leq \max_{\substack{u \neq v \\ (u,v)}} \leq \sum \leq n^2 \, o(n^{-2}) +$

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

Homework 3   수학적으로 깔끔하게 처리하는것

Theorem 2장 2절 → diameter $\stackrel{?}{\leq} f(n)$

u   v   D를 D만큼 연결하기

$\Pr(d_G(u,v) > 2D + 2c\log(n))$

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

$\leq 2D + 2c\log n$

$\Pr(d_G(u,v) > f(n))$ $\Pr(\text{Diameter} > f(n)) \leq n^2 \cdot$

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

Diameter $\leq f(n)$

사건: $\Pr(A) = \Pr(A\cap B) + \Pr(A\cap B^c)$ (예시바꿔 집리하기 / 예시바꿔 정리하기)

$\leq \Pr(A|B) + \Pr(B^c) \longrightarrow 0$

Homework 4 질문 $d_1 = |\Gamma_1(u)|$ 을 $c\log n$ 으로 선택했을까?

Lecture 8 (part 3) Lemma 6.2

Thm 6.1

$$\Pr(d_{G'}(u,v) > 2D + 2c\log n) \le \Pr\Big(\bigcap_{i=2} \mathcal{E}_i(u)\Big) + \Pr\Big(\bigcap_{i=2} \mathcal{E}_i(w)\Big) + T$$

$O(n^2)$

$\sum_{u,v} n^2$

Lemma 6.2

n^2 보다 더 빠르게

Lemma 6.2. $\Gamma_1(u)$: a group of nodes (clusters) containing u of size $c\log n$, such that all nodes in $\Gamma_1(u)$ are connected by only "grid edges"

$\Gamma_2(u)$: reachable from nodes in $\Gamma_1(u)$ via shortcuts generated from $\Gamma_1(u)$

$\vdots$

Similary, we define $\Gamma_i(u)$.

N' 는 set of nodes

def $d_1(u) = |\Gamma_1(u)|$, $d_2(u) = |\Gamma_2(u)|$ ... $d_i(u) = |\Gamma_i(u)|$.

Let $\varepsilon > 0$ be fixed, such that ① $kp(1-\varepsilon) \ge 1$, and

p는 사전에 주어진 값이다

$$\frac{\log(kp(1+\varepsilon))}{\log(kp(1-\varepsilon))} < 2 \iff ② \; kp(1+\varepsilon) < kp(1-\varepsilon)^2$$

The constant $c > 0$ can be chosen s.t for all $u \in N'$, with $1 - O(n^{-2})$ the following holds:

$$kp(1-\varepsilon) \le \frac{d_i(u)}{d_{i-1}(u)} \le kp(1+\varepsilon), \quad i = 2, \dots, D$$

where $D = \left\lceil \frac{\log n}{2\log(kp(1+\varepsilon))} \right\rceil + 1$

잘 이해 안됨

(길고 복잡합니다.) Recall $\mathcal{E}_i(u) = \{(1-\varepsilon)kp \leq d_i(u) / d_{i-1}(u) \leq (1+\varepsilon)kp\}$

$\Pr\left(\bigcap_{i=2}^{D} \mathcal{E}_i(u)\right)$ 가 크다. , $\Pr\left(\overline{\bigcap_{i=2}^{D} \mathcal{E}_i(u)}\right)$ 가 작다.

Approach 1.

$\leq \Pr\left(\overline{\mathcal{E}_2(u)}\right) + \Pr\left(\overline{\mathcal{E}_3(u)}\right) + \cdots + \Pr\left(\overline{\mathcal{E}_D(u)}\right) \rightarrow$ 작다.

$d_2(u)$, $d_1(u)$ $\quad$ $d_3(u)$, $d_2(u)$ $\quad$ $d_D(u)$, $d_{D-1}(u)$

$d_1(u), \ldots, d_{i-1}(u)$ 예측가능 $\rightarrow$ $d_i(u)$ 확률로 예측가능

$\Pr\left(\bigcap_{i=2}^{D} \mathcal{E}_i(u)\right) = \prod_{i=3}^{D} \mathcal{E}_2(u) \cdot \Pr\left(\mathcal{E}_i(u) \mid \mathcal{E}_{i-1}(u), \ldots, \mathcal{E}_2(u)\right)$

chain rule

$\hookrightarrow \Pr(A_4, A_3, A_2, A_1) = p(A_4 \mid A_1, A_2, A_3) \cdot p(A_3 \mid A_1 A_2)\ p(A_2 \mid A_1) \cdot p(A_1)$

$\Pr\left(\mathcal{E}_i(u) \mid \mathcal{E}_{i-1}(u), \ldots, \mathcal{E}_2(u)\right)$ 가 크다 ~~$o(n^{-2})$~~

$\geq 1 - n^{k}$

$\hookrightarrow \boxed{\Pr\left(\overline{\mathcal{E}_i(u)} \mid \mathcal{E}_{i-1}(u), \ldots \mathcal{E}_2(u)\right) \text{ 가 작다.}}$ $/\!/A$ $/\!/B$

$\leq n^{-2}$ for any $i \geq 2$

$\prod_{i=3}^{D} (1 - n^{k})$ $\boxed{\text{발산할 수도}}$ $\geq 1 - \cancel{B}\, o(n^{-2})$

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

엡실론

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

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

$\leq p(A \mid B) + p(B^c)$

어떤 event C를 선택 $P(C)$가 명확히 작게

$$P(A|B) = P(A\cap C|B) + P(A\cap C^c|B)$$

event C를 어떻게 선택

→ 우리의 확률계산을 편하게

도입

$$\le \boxed{\Pr(A|B,C) + \Pr(C^c|B)}$$

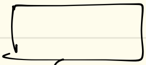
$P(C)$는 크다

왜 확률 계산 편하게??

$$\frac{P(A\cap B\cap C)}{P(B)} \le \frac{P(A\cap B\cap C)}{P(B\cap C)}$$

증명 = 그림그리기

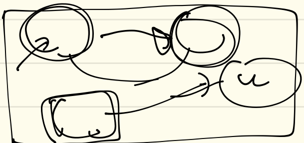
증명



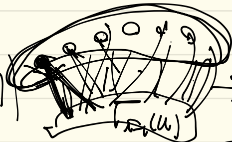
cafe, 랜덤성, 자연성

→ right 80% 이상 확인

확인



기본적인 사실들



$d_{i-1}(u) = |\Gamma_{i-1}(u)|$ → 개수

i) conditioned $d_1(u), d_2(u), \dots, d_{i-1}(u)$,

# of shortcuts from $\Gamma_{i-1}(u) \sim Bin(k\, d_{i-1}(u), p)$

Let such a random variable be T.

ii) T와 $d_i(u)$와의 관계 : ⇒ Balls and Bins 관계로 이해가능

$d_i(u)$: # of occupied bins, among $n' - d_1(u) - d_2(u) - \dots - d_{i-1}(u)$ after throwing T balls

앞에서 했던 : ~~A-H inequality~~

Conditioned on $T, \Gamma_1(u), \dots, \Gamma_{i-1}(u)$,

$$\Rightarrow \Pr\left(d_i(u) - \bar{d}_i(u) \ge x \mid T, \Gamma_1(u), \dots, \Gamma_{i-1}(u)\right) \le \boxed{\exp\left(-\frac{x^2}{2T}\right)}$$

⇓ A-H inequality

$d_i(u)$의 확률적 중앙

Let the event $C = \{ T \in I \}$, $I = [\, kp\, d_{i-1}(u)(1-\frac{\varepsilon}{2}),\ kp\, d_{i-1}(u)(1+\frac{\varepsilon}{2}) \,]$

$Pr(C) \longrightarrow$ 작다 (0, 크다

$E(T) = kp\, d_{i-1}(u)$

$$Pr\left(\overline{\mathcal{E}_i(u)} \mid \mathcal{E}_2(u), \ldots, \mathcal{E}_{i-1}(u)\right) \to 0$$

왜 C라는 event를 T의 조건으로 사용

$$\le Pr\left(\boxed{T \in I} \mid \mathcal{E}_2(u), \ldots, \mathcal{E}_{i-1}(u)\right)$$

$$+ Pr\left(\overline{\mathcal{E}_i(u)} \mid \mathcal{E}_2(u), \ldots, \mathcal{E}_{i-1}(u), T \in I\right)$$

$$= Pr\left(T \notin I \mid \mathcal{E}_2(u), \ldots, \mathcal{E}_{i-1}(u)\right) \longrightarrow ①$$

$$+ Pr\left(\frac{d_i(u)}{d_{i-1}(u)} \ge kp(1+\varepsilon) \mid \mathcal{E}_2(u), \ldots, \mathcal{E}_{i-1}(u), T \in I\right) \quad ②$$

$$+ Pr\left(\frac{d_i(u)}{d_{i-1}(u)} \le kp(1-\varepsilon) \mid \mathcal{E}_2(u), \ldots, \mathcal{E}_{i-1}(u), T \in I\right) \quad ③$$

책의 6.8

(Q) $\boxed{\text{왜 이렇게 하는가?}}$ ⟹ T를 기준으로 event C를 만든다는 사실 을 편하게 잡고

한꺼번에 처리하기?

①과 ②는 명확히 확률 계산

① $T \sim Bin(k\, d_{i-1}(u), p)$

$\Gamma_1(u)$ construction

↳ $c\log n$ 사이즈

$\Gamma_1(u)$

$|\Gamma_1(u)| = c\log n$ $\boxed{kp(1-\varepsilon) > 1}$

↓ 증가

$d_{i-1}(u) = |\Gamma_{i-1}(u)| \ge c\log n$

conditioned on $\bigcap_{j=2}^{i-1} \mathcal{E}_j(u)$

$kp(1-\varepsilon) \le \frac{d_j}{d_{j-1}(u)} \le kp(1+\varepsilon)$

① $\le 2\exp\left(-(c\log n)\, h(\varepsilon/2)\right) \longrightarrow$ C를 충분히 크게 선택해서

Lemma 2.4

n^{-k} for any desired $k > 0$ ⟹ check!

Homework 5.

② We choose $T \in I$, such that $\exp\left(-x^2/2T\right)$ is maximized,

i.e., $\underline{\sup_{T\in I} \exp\left(-x^2/2T\right)}$ $\boxed{\text{I는 퀄수있는한 크게}}$

$\Pr\left(d_i(u) \geq kp(1+\varepsilon)\cdot d_{i-1}(u)\right)$

$\Pr\left(d_i(u) - \bar{d}_i(u) \geq x\right)$

$$\bar{d}_i(u) = \underbrace{(n' - d_1(u) - d_2(u) \cdots - d_{i-1}(u))}_{\text{6ini 개수}} \cdot \underbrace{\left(1 - \left(1 - \frac{1}{n'}\right)^{T}\right)}_{\text{확률}}$$

$\therefore x = d_{i-1}(u)\,kp(1+\varepsilon) - d_i(u)$

$$\Pr\left(d_i(u) - \bar{d}_i(u) \geq \boxed{d_{i-1}(u)\,kp(1+\varepsilon) -} \;\middle|\; T \in I, \varepsilon_0(u) \ldots \varepsilon_0(u)\right)$$

~~$t \leq i$~~, Using condition (6.2) $kp(1+\varepsilon) < (kp(1+\varepsilon))^2$,

$\hookrightarrow \boxed{d_{i-1}(u) = o(n)}$.

will $x \geq \boxed{d_{i-1}(u)\,kp(1+\varepsilon) - (1+o(1))\,T}$.

At inequality을 적용하면 ② $\leq \exp(-\varepsilon'(\log n))$ for some constant $\varepsilon' > 0$

충분히 만큼 크게 선택

$\leq n^{-k}$, for any desired $k > 0$.

Homework 6 확률을 clarify 해볼것.