BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Trần Mỹ Đức

BAO CỦA ĐỒ THỊ NGẪU NHIÊN

LUẬN VĂN THẠC SĨ TOÁN HỌC

Hà Nội - 2023

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Trần Mỹ Đức

BAO CỦA ĐỒ THỊ NGẪU NHIÊN

LUẬN VĂN THẠC SĨ TOÁN HỌC

Chuyên ngành: Toán ứng dụng
Mã số: 8 46 01 12

NGƯỜI HƯỚNG DẪN KHOA HỌC:
1. TS. Cấn Văn Hảo

Hà Nội - 2023

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn chịu trách nhiệm trước phát luật.

Hà Nội, tháng 10 năm 2023

Học viên

Trần Mỹ Đức

LỜI CẢM ƠN

Đầu tiên, tôi xin được bày tỏ lòng biết ơn sâu sắc nhất của bản thân mình tới TS. Cấn Văn Hảo, người thầy đã trực tiếp hướng dẫn tôi trong quá trình tìm đề tài và hoàn thành luận văn. Luận văn này được hoàn thành dưới sự chỉ bảo tận tình của thầy trong suốt một thời gian dài. Xuyên suốt, thầy đã luôn quan tâm, giúp đỡ và động viên tôi.

Tiếp theo, tôi xin được trân trọng cảm ơn sự giúp đỡ và tạo điều kiện thuận lợi về môi trường học tập của ban Lãnh đạo, phòng Đào tạo, các phòng chức năng của Viện Toán học và Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Tôi cũng xin trân trọng cảm ơn Ban giám hiệu và Bộ môn Chuyên Toán Trường Trung học phổ thông Chuyên Khoa học Tự nhiên, Trường ĐH KHTN, ĐHQGHN. Các thầy cô ở Viện và đồng nghiệp ở Trường đã luôn hỗ trợ tôi trong học tập, công tác. Bên cạnh đó, trong quá trình học tập, nghiên cứu và thực hiện Luận văn, tôi còn nhận được nhiều sự quan tâm, góp ý, hỗ trợ quý báu của các thầy cô, anh chị và bạn bè trong và ngoài Viện Toán học. Tôi xin trân trọng cảm ơn.

Tôi muốn gửi lời cảm ơn tới Quỹ Đổi mới sáng tạo Vingroup - đơn vị hỗ trợ tài chính giúp tôi hoàn thành hai năm học thạc sĩ. Học bổng được Quỹ VINIF trao dưới mã số VINIF.2020.ThS.VTH.10 và VINIF.2021.ThS.VTH.06 là nguồn động viên tinh thần và vật chất to lớn dành cho tôi.

Cuối cùng, tôi xin cảm ơn gia đình, người thân và bạn bè đã luôn sát cánh, động viên và khích lệ tôi trong suốt quá trình học tập và nghiên cứu.

Học viên

Trần Mỹ Đức

Mục lục

DANH MỤC CÁC KÍ HIỆU

$\mathbb{N}$	Tập các số tự nhiên
$\mathbb{R}$	Tập các số thực
$\mathbb{R}_+$	Tập các số thực dương
$\inf E$	Cận dưới đúng của E
$\exp(n)$	e^n
$\|E\|$	Lực lượng của tập E
$u \sim v$	Đỉnh u kề với đỉnh v
∞	Dương vô cực
$[n]$	Tập các số nguyên dương không vượt quá n
$\deg(v)$	Bậc (số đỉnh kề) của đỉnh v
$A \times B$	Tích Descartes của tập A và tập B
$G_{n,p}$	Đồ thị ngẫu nhiên n đỉnh, và hai đỉnh bất kỳ được nối với nhau một cách độc lập với xác suất p
K_n	Đồ thị đầy đủ n đỉnh (tất cả các đỉnh đều được nối với nhau)
$A_n \approx B_n$	$A_n = (1 + o(1))B_n$
$A_n \gg B_n$	$A_n/B_n \to \infty$ khi $n \to \infty$
$Exp(\lambda)$	Phân phối mũ tham số λ (kì vọng bằng $1/\lambda$)
$Bin(n,p)$	Phân phối nhị thức.
$U[a,b]$	Phân phối đều trên đoạn $[a,b]$.
$\mathbf{1}_{\{.\}}$	Hàm chỉ tiêu
$i.i.d$	Độc lập, cùng phân phối
$w.h.p$	With high probability - Với xác suất cao
$h.c.c$	Hầu chắc chắn
$X \prec Y$	$\forall x : \mathbb{P}(X \geq x) \leq \mathbb{P}(Y \geq x)$.

Giới thiệu

Xét đồ thị vô hướng, liên thông $G = (V, E)$ và một hàm trọng $l : E \to \mathbb{R}_+$. Kí hiệu $d_G(u, v)$ là độ dài đường đi ngắn nhất giữa u và v trong G, tức là:

$$d_G(u, v) = \inf \Big\{ \sum_{e \in \gamma} l(e),\, \gamma : \text{đường đi từ } u \to v \Big\},$$

trong đó inf được lấy trên tất cả các đường đi γ có thể từ u đến v. Nếu $l(e) = 1$ với mọi $e \in E$ thì $d_G(u, v)$ đơn giản là khoảng cách đồ thị giữa u và v.

Một đồ thị con $G' = (V', E')$ được gọi là một *đồ thị bao* (*graph spanner*) của G (hay đơn giản *bao*) với hàm hiệu chỉnh f trên $P \subseteq V \times V$ nếu

$$d_{G'}(u, v) \leq f(d_G(u, v)), \quad \forall (u, v) \in P.$$

Chú ý rằng $d_G(u, v) \leq d_{G'}(u, v)$ với mọi đồ thị con G', và mọi đỉnh u, v. Do đó, bài toán chỉ có ý nghĩa khi $f(x) \geq x$ với mọi $x \geq 0$. Như vậy các *bao* được xác định cụ thể qua hàm hiệu chỉnh f và tập con $P \subseteq V \times V$ quy định các cặp đỉnh mà cần được "bảo toàn" độ dài. Một số những lựa chọn hàm hiệu chỉnh f điển hình (xem thêm tại [1]) là:

- Hàm nhân tính: $f(x) = tx$ với $t \geq 1$. Khi đó G' được gọi là $t-$bao.
- Hàm cộng tính: $f(x) = x + \beta$ với $\beta \geq 0$. Khi đó G' được gọi là $+\beta-$bao.

- Hàm tuyến tính: $f(x) = \alpha x + \beta$ với $\alpha, \beta \geq 0$. Khi đó G' được gọi là $(\alpha, \beta)-$bao.

Với $t-$bao, nếu như không nói thêm gì, ta luôn xét $V' = V$ và $P = V \times V$. Như vậy, một câu hỏi tự nhiên mà ta có thể quan tâm là: Cho $G = (V, E)$, và cố định $t \geq 1$, tìm $E' \subseteq E$ sao cho $G' = (V, E')$ là một $t-$bao của G, tức:

$$d_{G'}(u, v) \leq t \cdot d_G(u, v), \quad \forall u, v \in G. \tag{0.0.1}$$

Để thuận tiện, đôi khi ta cũng coi chính E' là $t-$bao. Nhận xét rằng G' phải liên thông; và nếu t càng gần 1 thì E' càng gần với E.

Như vậy, với việc cho trước G là một đồ thị vô hướng, có trọng, và liên thông, ta có thể hiểu một bao của G là một đồ thị con G' mà nó bảo toàn độ dài của các đường đi ngắn nhất trong G với mức độ hiệu chỉnh hay sai số đã được định sẵn. Khái niệm *bao* (cụ thể là *t-bao*) được giới thiệu bởi Peleg và Schäffer [2] từ năm 1989. Cũng trong bài báo đó, hai tác giả đã chỉ ra rằng với các đồ thị không trọng, việc chứng minh một $t-$bao của G phải chứa tối thiểu m cạnh là một bài toán *NP-đầy đủ*.

Cho tới nay, khái niệm này cùng với các mở rộng của nó có nhiều ứng dụng lý thuyết trong lĩnh vực tổ hợp, khoa học máy tính cũng như ứng dụng thực hành trong các vấn đề thiết kế mạng khác nhau, xem thêm [1]. Trong luận văn này, chúng tôi tìm hiểu bài toán về kích thước $t - bao$ (cỡ $|E'|$) của các đồ thị ngẫu nhiên lớn qua hai kết quả được đưa ra gần đây của Alan Frieze và Wesley Pegden [3, 4].

Chương 1

Kiến thức chuẩn bị

Trong chương này, tôi xin trình bày một số kiến thức cơ sở và kết quả được sử dụng trong quá trình chứng minh các kết quả chính của luận văn.

1.1 Một số kiến thức cơ sở trong Lý thuyết đồ thị

Chúng ta có thể hiểu rằng đồ thị là một tập các đối tượng mà giữa hai đối tượng có thể có hoặc không có mối liên hệ nào đó. Trong luận văn ta chỉ bàn tới đồ thị đơn, vô hướng do đó nếu không nói gì thêm thì ta sẽ hiểu một đồ thị là đơn, vô hướng.

Định nghĩa 1.1. *Cho n là một số nguyên dương, một* ***đồ thị đơn, vô hướng*** *gồm n đỉnh là một gặp tập hợp $G := (V, E)$, trong đó V là một tập hợp gồm n phần tử và E là một họ các tập con hai phần tử của V mà mỗi tập con chỉ xuất hiện đúng một lần.*
Mỗi phần tử $x \in V$ được gọi là một đỉnh. Mỗi phần tử $\{x, y\} \in E$ được gọi là một cạnh nối x với y, khi đó ta kí hiệu $x \sim y$ và nói hai đỉnh này kề nhau. Do đó, tập V, E còn tương ứng được gọi là tập đỉnh và tập cạnh của đồ thị G.

Nhận xét 1.1. *Ta có một số lưu ý quan trọng:*

- *Người ta thường mô tả mỗi đỉnh $x \in V$ bởi một điểm trong mặt phẳng và mỗi cạnh $\{x, y\} \in E$ bởi một đường nối trực tiếp giữa hai điểm x và y.*

- *Đồ thị đơn, vô hướng có n đỉnh mà hai đỉnh bất kỳ đều kề nhau được gọi là đồ thị đầy đủ và được kí hiệu là K_n. Rõ ràng số cạnh của K_n là $\binom{n}{2}$.*

Định nghĩa 1.2. *Số đỉnh kề với một đỉnh $x \in V$ được gọi là **bậc** của x và kí hiệu là $\deg(x)$.*

Nhận xét 1.2. *Mối quan hệ giữa số cạnh của đồ thị với bậc của các đỉnh được thể hiện qua công thức:*

$$\sum_{x \in V} \deg(x) = 2|E|.$$

Định nghĩa 1.3. *Trong đồ thị G, **một đường đi (a path)** là một chuỗi luân phiên giữa đỉnh và cạnh, bắt đầu và kết thúc bởi hai đỉnh, sao cho không có đỉnh nào xuất hiện nhiều hơn một lần. Ta thường kí hiệu và biểu diễn đường đi là $P = v_1 \to v_2 \to \cdots \to v_n$ hoặc đơn giản $P = (v_1, v_2, \ldots, v_n)$ hay $P : v_1 \to v_n$, và nói đường đi này đi từ v_1 tới v_n. Ở trên, lưu ý rằng $\{v_i, v_{i+1}\} \in E$ $\forall i$, và trong trường hợp $v_n \sim v_1$ thì ta thu được **một chu trình** $(v_1, v_2, \ldots, v_n, v_1)$. Nói cách khác, chu trình là hợp của một đường đi và cạnh nối đỉnh đầu và đỉnh cuối của đường đi đó.*

Định nghĩa 1.4. *Một đồ thị được gọi là liên thông nếu với bất kỳ hai đỉnh x, y nào cũng tồn tại một đường đi từ x tới y.*

Nhận xét 1.3. *Một đồ thị bất kỳ luôn có thể được phân hoạch thành các thành phần liên thông, tức là với hai phần tử x, y thuộc một thành phần liên thông thì có đường đi nối x với y, còn nếu chúng thuộc hai thành phần liên thông khác nhau thì không có đường đi nối giữa chúng.*

Định nghĩa 1.5. *Xét một đồ thị vô hướng, liên thông $G = (V, E)$. Trên mỗi cạnh của đồ thị ta trang bị thêm một đặc trưng là **trọng số** hay **độ dài** của cạnh, được thể hiện qua một hàm trọng $l : E \to \mathbb{R}_+$. Khi đó ta nói mỗi cạnh $e \in E$ có trọng số hay độ dài là $l(e)$. Khi đó xét một đường đi $P = v_1 \to v_2 \to \cdots \to v_n$, ta có độ dài của đường đi này chính bằng tổng độ dài các cạnh trong đường đi, và kí hiệu là $l(P)$. Cụ thể:*

$$l(P) = \sum_{e \in P} l(e).$$

Trường hợp $l(e) = 1 \ \forall e \in E$ thì độ dài của đường đi hay chu trình chính bằng số canh của đường đi hay chu trình đó.

Định nghĩa 1.6. *Kí hiệu $d_G(u, v)$ là **độ dài đường đi ngắn nhất** giữa hai đỉnh u và v trong G. Khi đó:*

$$d_G(u, v) = \inf \Big\{ \sum_{e \in P} l(e),\ P : u \to v \Big\},$$

trong đó inf *được lấy trên tất cả các đường đi P có thể có từ u đến v. Nếu $l(e) = 1$ với mọi $e \in E$ thì $d_G(u, v)$ được gọi là khoảng cách đồ thị giữa u và v.*

1.2 Một số bất đẳng thức xác suất

1.2.1 Bất đẳng thức Markov

Định lý 1.1. *Cho X là biến ngẫu nhiên không âm. Khi đó:*

$$\forall u > 0 : \mathbb{P}(X \geq u) \leq \frac{\mathbb{E}(X)}{u}.$$

1.2.2 Bất đẳng thức Chebyshev

Định lý 1.2. *Cho X là biến ngẫu nhiên bất kỳ. Khi đó:*

$$\forall u > 0 : \mathbb{P}(|X - \mathbb{E}(X)| \geq u) \leq \frac{Var(X)}{u^2}.$$

1.2.3 Bất đẳng thức Chernoff cho phân phối nhị thức

Định lý 1.3. *[Chương 2, [3]] Xét biến ngẫu nhiên $X \sim Bin(n,p)$. Khi đó với mọi $0 \le \varepsilon \le 1$ và $\alpha > 0$, ta có:*

$$\mathbb{P}(X \le (1-\varepsilon)np) \le e^{-\varepsilon^2 np/2},$$
$$\mathbb{P}(X \ge (1+\varepsilon)np) \le e^{-\varepsilon^2 np/3},$$
$$\mathbb{P}(X \ge \alpha np) \le \left(\frac{e}{\alpha}\right)^{\alpha np}.$$

1.3 Một số khái niệm và kết quả được sử dụng

1.3.1 Khái niệm với xác suất cao

Định nghĩa 1.7. *Ta nói một sự kiện (hay biến cố) xảy ra **với xác suất cao** (viết tắt là w.h.p) nếu xác suất xảy ra sự kiện đó phụ thuộc vào một chỉ số n, và tiến tới 1 nếu n tiến tới dương vô cùng.*

1.3.2 Thuật toán Dijkstra tìm đường đi ngắn nhất

Thuật toán Dijkstra, mang tên của nhà khoa học máy tính người Hà Lan Edsger Dijkstra, là một trong những thuật toán cổ điển để giải quyết bài toán tìm đường đi ngắn nhất từ một điểm cho trước tới tất cả các điểm còn lại trong một đồ thị có trọng số.

Thuật toán 1.1. *[5] Cho đồ thị tất định $G = (V, E)$ vô hướng, liên thông, có trọng không âm. Ý tưởng của thuật toán Dijkstra tìm đường đi ngắn nhất trong G xuất phát từ đỉnh gốc s như sau:*

B1. Từ đỉnh gốc s, gán khoảng cách tới chính nó là 0, gán khoảng cách nhỏ nhất ban đầu tới các đỉnh khác là ∞. Ta được danh sách khoảng cách tới các đỉnh.

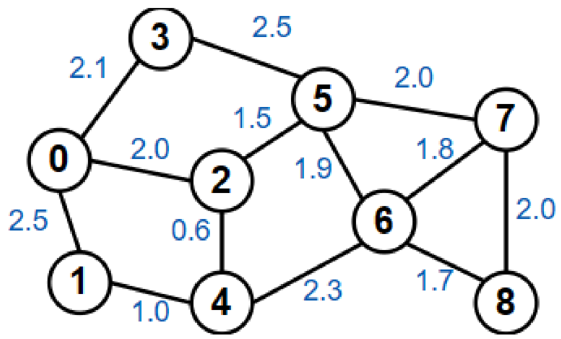
B2. Chọn đỉnh a có khoảng cách nhỏ nhất trong danh sách này và ghi nhận đây chính là khoảng cách nhỏ nhất từ đỉnh nguồn s tới a trong G. Các lần sau ta không xét tới đỉnh này nữa.

B3. Lần lượt xét các đỉnh b kề với a. Nếu khoảng cách từ đỉnh gốc tới đỉnh b nhỏ hơn khoảng cách hiện tại đang được ghi nhận thì cập nhật giá trị và đỉnh kề a vào khoảng cách hiện tại của b.

B4. Sau khi xét tất cả đỉnh kề b của đỉnh a. Lúc này ta được danh sách khoảng cách tới các điểm đã được cập nhật. Quay lại B2 với danh sách này. Thuật toán kết thúc khi chọn được khoảng cách nhỏ nhất (tức độ dài đường đi ngắn nhất) từ đỉnh nguồn s tới tất cả các đỉnh.

Để hiểu hơn về thuật toán Dijkstra nêu trên, ta cùng xem ví dụ sau đây.

Ví dụ 1.1. *Xét đồ thị vô hướng G:*



Hình 1.1: Đồ thị G.

Ta sẽ sử dụng thuật toán Dijkstra sẽ tìm khoảng cách từ đỉnh gốc 0 *tới tất cả các đỉnh còn lại trong đồ thị G như sau.*

- *Đầu tiên, ta khởi tạo khoảng cách từ đỉnh gốc* 0 *tới các đỉnh khác là* ∞*, khoảng cách tới chính nó bằng* 0*. Ta được danh sách khoảng cách từ đỉnh gốc tới các đỉnh:*

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$

- *Chọn đỉnh* 0 *có khoảng cách nhỏ nhất trong danh sách trên (in đậm), ghi nhận* $d_G(0,0) = 0$.

- *Xét các đỉnh kề với đỉnh* 0 *là đỉnh* $1, 2, 3$. *Với đỉnh* 1, *khoảng cách từ đỉnh gốc tới đỉnh* 1 *là* $2.5 < \infty$ *nên ta ghi nhận giá trị mới là* $(2.5, 1)$ *(nghĩa là khoảng cách tới đỉnh gốc là* 2.5, *đỉnh liền kề trước nó là đỉnh* 0*). Làm tương tự cho đỉnh* 2 *và đỉnh* 3. *Ta thu được danh sách mới khoảng cách từ đỉnh gốc tới các đỉnh:*

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$

- *Lúc này ta chọn đỉnh* 2 *có khoảng cách nhỏ nhất trong danh sách trên (in đậm), ghi nhận* $d_G(0,2) = 2.0$.
  Tiếp tục xét đỉnh kề với 2 *là đỉnh* 4 *và* 5. *Xét đỉnh* 4, *khoảng cách từ đỉnh gốc tới đỉnh* 4 *sẽ bằng khoảng cách từ đỉnh gốc tới đỉnh* 2 *cộng khoảng cách từ* 2 *tới* 4, *ghi nhận khoảng cách tại đỉnh* 4 *là* $(2.6, 2)$. *Làm tương tự cho đỉnh* 5. *Ta thu được danh sách mới khoảng cách từ đỉnh gốc tới các đỉnh:*

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	$-$	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$

- *Tiếp theo ta chọn đỉnh* 3 *có khoảng cách nhỏ nhất trong danh sách trên (in đậm), ghi nhận* $d_G(0,3) = 2.1$.
  Tiếp tục xét đỉnh kề với 3 *là đỉnh* 5. *Với đỉnh* 5, *khoảng cách từ đỉnh gốc tới đỉnh* 5 *mà đi qua đỉnh* 3 *liền kề ngay trước là* $2.1 + 2.5 = 4.6 > 3.5$ *nên giá trị tại đỉnh* 5 *không đổi.*

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	$-$	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	***(2.5, 0)***	$-$	$-$	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$

- *Đỉnh* 1 *là đỉnh tiếp theo được chọn (in đậm), ghi nhận* $d_G(0,1) = 2.5$.

Tiếp tục xét đỉnh kề với 1 là đỉnh 4. Với đỉnh 4, khoảng cách từ đỉnh gốc tới đỉnh 4 mà đi qua đỉnh 1 liền kề ngay trước là $2.5 + 1.0 = 3.5 > 2.6$ *nên giá trị tại đỉnh 4 không đổi.*

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	$(2.5, 0)$	–	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	***(2.5, 0)***	–	–	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	–	–	–	***(2.6, 2)***	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$

- *Xong đỉnh 1, ta chọn đỉnh 4 và ghi nhận* $d_G(0, 4) = 2.6$.
  Xét đỉnh kề với 4 là đỉnh 6: khoảng cách từ đỉnh gốc tới đỉnh 6 mà đi qua đỉnh 4 liền kề ngay trước là $2.6 + 2.3 = 4.9 < \infty$ *nên cập nhật giá trị tại đỉnh 6 là* $(4.9, 4)$.

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	$(2.5, 0)$	–	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	***(2.5, 0)***	–	–	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	–	–	–	***(2.6, 2)***	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	–	–	–	–	***(3.5, 2)***	$(4.9, 4)$	$(\infty, -)$	$(\infty, -)$

- *Tiếp tục ta chọn đỉnh 5 và ghi nhận* $d_G(0, 5) = 3.5$.
  Xét đỉnh kề với 5, có 4 đỉnh là $2, 3, 6$ *và* 7. *Tuy nhiên đỉnh* 2 *và đỉnh* 3 *ta đã ghi nhận kết quả và không xét tới nữa, do vậy ta chỉ quan tới việc có cập nhật giá trị cho đỉnh* 6 *và đỉnh* 7 *hay không. Với đỉnh* 6, *vì* $3.5 + 1.9 = 5.4 > 4.9$ *nên câu trả lời là không. Còn với đỉnh* 7, *vì* $3.5 + 2.0 = 5.5 < \infty$ *nên câu trả lời là có.*

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	$(2.5, 0)$	–	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	***(2.5, 0)***	–	–	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	–	–	–	***(2.6, 2)***	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	–	–	–	–	–	***(3.5, 2)***	$(4.9, 4)$	$(\infty, -)$	$(\infty, -)$
	–	–	–	–	–	–	***(4.9,4)***	$(5.5, 5)$	$(\infty, -)$

- *Chọn đỉnh 6 tiếp theo, và ghi nhận* $d_G(0, 6) = 4.9$.
  Xét đỉnh kề với 6, ta thấy ta sẽ không cập nhật giá trị cho đỉnh 7 nhưng cập nhật giá trị cho đỉnh 8 là $(6.6, 6)$.

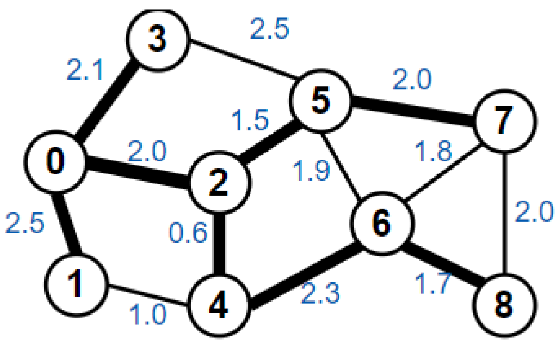
Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	$-$	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	***(2.5, 0)***	$-$	$-$	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	***(2.6, 2)***	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	$-$	***(3.5, 2)***	$(4.9, 4)$	$(\infty, -)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	$-$	$-$	***(4.9,4)***	$(5.5, 5)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	$-$	$-$	$-$	***(5.5, 5)***	$(6.6, 6)$

- *Chọn đỉnh 7, và ghi nhận* $d_G(0, 7) = 5.5$.

  Xét đỉnh kề với 7, ta thấy ta sẽ giữ nguyên giá trị cho đỉnh 8.

Đỉnh	0	1	2	3	4	5	6	7	8
K/c tới đỉnh gốc	***0***	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	***(2.0, 0)***	$(2.1, 0)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$(2.5, 0)$	$-$	***(2.1, 0)***	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	***(2.5, 0)***	$-$	$-$	$(2.6, 2)$	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	***(2.6, 2)***	$(3.5, 2)$	$(\infty, -)$	$(\infty, -)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	$-$	***(3.5, 2)***	$(4.9, 4)$	$(\infty, -)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	$-$	$-$	***(4.9,4)***	$(5.5, 5)$	$(\infty, -)$
	$-$	$-$	$-$	$-$	$-$	$-$	$-$	***(5.5, 5)***	$(6.6, 6)$
	$-$	$-$	$-$	$-$	$-$	$-$	$-$	$-$	***(6.6, 6)***

- *Thuật toán kết thúc khi đã chọn được khoảng cách nhỏ nhất (tức độ dài đường đi ngắn nhất) từ đỉnh gốc tới tất cả các đỉnh.*



Hình 1.2: Đường đi ngắn nhất từ đỉnh 0 tới tất cả các đỉnh trong G.

Nhận xét 1.4. *Bằng cách cập nhật cả giá trị và đỉnh kề vào danh sách tại B3, ta không chỉ biết được độ dài đường đi ngắn nhất từ đỉnh gốc tới các đỉnh, mà còn biết chính xác đường đi đó đi qua những đỉnh nào.*

1.3.3 Một kết quả của phân phối đều

Định lý 1.4. *[Bổ đề 9, [6]] Cho* $\gamma > 0; U_1, U_2, \ldots, U_k$ *i.i.d* $\sim U[0,1]$. *Khi đó với mọi* $u \geq 0$, *ta có:*

$$\mathbb{P}(U_1^\gamma + U_2^\gamma + \cdots + U_k^\gamma \leq u) \leq \frac{u^{k/\gamma}\left(\Gamma\left(\frac{1}{\gamma}+1\right)\right)^k}{\Gamma\left(\frac{k}{\gamma}+1\right)};$$

trong đó $\Gamma(.)$ *là hàm Gamma.*

Hệ quả 1.1. *Cho* $\gamma = 1$, *ta suy ra* $\mathbb{P}(U_1 + U_2 + \cdots + U_k \leq u) \leq \dfrac{u^k}{k!}$.

1.3.4 So sánh giữa phân phối mũ và phân phối đều

Nhận xét 1.5. *Xét* $X \sim Exp(1), Y \sim U[0,1]$, *khi đó* $Y \prec X$ *vì*

$$\forall x \geq 0 : \mathbb{P}(X \geq x) = e^{-x} \geq 1 - x = \mathbb{P}(Y \geq x).$$

Hơn nữa $\forall x \geq 0 :$

$$\mathbb{P}(X_1 + X_2 + \cdots + X_n \leq x) \leq \mathbb{P}(Y_1 + Y_2 + \cdots + Y_n \leq x),$$

với $X_1, X_2, \ldots, X_n$ *i.i.d* $\sim Exp(1)$ *và* $Y_1, Y_2, \ldots, Y_n$ *i.i.d* $\sim U[0,1]$.
Theo Hệ quả 1.1 ta suy ra:

$$\mathbb{P}(X_1 + X_2 + \cdots + X_n \leq x) \leq \frac{x^k}{k!}. \tag{1.3.1}$$

1.3.5 Xấp xỉ Stirling

Định lý 1.5. *Cho* n *là một số nguyên dương. Khi đó:*

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Hệ quả 1.2. $n! > \left(\frac{n}{e}\right)^n \ \forall n \in \mathbb{Z}_+$.

Chương 2

Bao của đồ thị có trọng ngẫu nhiên

Khái niệm đồ thị ngẫu nhiên - cụ thể là $G_{n,p}$[1] - lần đầu được giới thiệu bởi P. Erdös và A. Rényi vào năm 1959, sau đó phổ biến và phát triển nhiều mô hình khác nhau, xem thêm tại [7]. Một trong số đó là đồ thị có trọng ngẫu nhiên. Đồ thị có trọng ngẫu nhiên là đồ thị có tập đỉnh, tập cạnh cố định, nhưng độ dài các cạnh là các biến ngẫu nhiên. Thông thường chúng độc lập và cùng phân phối, chủ yếu là phân phối mũ hoặc phân phối đều.

Năm 1999, S. Janson đã chứng minh một kết quả về đường đi ngắn nhất trong đồ thị đầy đủ K_n[2] có trọng ngẫu nhiên $Exp(1)$ (xem [8]). Cụ thể ông chứng minh được với xác suất cao thì:

$$d_{K_n}(1,2) \approx \frac{\log n}{n}; \quad \max_{j>1} d_{K_n}(1,j) \approx \frac{2\log n}{n}; \quad \max_{i,j} d_{K_n}(i,j) \approx \frac{3\log n}{n}.$$

Tại chương này, ta nêu các kết quả về $t - bao$ trên lớp rộng hơn các đồ thị $\mathcal{G}(d)$ (xem Định nghĩa 2.1), theo một khía cạnh nào đó được hiểu như mở rộng các kết quả của Jason.

---

[1] $G_{n,p}$ là đồ thị ngẫu nhiên n đỉnh, và hai đỉnh bất kỳ được nối với nhau một cách độc lập với xác suất p.

[2] K_n là đồ thị đơn, vô hướng có n đỉnh, hai đỉnh bất kì đều được nối với nhau.

2.1 Kết quả chính

Cho trước số nguyên dương n đủ lớn, đặt $\theta = \dfrac{1}{\sqrt{\log n}}$.

Định nghĩa 2.1. *$\mathcal{G}(d)$ là lớp các đồ thị vô hướng, liên thông $G = ([n], E)$ với $[n] = \{1, 2, \ldots, n\}$, thỏa mãn hai tính chất chính quy dưới đây:*

(i) G là dn- chính quy:
Nghĩa là tồn tại $d = d(n)$ sao cho $1 \geq d \gg \theta \log(\log n)$, và

$$|\deg(i) - dn| \leq \frac{dn}{\sqrt{\log n}} \ \forall\, i \in [n]; \tag{2.1.1}$$

(ii) với mọi cặp hai tập con rời nhau $S, T \subset [n]$ với $|S|, |T| \geq \theta n$ ta có

$$\frac{|E(S,T)|\sqrt{\log n}}{|S||T| \log(\log n)} \to \infty, \tag{2.1.2}$$

ở đây $E(S,T) = \{e = \{u, v\} \in E \mid u \in S, v \in T\}$.

Nhận xét 2.1. *Đồ thị đầy đủ $K_n \in \mathcal{G}(1)$ và nếu $np \gg \log n$ thì với xác suất cao $G_{n,p} \in \mathcal{G}(p)$.*

Định lý 2.1. *[3] Xét đồ thị $G = ([n], E) \in \mathcal{G}(d)$ hoặc G là dn-chính quy với $d > 1/2$. Giả thiết rằng các cạnh $\{i, j\} \in E$ có độ dài $l_{i,j}$ là các biến ngẫu nhiên độc lập cùng phân phối $Exp(1)$. Khi đó với xác suất cao thì kích thước tối thiểu của $1 - bao$ là xấp xỉ $\frac{1}{2} n \log n$, tức là:*

$$\min\{|E'| : G' = (V, E') \text{ là } 1 - bao\} = \left(\frac{1}{2} + o(1)\right) n \log n.$$

2.2 Chứng minh Định lý 2.1

Trong suốt phần chứng minh này, ta gọi $(E_j)_{j\geq 1}$ là dãy các biến ngẫu nhiên độc lập, cùng phân phối $Exp(1)$. Gọi $(U_j)_{j\geq 1}$ là dãy các biến ngẫu nhiên độc lập, cùng phân phối $U[0, 1]$.

Trước hết, ta nhắc lại định nghĩa của $t - bao$ đã nêu ở phần giới thiệu:

Cho đồ thị liên thông, vô hướng $G = (V, E)$, và cố định $t \geq 1$. Đồ thị con $G' = (V, E') \subseteq G$ được gọi là một $t-$bao của G, nếu:

$$d_{G'}(u, v) \leq t.d_G(u, v), \quad \forall u, v \in G.$$

Để thuận tiện, đôi khi ta cũng coi chính $E' \subseteq E$ là $t - bao$. Với $t = 1$, rõ ràng $d_{G'}(u, v) = d_G(u, v), \quad \forall u, v \in G$.

2.2.1 Cận dưới kích thước tối thiểu của 1-*bao*

Để tìm được cận dưới kích thước tối thiểu của $1 - bao$ của G, ta thực hiện hai bước:

1. Chỉ ra rằng bất kỳ $1 - bao$ nào cũng phải chứa các tập X_v (xem định nghĩa tại Bổ đề 2.4) với "hầu hết" $v \in [n]$. Một cách ngắn gọn, X_v là tập các cạnh kề với v có trọng không quá lớn so với các cạnh khác cùng kề với v.

2. Ước lượng kích thước tối thiểu của $1 - bao$ qua việc đánh giá kích thước của các tập X_v.

Trước tiên, ta đi chứng minh kết luận ở bước đầu tiên qua các bổ đề sau.

Bổ đề 2.1. *Đặt $\alpha = 1 - 2\theta$. Cố định các đỉnh $v, w_1, w_2, \dots, w_l \in [n]$ với $l = O(\log n)$. Khi đó:*

$$\mathbb{P}\left(\exists\ 1 \leq i \leq l : d_G(v, w_i) \leq \frac{\alpha \log n}{dn}\right) = o(1). \qquad (2.2.3)$$

Chứng minh. Ta chỉ quan tâm tới trường hợp $\alpha > 0$.
Cố định $i \in \{1, 2, \dots, l\}$ bất kỳ, vì $\deg(u) \leq (1 + \theta)dn\ \forall u \in [n]$ (theo (2.1.1)) nên có tối đa $((1 + \theta)dn)^{k-1}$ đường đi k cạnh từ đỉnh v tới đỉnh w_i. Cùng với Nhận xét 1.5, ta thu được:

$$\mathbb{P}\left(d_G(v, w_i) \leq \frac{\alpha \log n}{dn}\right)$$

$$\leq \sum_{k=1}^{n-1} \mathbb{P}\left(d_G(v,w_i) \leq \frac{\alpha \log n}{dn} : \text{đường đi tối ưu có } k \text{ cạnh}\right)$$
$$\leq \sum_{k=1}^{n-1} \mathbb{P}\left(E_1 + E_2 + \cdots + E_k \leq \frac{\alpha \log n}{dn}\right) ((1+\theta)dn)^{k-1}$$
$$\leq \sum_{k=1}^{n-1} \mathbb{P}\left(U_1 + U_2 + \cdots + U_k \leq \frac{\alpha \log n}{dn}\right) ((1+\theta)dn)^{k-1}$$
$$\leq \sum_{k=1}^{n-1} \frac{(\alpha \log n)^k}{k!(dn)^k} ((1+\theta)dn)^{k-1}.$$

Do vậy:

$$\mathbb{P}\left(\exists\ 1 \leq i \leq l : d_G(v,w_i) \leq \frac{\alpha \log n}{dn}\right)$$
$$\leq \sum_{i=1}^{l} \mathbb{P}\left(d_G(v,w_i) \leq \frac{\alpha \log n}{dn}\right)$$
$$\leq l. \sum_{k=1}^{n-1} \frac{(\alpha \log n)^k}{k!(dn)^k} ((1+\theta)dn)^{k-1}$$
$$\leq \frac{l}{dn} \sum_{k=1}^{n-1} \frac{(1+\theta)^k (\alpha \log n)^k}{k!}.$$

Sử dụng Hệ quả 1.2 về xấp xỉ Stirling, ta thu được:

$$\mathbb{P}\left(\exists\ 1 \leq i \leq l : d_G(v,w_i) \leq \frac{\alpha \log n}{dn}\right)$$
$$\leq \frac{l}{dn} \sum_{k=1}^{n-1} \left(\frac{e(1+\theta)(\alpha \log n)}{k}\right)^k := S.$$

Tới đây, ta tách tổng S thành hai phần:

$$S = \frac{l}{dn}\left[\underbrace{\sum_{k=1}^{10\log n} \left(\frac{e(1+\theta)(\alpha \log n)}{k}\right)^k}_{S_1} + \underbrace{\sum_{k=10\log n}^{n-1} \left(\frac{e(1+\theta)(\alpha \log n)}{k}\right)^k}_{S_2}\right].$$

Xét $1 \leq k \leq 10 \log n$, đặt $a = e(1+\theta)\alpha \log n$, ta có:

$$\left(\frac{e(1+\theta)(\alpha \log n)}{k}\right)^k = \left(\frac{a}{k}\right)^k := f(k).$$

Ta có $f'(x) = \left(\frac{a}{x}\right)^x (\log a - 1 - \log x)$ nên ta dễ dàng suy ra được f đạt cực đại tại $x = a/e = (1+\theta)\alpha \log n$. Khi đó

$$\begin{aligned}\left(\frac{e(1+\theta)(\alpha \log n)}{k}\right)^k &= f(k) \leq f((1+\theta)\alpha \log n) \\ &= e^{(1+\theta)\alpha \log n} = n^{(1+\theta)\alpha},\end{aligned}$$

suy ra $S_1 \leq 10 \log n \cdot n^{(1+\theta)\alpha}$.

Với $k \geq 10 \log n$, sử dụng tính đơn điệu giảm của hàm f trên khoảng $(a/e, \infty)$, ta có:

$$\begin{aligned}\left(\frac{e(1+\theta)(\alpha \log n)}{k}\right)^k &= \left(\frac{e(1+\theta)(1-2\theta)\log n}{k}\right)^k \\ &\leq \left(\frac{e(1+\theta)(1-2\theta)\log n}{10 \log n}\right)^{10 \log n} \\ &= \left(\frac{e(1+\theta)(1-2\theta)}{10}\right)^{10 \log n} \\ &\leq \left(\frac{1}{e}\right)^{10 \log n} = \frac{1}{n^{10}},\end{aligned}$$

suy ra $S_2 \leq (n - 10 \log n) \cdot n^{-10} \leq n^{-9}$.

Như vậy:

$$\begin{aligned}\mathbb{P}\left(\exists\, 1 \leq i \leq l : d_G(v, w_i) \leq \frac{\alpha \log n}{dn}\right) &\leq S = \frac{l}{dn}(S_1 + S_2) \\ \leq \frac{l}{dn}\left(10 \log n \cdot n^{(1+\theta)\alpha} + n^{-9}\right) &= \frac{10 l \log n}{dn^{2\theta^2+\theta}} + \frac{l}{dn^{10}}.\end{aligned}$$

Chú ý rằng $\begin{cases} \theta = \frac{1}{\sqrt{\log n}} \Leftrightarrow n = e^{1/\theta^2} \Rightarrow n^{2\theta^2+\theta} = e^{2+\sqrt{\log n}} \\ l = O(\log n) \\ \frac{d\sqrt{\log n}}{\log(\log n)} = \frac{d}{\theta \log(\log n)} \to \infty \end{cases}$, ta thu được kết luận:

$$
\begin{aligned}
\mathbb{P}\left(\exists\, 1 \le i \le l : d_G(v, w_i) \le \frac{\alpha \log n}{dn}\right) &\le \frac{10 l \log n}{dn^{2\theta^2+\theta}} + \frac{l}{dn^{10}} \\
&\le \frac{10}{e^2} \cdot \frac{l \log^{3/2} n \log(\log n)}{\frac{d\sqrt{\log n}}{\log(\log n)} e^{\sqrt{\log n}}} + \frac{l}{dn^{10}} \\
&= o(1).
\end{aligned}
$$

□

Bổ đề 2.2. *Với mỗi $v \in [n]$, ta định nghĩa*

$$A_v = \left\{ w \neq v : w \sim v;\ l_{v,w} \le \frac{\log n}{dn} \right\}.$$

Khi đó với xác suất cao thì $|A_v| \le 4 \log n\ \forall v$.

Chứng minh. Xét $v \in [n]$ cố định bất kỳ. Ta có:

- Theo (2.1.1) : $|A_v| \le \deg(v) \le (1+\theta)dn$.
- Với $w \sim v$ thì

$$
\begin{aligned}
\mathbb{P}(w \in A_v) = \mathbb{P}\left(l_{v,w} \le \frac{\log n}{dn}\right) &= \mathbb{P}\left(E_1 \le \frac{\log n}{dn}\right) \\
&\le \mathbb{P}\left(U_1 \le \frac{\log n}{dn}\right) = \frac{\log n}{dn}.
\end{aligned}
$$

Suy ra $|A_v| \prec Bin\left((1+\theta)dn, \frac{\log n}{dn}\right)$. Sử dụng thêm Bất đẳng thức Chernoff 1.3, ta thu được:

$$
\begin{aligned}
\mathbb{P}(|A_v| \ge 4 \log n) &\le \mathbb{P}\left(Bin((1+\theta)dn, \frac{\log n}{dn}) \ge 4 \log n\right) \\
&\le \underbrace{\left(\frac{e(1+\theta)}{4}\right)^{4 \log n} \le \left(\frac{1}{e^{1,1}}\right)^{\log n}}_{\text{khi } n \text{ đủ lớn}} \le \frac{1}{n^{1,1}}.
\end{aligned}
$$

Do đó:

$$\begin{aligned}\mathbb{P}(\exists v : |A_v| \geq 4\log n) &= \mathbb{P}\left(\bigcup_{v\in[n]}(|A_v| \geq 4\log n)\right)\\ &\leq \sum_{v\in[n]} \mathbb{P}(|A_v| \geq 4\log n)\\ &\leq n.\frac{1}{n^{1,1}} = \frac{1}{n^{0,1}}.\end{aligned}$$

Từ đây ta có thể thu được kết luận của Bổ đề. □

Với mỗi $v \in [n]$, ta kí hiệu $\delta_v = \min\{l_{w,v} : w \sim v\}$ là khoảng cách từ v đến đỉnh kề gần nhất.

Bổ đề 2.3. *Đặt*

$$B = \left\{v \in [n] : \delta_v \geq \frac{\sqrt{\log n}}{dn}\right\}.$$

Khi đó với xác suất cao thì $|B| \leq n\exp(-\sqrt[3]{\log n})$.

Chứng minh. Xét $v \in [n]$, theo (2.1.1) thì $\deg(v) \geq (1-\theta)dn$. Ta có:

$$\begin{aligned}\mathbb{P}(v \in B) &= \mathbb{P}\left(\delta_v \geq \frac{\sqrt{\log n}}{dn}\right) = \mathbb{P}\left(l_{w,v} \geq \frac{\sqrt{\log n}}{dn} \quad \forall w \sim v\right)\\ &= \prod_{w\sim v}\mathbb{P}\left(l_{w,v} \geq \frac{\sqrt{\log n}}{dn}\right) = \prod_{w\sim v}\mathbb{P}\left(E_1 \geq \frac{\sqrt{\log n}}{dn}\right)\\ &= \left(\exp\left(-\frac{\sqrt{\log n}}{dn}\right)\right)^{\deg(v)} = \exp\left(-\frac{\sqrt{\log n}}{dn}\cdot\deg(v)\right)\\ &\leq \exp\left(-\frac{\sqrt{\log n}}{dn}\cdot(1-\theta)dn\right).\end{aligned}$$

Suy ra $\mathbb{E}(|B|) \leq n\exp\left(-(1-\theta)\sqrt{\log n}\right)$.

Áp dụng Bất đẳng thức Markov 1.1 ta thu được:

$$\begin{aligned}\mathbb{P}\left(|B| \geq n\exp(-\sqrt[3]{\log n})\right) &\leq \frac{\mathbb{E}(|B|)}{n\exp(-\sqrt[3]{\log n})}\\ &\leq \underbrace{\exp(\sqrt[3]{\log n} - (1-\theta)\sqrt{\log n})}_{\to 0 \text{ khi } n\to\infty}.\end{aligned}$$

Vậy ta có điều phải chứng minh. □

Bổ đề 2.4. *Với mỗi $v \in [n]$, ta định nghĩa*

$$X_v = \left\{ e = \{v, x\} \in E : l(e) \le \delta_v + \frac{\alpha \log n}{dn} \right\}.$$

Gọi $G_S = ([n], S)$ là $1-bao$ bất kỳ của G. Khi đó với xác suất cao thì $X_v \subseteq S$ $\forall v$ ngoại trừ $o(n)$ đỉnh.

Chứng minh. Ta sẽ chứng minh $w.h.p$ $X_v \subseteq S$ với hầu hết $v \notin B$.
Ta có:

$$\mathbb{P}(X_v \nsubseteq S \mid v \notin B) = \mathbb{P}\left(\mathcal{E} \mid v \notin B\right),$$

trong đó $\mathcal{E} = \{\exists\ e = \{v, x\} \in X_v\}$.
Lại có G_S là $1-$*bao* của G, nên nếu ta xét đường đi tối ưu từ x tới v trong G_S thì:

$$\begin{aligned}
\mathbb{P}\left(\mathcal{E} \mid v \notin B\right) &\le \mathbb{P}(\mathcal{E}, \exists\ w \sim v : d_{G_S}(x, v) = d_{G_S \setminus \{v\}}(x, w) + l_{w,v} \mid v \notin B) \\
&= \mathbb{P}(\mathcal{E}, \exists\ w \sim v : d_G(x, v) = d_{G_S \setminus \{v\}}(x, w) + l_{v,w} \mid v \notin B) \\
&:= p_1 + p_2,
\end{aligned}$$

trong đó:

$$\begin{aligned}
p_1 &= \mathbb{P}(\mathcal{E}, \exists\ w \sim v, w \in A_v : d_G(x, v) = d_{G_S \setminus \{v\}}(x, w) + l_{v,w} \mid v \notin B), \\
p_2 &= \mathbb{P}(\mathcal{E}, \exists\ w \sim v, w \notin A_v : d_G(x, v) = d_{G_S \setminus \{v\}}(x, w) + l_{v,w} \mid v \notin B).
\end{aligned}$$

Ta sẽ đi ước lượng p_1, p_2.

- Tính p_2.
  Khi $v \notin B$ thì $\delta_v < \frac{\sqrt{\log n}}{dn}$, kết hợp với việc $\alpha = 1 - 2\theta = 1 - \frac{2}{\sqrt{\log n}}$ ta có

$$\begin{aligned}
\delta_v + \frac{\alpha \log n}{dn} &< \frac{\sqrt{\log n} + \alpha \log n}{dn} \\
&= \frac{\log n - \sqrt{\log n}}{dn} < \frac{\log n}{dn},
\end{aligned}$$

lại có $\{v, x\} \in X_v$ suy ra

$$l_{x,v} \leq \delta_v + \frac{\alpha \log n}{dn},$$

mà $w \notin A_v$ nên

$$l_{v,w} > \frac{\log n}{dn},$$

do đó

$$d_G(x, v) \leq l_{x,v} \leq \delta_v + \frac{\alpha \log n}{dn} < \frac{\log n}{dn} < l_{v,w}.$$

Ta suy ra sự kiện trong p_2 không thể xảy ra vì:

$$l_{v,w} > d_G(x, v) = d_{G_S \setminus \{v\}}(x, w) + l_{v,w} \geq l_{v,w}.$$

Vậy $p_2 = 0$.

- Tính p_1.

  Sử dụng các định nghĩa của δ_v, A_v trong Bổ đề 2.2, và các tính toán khi tính p_2, ta có:

$$\begin{aligned}
p_1 &= \mathbb{P}(\mathcal{E}, \exists\, w \sim v, w \in A_v : d_G(x, v) = d_{G_S \setminus \{v\}}(x, w) + l_{v,w} \mid v \notin B) \\
&= \mathbb{P}(\mathcal{E}, \exists\, w \in A_v : d_{G_S \setminus \{v\}}(x, w) + l_{v,w} = d_G(x, v) \mid v \notin B) \\
&\leq \mathbb{P}\left(\mathcal{E}, \exists\, w \in A_v : d_{G_S \setminus \{v\}}(x, w) + l_{v,w} \leq \delta_v + \frac{\alpha \log n}{dn} \mid v \notin B\right) \\
&\leq \mathbb{P}\left(\mathcal{E}, \exists\, w \in A_v : d_{G_S \setminus \{v\}}(x, w) \leq \frac{\alpha \log n}{dn} \mid v \notin B\right) \\
&\leq \sum_{w \sim v} \mathbb{P}\left(\mathcal{E}, w \in A_v : d_{G_S \setminus \{v\}}(x, w) \leq \frac{\alpha \log n}{dn} \mid v \notin B\right) \\
&:= p_{1,1} + p_{1,2},
\end{aligned}$$

  trong đó:

$$p_{1,1} = \sum_{w \sim v} \mathbb{P}\left(\mathcal{E}, w \in A_v, |A_v| \leq 4 \log n : d_{G_S \setminus \{v\}}(x, w) \leq \frac{\alpha \log n}{dn} \mid v \notin B\right),$$

$$p_{1,2} = \sum_{w \sim v} \mathbb{P}\left(\mathcal{E}, w \in A_v, |A_v| > 4 \log n : d_{G_S \setminus \{v\}}(x, w) \leq \frac{\alpha \log n}{dn} \mid v \notin B\right).$$

Đến đây, để ý rằng $\deg v \leq (1+\theta)dn$, đồng thời sử dụng kết quả tại Bổ đề 2.1, ta thu được:

$$
\begin{aligned}
&p_{1,1} \\
&= \sum_{w\sim v} \mathbb{P}\left(\mathcal{E}, w \in A_v, |A_v| \leq 4\log n : d_{G_S\setminus\{v\}}(x,w) \leq \frac{\alpha \log n}{dn} \mid v \notin B\right) \\
&\leq \sum_{w\sim v} \mathbb{P}\left(\mathcal{E}, d_{G_S\setminus\{v\}}(x,w) \leq \frac{\alpha \log n}{dn} \mid w \in A_v, |A_v| \leq 4\log n\right) \\
&\qquad \times \mathbb{P}(w \in A_v, |A_v| \leq 4\log n \mid v \notin B) \\
&\leq \sum_{w\sim v} \mathbb{P}\left(\mathcal{E}, d_{G_S\setminus\{v\}}(x,w) \leq \frac{\alpha \log n}{dn} \mid w \in A_v, |A_v| \leq 4\log n\right) \\
&\qquad \times \mathbb{P}(w \in A_v) \\
&\leq (1+\theta)dn \cdot \left[\frac{10}{d}\left(\frac{\log n}{\exp(2+\sqrt{\log n})} + \frac{1}{10n^{10}}\right) \cdot \frac{\log n}{dn}\right] \\
&\leq (1 + \frac{1}{\sqrt{\log n}}) \cdot \frac{10}{d\sqrt{\log n}} \left(\frac{\log^{3/2} n}{\exp(2+\sqrt{\log n})} + \frac{\log^{1/2} n}{10n^{10}}\right) \to 0.
\end{aligned}
$$

Mặt khác, áp dụng các kết quả của Bổ đề 2.2 và Bổ đề 2.3, ta cũng có:

$$
\begin{aligned}
p_{1,2} &\leq \sum_{w\sim v} \mathbb{P}(|A_v| > 4\log n \mid v \notin B) \\
&\leq \sum_{w\sim v} \frac{\mathbb{P}(|A_v| > 4\log n)}{\mathbb{P}(v \notin B)} \\
&\leq \deg(v) \cdot \frac{\mathbb{P}(|A_v| > 4\log n)}{\mathbb{P}(v \notin B)} \\
&\leq (1+\theta)dn \cdot \frac{1}{n^{1,1}(1-\exp\left(-(1-\theta)\sqrt{\log n}\right))} \to 0.
\end{aligned}
$$

Như vậy, $p_1 = p_{1,1} + p_{1,2} \to 0$ khi $n \to \infty$.

Tóm lại

$$\mathbb{P}(X_v \nsubseteq S \mid v \notin B) \leq p_1 + p_2 \to 0 \text{ khi } n \to \infty,$$

suy ra:

$$\mathbb{P}(v \notin B, X_v \nsubseteq S) \leq \mathbb{P}(X_v \nsubseteq S \mid v \notin B) \to 0 \text{ khi } n \to \infty.$$

Khi đó đặt $C = \{v \notin B, X_v \nsubseteq S\}$ thì $\begin{cases} \mathbb{E}(|C|) = o(n) \\ X_v \subseteq S \ \forall v \in [n] \setminus (B \cup C) \end{cases}$

đồng thời ta đã biết $\mathbb{E}(|B|) \leq n.\exp\left(-(1-\theta)\sqrt{\log n}\right) = o(n)$.

Vậy ta kết luận được với xác suất cao thì $X_v \subseteq S \ \forall v$ ngoại trừ $o(n)$ đỉnh. □

Tới đây, ta đã chứng minh xong khẳng định ở bước đầu tiên. Tiếp theo, ta đi ước lượng $\sum_{v\in[n]\setminus(B\cup C)} |X_v|$. Ta có:

$$|X_v| = \sum_{w\sim v} \mathbf{1}_{\left\{l_{w,v} \leq \delta_v + \frac{\alpha \log n}{dn}\right\}} \geq \sum_{w\sim v} \mathbf{1}_{\left\{l_{w,v} \leq \frac{\alpha \log n}{dn}\right\}}.$$

Đặt $Y_v = \sum_{w\sim v} \mathbf{1}_{\left\{l_{w,v} \leq \frac{\alpha \log n}{dn}\right\}} \Rightarrow |X_v| \geq Y_v$ $h.c.c$ nên

$$\sum_{v\in[n]\setminus(B\cup C)} |X_v| \geq \sum_{v\in[n]\setminus(B\cup C)} Y_v \quad h.c.c\ . \tag{2.2.4}$$

Ta đi ước lượng $\sum_{v\in[n]\setminus(B\cup C)} |X_v|$ qua $\sum_{v\in[n]\setminus(B\cup C)} Y_v$ bằng hai bước.

Bước 1: Ta có

$$\begin{aligned}\mathbb{E}(Y_v) = \mathbb{E}\left(\sum_{w\sim v} \mathbf{1}_{\left\{l_{w,v} \leq \frac{\alpha \log n}{dn}\right\}}\right) &= \sum_{w\sim v} \mathbb{P}\left(l_{w,v} \leq \frac{\alpha \log n}{dn}\right) \\ &= \deg(v) \cdot \left(1 - \exp\left(\frac{-\alpha \log n}{dn}\right)\right);\end{aligned}$$

mà $\deg(v) \geq (1-\theta)dn \ \forall v$ và $1 - e^{-t} \geq t(1 - t/2) \ \forall t$ nên ta suy ra

$$\begin{aligned}\mathbb{E}\left(\sum_{v\in[n]} Y_v\right) &= \sum_{v\in[n]} \deg(v) \cdot \left(1 - \exp\left(\frac{-\alpha \log n}{dn}\right)\right) \\ &\geq n \cdot (1-\theta)dn \cdot \frac{\alpha \log n}{dn}\left(1 - \frac{\alpha \log n}{2dn}\right).\end{aligned} \tag{2.2.5}$$

Từ định nghĩa của Y_v ta thấy nếu $i \nsim j$ thì Y_i độc lập với Y_j, nên:

$$\operatorname{Var}\left(\sum_{v\in[n]} Y_v\right) = \sum_{v,w\in[n]} \operatorname{Cov}(Y_v, Y_w) = \sum_{v\in[n]}\sum_{w\sim v} \operatorname{Cov}(Y_v, Y_w).$$

Mặt khác nếu $w \sim v$ thì

$$\begin{aligned}
\operatorname{Cov}(Y_v, Y_w) &= \operatorname{Cov}\left(\sum_{x\sim v} \mathbf{1}_{\left\{l_{x,v}\le\frac{\alpha\log n}{dn}\right\}}; \sum_{y\sim w} \mathbf{1}_{\left\{l_{y,w}\le\frac{\alpha\log n}{dn}\right\}}\right)\\
&= \sum_{x\sim v}\sum_{y\sim w} \underbrace{\operatorname{Cov}\left(\mathbf{1}_{\left\{l_{x,v}\le\frac{\alpha\log n}{dn}\right\}}; \mathbf{1}_{\left\{l_{y,w}\le\frac{\alpha\log n}{dn}\right\}}\right)}_{\begin{cases}\neq 0 \text{ khi } x = w \text{ và } y = v\\ = 0 \text{ trong mọi TH còn lại}\end{cases}}\\
&= \operatorname{Cov}\left(\mathbf{1}_{\left\{l_{w,v}\le\frac{\alpha\log n}{dn}\right\}}; \mathbf{1}_{\left\{l_{v,w}\le\frac{\alpha\log n}{dn}\right\}}\right)\\
&= \operatorname{Var}\left(\mathbf{1}_{\left\{l_{w,v}\le\frac{\alpha\log n}{dn}\right\}}\right)\\
&= \mathbb{E}\left(\mathbf{1}^2_{\left\{l_{w,v}\le\frac{\alpha\log n}{dn}\right\}}\right) - \left(\mathbb{E}\left(\mathbf{1}_{\left\{l_{w,v}\le\frac{\alpha\log n}{dn}\right\}}\right)\right)^2\\
&= \mathbb{E}\left(\mathbf{1}_{\left\{l_{w,v}\le\frac{\alpha\log n}{dn}\right\}}\right) - \left(\mathbb{E}\left(\mathbf{1}_{\left\{l_{w,v}\le\frac{\alpha\log n}{dn}\right\}}\right)\right)^2\\
&= \exp\left(-\frac{\alpha\log n}{dn}\right)\cdot\left(1 - \exp\left(-\frac{\alpha\log n}{dn}\right)\right)\\
&\le 1\cdot\frac{\alpha\log n}{dn} = \frac{\alpha\log n}{dn}.
\end{aligned}$$

Ta suy ra

$$\begin{aligned}
\operatorname{Var}\left(\sum_{v\in[n]} Y_v\right) &= \sum_{v\in[n]}\sum_{w\sim v} \operatorname{Cov}(Y_v, Y_w)\\
&\le \sum_{v\in[n]} \deg(v)\cdot\frac{\alpha\log n}{dn}\\
&\le n\cdot(1+\theta)dn\cdot\frac{\alpha\log n}{dn} = n(1+\theta)\alpha\log n. \qquad (2.2.6)
\end{aligned}$$

Áp dụng Bất đẳng thức Chebyshev 1.2 với $u > 0$:

$$\mathbb{P}\left[\sum_{v\in[n]} Y_v - \mathbb{E}\left(\sum_{v\in[n]} Y_v\right) \le -u\right] \le \mathbb{P}\left[\left|\sum_{v\in[n]} Y_v - \mathbb{E}\left(\sum_{v\in[n]} Y_v\right)\right| \ge u\right] \le \frac{\mathrm{Var}\left(\sum_{v\in[n]} Y_v\right)}{u^2}.$$

Sử dụng (2.2.5), (2.2.6), và từ định nghĩa của các tham số θ, α, d ta có

$$\mathbb{P}\left[\sum_{v\in[n]} Y_v \le \underbrace{n(1-\theta)\alpha\log n\left(1 - \frac{\alpha\log n}{dn}\right)}_{\approx\, n\log n} - u\right] \le \underbrace{n(1+\theta)\alpha\log n}_{\approx\, n\log n} \cdot \frac{1}{u^2}.$$

Chọn $\sqrt{n\log n} \ll u \ll n\log n$ ta thu được

$$\mathbb{P}\left[\sum_{v\in[n]} Y_v \ge [1 - o(1)]\, n\log n\right] \to 1 \text{ khi } n \to \infty. \tag{2.2.7}$$

<u>Bước 2:</u> Ta có $\sum_{v\in(B\cup C)} Y_v = \sum_{v\in[n]}\left(Y_v \mathbf{1}_{\{v\in(B\cup C)\}}\right)$, và

$$\mathbb{E}\left(Y_v \mathbf{1}_{\{v\in(B\cup C)\}}\right) \le \sqrt{\mathbb{E}\left(Y_v^2\right)\mathbb{E}\left(\mathbf{1}^2_{\{v\in(B\cup C)\}}\right)}. \tag{2.2.8}$$

Ta cũng có

- $\mathbb{E}\left(Y_v^2\right) = \mathrm{Var}(Y_v) + \left(\mathbb{E}(Y_v)\right)^2$, trong đó:

  (i). $\mathrm{Var}(Y_v) \le \mathbb{E}(Y_v)$ vì Y_v là tổng các biến ngẫu nhiên chỉ thị độc lập.

  (ii). $\mathbb{E}(Y_v) = \deg(v)\left(1 - \exp\left(\frac{-\alpha\log n}{dn}\right)\right)$: chứng minh tại bước 1.

  (iii). $\deg(v) \le (1+\theta)dn$ và $\left[1 - \exp\left(\frac{-\alpha\log n}{dn}\right)\right] \le \frac{\alpha\log n}{dn}$.

  Ta suy ra:

$$\begin{aligned}\mathbb{E}\left(Y_v^2\right) &\le \mathbb{E}(Y_v).\left(1 + \mathbb{E}(Y_v)\right)\\ &\le \left[(1+\theta)\alpha\log n\right]\left[1 + (1+\theta)\alpha\log n\right]\\ &\le \left(2\alpha\log n\right)^2.\end{aligned} \tag{2.2.9}$$

- $\mathbb{E}\left(\mathbf{1}^2_{\{v\in(B\cup C)\}}\right) = \mathbb{E}\left(\mathbf{1}_{\{v\in(B\cup C)\}}\right) = \mathbb{P}(v \in (B \cup C))$, trong đó:
  (i). $B \cap C = \emptyset$.
  (ii). $\mathbb{P}(v \in B) \leq \exp\left[-(1-\theta)\sqrt{\log n}\right]$: chứng minh tại Bổ đề 2.3.
  (iii). $\mathbb{P}(v \in C) = o(1)$: chứng minh tại Bổ đề 2.4.

  Ta suy ra:

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}^2_{\{v\in(B\cup C)\}}\right) &\leq \exp\left[-(1-\theta)\sqrt{\log n}\right] + o(1)\\ &= o(1). \end{aligned} \tag{2.2.10}$$

Thế (2.2.9) và (2.2.10) vào (2.2.8) ta thu được:

$$\mathbb{E}\left(Y_v \mathbf{1}_{\{v\in(B\cup C)\}}\right) = o(\log n).$$

Suy ra:

$$\begin{aligned}\mathbb{E}\left(\sum_{v\in(B\cup C)} Y_v\right) &= \mathbb{E}\left(\sum_{v\in[n]}\left(Y_v \mathbf{1}_{\{v\in(B\cup C)\}}\right)\right)\\ &= \sum_{v\in[n]} \mathbb{E}\left(Y_v \mathbf{1}_{\{v\in(B\cup C)\}}\right) = o(n\log n).\end{aligned}$$

Khi đó, theo Bất đẳng thức Markov 1.1, ta có:

$$\mathbb{P}\left(\sum_{v\in(B\cup C)} Y_v \leq o(n\log n)\right) \to 1 \text{ khi } n \to \infty. \tag{2.2.11}$$

Vậy ta kết thúc hai bước để ước lượng $\sum_{v\in[n]\setminus(B\cup C)} Y_v$. Đến đây áp dụng (2.2.7) và (2.2.11) vào so sánh (2.2.4) ta thu được:

$$\sum_{v\in[n]\setminus(B\cup C)} |X_v| \geq \sum_{v\in[n]\setminus(B\cup C)} Y_v \geq (1-o(1))n\log n \quad w.h.p\ .$$

Mặt khác, rõ ràng $\forall v \neq w : |X_v \cap X_w| \leq 1$ nên ta có thể kết luận: với $G_S = ([n], S)$ là $1-bao$ bất kỳ của G thì

$$|S| \geq \left|\bigcup_{v\in[n]\setminus(B\cup C)} X_v\right| \geq \frac{1}{2}\sum_{v\in[n]\setminus(B\cup C)} |X_v| \geq \left(\frac{1}{2}+o(1)\right) n\log n \quad w.h.p\ .$$

Vậy với xác suất cao thì kích thước tối thiểu của $1 - bao$ của G là $\left(\frac{1}{2} + o(1)\right) n \log n$.

2.2.2 Cận trên kích thước tối thiểu của 1-*bao*

Để tìm được cận trên kích thước tối thiểu của $1 - bao$, trước tiên ta xây dựng $1 - bao$ của G qua bổ đề sau:

Bổ đề 2.5. *Đặt* $l_0 = \frac{(1+\sqrt{\theta})\log n}{dn}$, *và*

$$E' = E_0 \cup E'_0$$

với

$$\begin{aligned}
E_0 &= \{e \in E : l(e) \leq l_0\}, \\
E'_0 &= \{e = \{u, v\} \in E : l_0 < l(e) = d_G(u, v)\}.
\end{aligned}$$

Khi đó $G' = ([n], E')$ *là* $1 - bao$ *của* G.

Chứng minh. Với $u, v \in [n]$ bất kỳ, xét đường đi ngắn nhất từ $u \to v$ trong G là $u = w_0 \to w_1 \to w_2 \to \cdots \to w_k = v,\ k \geq 1, w_j \sim w_{j+1}$.
Khi đó:

$$d_G(u, v) = \sum_{j=0}^{k-1} l_{w_j, w_{j+1}}.$$

Đặt $e_j = \{w_j, w_{j+1}\} \in E$, ta có các trường hợp:

- Nếu $l_{w_j, w_{j+1}} \leq l_0$ thì $e_j \in E_0 \subseteq E'$.
- Nếu $l_{w_j, w_{j+1}} > l_0$ thì do e_j nằm trong đường đi ngắn nhất từ $u \to v$ nên e_j phải là đường đi ngắn nhất từ $w_j \to w_{j+1}$, suy ra $e_j \in E'_0$.

Tóm lại, $e_j \in E'\ \forall j$ nên $u = w_0 \to w_1 \to w_2 \to \cdots \to w_k = v$ cũng là một đường đi từ $u \to v$ trong $G' \subseteq G$. Do đó:

$$d_G(u, v) = \sum_{j=0}^{k-1} l_{w_j, w_{j+1}} \geq d_{G'}(u, v) \geq d_G(u, v),$$

hay

$$d_{G'}(u,v) = d_G(u,v) \quad \forall u, v \in [n].$$

Ta kết luận G' là $1 - \textit{bao}$ của G. $\square$

Bây giờ ta sẽ ước lượng kích thước của G', tức cỡ của $|E'|$ hay $|E_0| + |E'_0|$.

Bổ đề 2.6. *Với xác suất cao thì:*

$$|E_0| \approx \frac{1}{2} n \log n.$$

Chứng minh. Ta có:

$$\begin{aligned}\mathbb{P}(e \in E_0) = \mathbb{P}(l(e) \leq l_0) &= \mathbb{P}\left(E_1 \leq \frac{(1+\sqrt{\theta})\log n}{dn}\right)\\ &= 1 - \exp\left(-\frac{(1+\sqrt{\theta})\log n}{dn}\right),\end{aligned}$$

suy ra

$$|E_0| \sim Bin\left[|E|, \underbrace{1 - \exp\left(-\frac{(1+\sqrt{\theta})\log n}{dn}\right)}_{t}\right].$$

Áp dụng Bất đẳng thức Chernoff 1.3 ta thu được:

$$\mathbb{P}\left((1-\varepsilon)|E|t \leq |E_0| \leq (1+\varepsilon)|E|t\right) \to 1$$

Mặt khác, vì $\forall v : (1-\theta)dn \leq \deg(v) \leq (1+\theta)dn$, nên:

$$\frac{(1-\theta)dn^2}{2} \leq |E| \leq \frac{(1+\theta)dn^2}{2},$$

và cũng vì $x - x^2/2 \leq 1 - e^{-x} = t \leq x$ nên không khó để rút ra được:

$$|E|.t \approx \frac{n \log n}{2}.$$

Kết hợp những điều bên trên, ta có điều phải chứng minh. $\square$

Tiếp theo, ta tìm cách ước lượng $|E_0'|$. Chú ý rằng:

$$E_0' = \{e = \{u, v\} \in E : l_0 < l(e) = d_G(u, v)\}.$$

Để làm được điều này, ta sẽ tìm chặn trên l_1 cho $\max_{i,j} d_G(i, j)$, rồi từ đó ước lượng $\mathbb{P}(e \in E_0' \mid \max_{i,j} d_G(i, j) \leq l_1)$.
Cụ thể, đặt

$$l_1 = \frac{5 \log n}{dn}.$$

Bổ đề 2.7. *Với xác suất cao thì:*

$$\max_{i,j} d_G(i, j) \leq l_1.$$

Chứng minh. Ta xây dựng thuật toán tìm đường đi ngắn nhất trong G. Trong [8], Jason đã phân tích hiệu suất hoạt động của Thuật toán Dijkstra [5] trên đồ thị đầy đủ K_n với độ dài cạnh là biến ngẫu nhiên phân phối mũ. Về cơ bản, thuật toán của chúng ta sẽ điều chỉnh một vài chi tiết nhỏ trong lập luận của ông để phù hợp với đồ thị G thỏa mãn hai điều kiện 2.1.1 và 2.1.2.
Cụ thể, ta phân tích thuật toán Dijkstra để tìm đường đi ngắn nhất từ đỉnh nguồn 1 tới các đỉnh khác, mà ở đó độ dài các cạnh là các biến ngẫu nhiên độc lập, cùng phân phối mũ $Exp(1)$. Nhắc lại rằng sau bước i, ta có tập các đỉnh là cây T_i và tập các giá trị $d_v, v \in [n]$ thỏa mãn:

- Với $u \in T_i : d_u = d_G(1, u)$.
- Với $v \in \bar{T}_i = [n] \setminus T_i : d_v = \min\{d_u + l_{u,v}, u \sim v, u \in T_i\}$.
- Đặt $\delta_i = \max\{d_u, u \in T_i\}$ thì theo cách thuật toán vận hành, ta có ràng buộc $d_v \geq \delta_i \ \forall v \in \bar{T}_i$. Khi đó:

  $$d_u + l_{u,v} \geq \delta_i \Leftrightarrow l_{u,v} \geq \delta_i - d_u \quad \forall \{u, v\} \in E(T_i, \bar{T}_i).$$

  Mặt khác, do phân phối mũ có tính chất không nhớ *(memory less)* nên tồn tại $E_{u,v} \sim Exp(1)$ (độc lập với tất cả các $E(u', v')$ khác) thỏa mãn

$$l_{u,v} = \delta_i - d_u + E_{u,v} \Leftrightarrow d_u + l_{u,v} = \delta_i + E_{u,v} \quad \forall\{u,v\} \in E(T_i, \bar{T}_i).$$

Ta suy ra thuật toán này tương đương phân phối, tức bảo toàn phân phối độ dài đường đi ngắn nhất từ đỉnh nguồn 1 tới các đỉnh khác, với quá trình rời rạc sau.

Quá trình 1:

- Đặt $v_1 = 1, T_1 = 1, d_{v_1} = 0, \delta_1 = \max\{d_u, u \in T_1\}$.
- Sau khi có T_i, δ_i, với mỗi cạnh $\{u,v\} \in E(T_i, \bar{T}_i)$ ta đặt tương ứng độ dài là $E_{u,v}$, các độ dài độc lập, cùng phân phối $Exp(1)$. Xét $\{\hat{u}, \hat{v}\} = \arg\min_{\{u,v\}\in E(T_i,\bar{T}_i)}\{\delta_i + E_{u,v}\}$, khi đó ta đặt $v_{i+1} = \hat{v}, T_{i+1} = T_i \cup \{v_{i+1}\}, d_{v_{i+1}} = \delta_i + E_{\hat{u},\hat{v}}$.

Để ý rằng, $\min\{E_1, E_2, \ldots, E_r\} \sim Exp(r)$, ta suy ra quá trình 1 tương đương quá trình 2.

Quá trình 2:

- Đặt $v_1 = 1, T_1 = 1, d_{v_1} = 0$.
- Sau khi có v_i, T_i, chọn ngẫu nhiên theo phân phối đều cạnh $\{\hat{u}, \hat{v}\} \in E(T_i, \bar{T}_i)$. Ta đặt $v_{i+1} = \hat{v}, T_{i+1} = T_i \cup \{v_{i+1}\}, d_{v_{i+1}} = d_{v_i} + E^{\gamma_i}$, với E^{γ_i} là một biến ngẫu nhiên có phân phối mũ tham số $\gamma_i = |E(T_i, \bar{T}_i)|$.

Sử dụng quá trình 2, ta suy ra:

$$\mathbb{E}(d_G(1,m)) = S_m := \sum_{i=1}^{m-1} \mathbb{E}\left(\frac{1}{\gamma_i}\right), \quad \mathrm{Var}(d_G(1,m)) = \sum_{i=1}^{m-1} \mathbb{E}\left(\frac{1}{\gamma_i^2}\right).$$

Mặt khác, sử dụng giả thiết $\forall i \in [n] : (1-\theta)dn \le \deg(i) \le (1+\theta)dn$:

$$(1-\theta)dni - \binom{i}{2} \le |E(T_i, \bar{T}_i)| \le (1+\theta)dni \quad h.c.c\ ,$$

Mà $\theta = 1/\sqrt{\log n}$ nên $\forall i$:

$$(1-\theta)i(dn-i) \le \gamma_i \le (1+\theta)idn \quad h.c.c\ .$$

Do đó với xác suất cao thì:

$$\begin{gathered}\forall\ 1 \le i \le \theta n : \gamma_i = idn(1+\zeta_i) \text{ với } |\zeta_i| = O(\theta),\\ \forall\ n-\theta n \le i \le n : \gamma_i = (n-i)dn(1+\zeta_i) \text{ với } |\zeta_i| = O(\theta).\end{gathered}$$

Ta suy ra:

$$S_{\theta n} = (1+O(\theta))\sum_{i=1}^{\theta n}\frac{1}{idn} = \frac{\log n}{dn} + O\left(\frac{\sqrt{\log n}}{n}\right) \quad w.h.p\ .$$

Quay trở lại với việc chứng minh bổ đề, ta làm tương tự như trong [8]. Đặt $k_1 = \theta n$ và $Y_i = E_i^{\gamma_i}, 1 \le i < n \Rightarrow Z_1 = d_G(1,k_1) = \sum_{i=1}^{k_1-1} Y_i$. Khi đó với $t < 1 - \frac{1+o(1)}{dn}$, và $m = k_1 - 1$ thì xác suất cao:

$$\begin{aligned}\mathbb{E}(e^{tdnZ_1}) = \mathbb{E}\left(\prod_{i=1}^{m} e^{tdnY_i}\right) &= \sum_x \mathbb{E}\left(\prod_{i=1}^{m} e^{tdnY_i} \mid \gamma_m = x\right)\mathbb{P}(\gamma_m = x)\\ &\big(\text{Biết } \gamma_m \text{ thì } Y_m \text{ độc lập với } Y_1, Y_2, \ldots, Y_{m-1}\big)\\ &= \mathbb{E}\left(\prod_{i=1}^{m-1} e^{tdnY_i}\right)\sum_x \mathbb{E}\left(e^{tdnY_m} \mid \gamma_m = x\right)\mathbb{P}(\gamma_m = x)\\ &= \mathbb{E}\left(\prod_{i=1}^{m-1} e^{tdnY_i}\right)\sum_x \frac{x}{x-tdn}\mathbb{P}(\gamma_m = x)\\ &= \mathbb{E}\left(\prod_{i=1}^{m-1} e^{tdnY_i}\right)\left(1 - \frac{(1+o(1))t}{m}\right)^{-1}.\end{aligned}$$

Do đó với mọi $\beta > 0$ ta có

$$\begin{aligned}\mathbb{P}\left(Z_1 \ge \frac{\beta\log n}{dn}\right) &\le \mathbb{E}\left(e^{tdnZ_1 - t\beta\log n}\right)\\ &\le e^{-t\beta\log n}\prod_{i=1}^{m}\left(1 - \frac{(1+o(1))t}{i}\right)^{-1}\end{aligned}$$

$$\leq e^{-t\beta \log n} \exp\left(\sum_{i=1}^{m}\left(\frac{(1+o(1))t}{i} + O(i^{-2})\right)\right)$$
$$\leq \exp\left((1+o(1)-\beta)t\log n\right).$$

Chọn $\beta = 2 + o(1)$ thì

$$w.h.p\ \forall j \in [n] : d_G(j, k_1) \leq \frac{(2+o(1))\log n}{dn}.$$

Đặt $\hat{T}_{k_1}$ là tập tương ứng với T_{k_1} khi chúng ta sử dụng quá trình 2 với điểm xuất phát là đỉnh số 2. Trước tiên ta xét $d \leq 1/2$ và $G \in \mathcal{G}(d)$. Khi đó sử dụng điều kiện 2.1.2, ta suy ra hoặc $T_{k_1} \cap \hat{T}_{k_1} = \emptyset$, hoặc

$$\mathbb{P}\left(\nexists e \in E(T_{k_1}, \hat{T}_{k_1}) : l(e) \leq \frac{1}{n}\right) \leq \exp\left(-\frac{-\psi\theta^2 n^2}{n}\right) = o(n^{-2}).$$

Điều này chứng tỏ rằng với xác suất $1 - o(n^{-2})$ta không thể tìm được một đường đi có tổng độ dài $> \frac{(4+o(1))\log n}{dn} + \frac{1}{n}$ giữa một cặp điểm cố định. Do đó:

$$w.h.p\ \max_{i,j} d_G(i,j) \leq \frac{(4+o(1))\log n}{dn} + \frac{1}{n}.$$

Xét trường hợp còn lại $d > 1/2 \Rightarrow d = 1/2 + \varepsilon, \varepsilon > 0$. Theo Điều kiện (2.1.1) thì $\forall i : \deg(i) \geq (1-\theta)dn \Rightarrow$ mọi cặp đỉnh đều có tối thiểu $(2\varepsilon - 2\theta)n$ đỉnh kề chung. Ta ghép cặp các đỉnh thuộc T_{k_1} và $\hat{T}_{k_1}$, suy ra xác suất để không tồn tại một đường đi độ dài 2 giữa một cặp đỉnh, chỉ sử dụng các cạnh độ dài không vượt quá $\frac{\log n}{n\log(\log n)}$, bằng

$$\left(e^{-\left(\frac{\log n}{n\log(\log n)}\right)^2}\right)^{-\theta n((2\varepsilon-2\theta)n)} = o(n^{-2}).$$

Lập luận tương tự như trên bằng việc sử dụng chặn hợp trên tất cả các cặp đỉnh, ta thu được điều phải chứng minh. $\square$

Bổ đề 2.8. *Cố định $e = \{u, v\} \in E$, ta có:*

$$\mathbb{P}(e \in E_0' \mid \max_{i,j} d_G(i,j) \leq l_1) = o\left(\frac{\log n}{n}\right)$$

Chứng minh. Để ý rằng $e \in E_0'$ tức là $e = \{u, v\} \in E$ thỏa mãn $l(e) > l_0$ và $l(e)$ chính bằng độ dài đường đi ngắn nhất giữa u và v. Khi đó:

$$d_G(u, w) + l_{w,v} \geq l(e) = l_{u,v} = d_G(u, v) \quad \forall w \sim v.$$

Đặt $\xi = l(e), x = \max_{w \sim v} d_G(u, w)$ ta suy ra

$$l_{w,v} \geq l(e) - d_G(u, w) \geq \xi - x \quad \forall w \sim v.$$

Do đó:

$$\begin{aligned}
&\mathbb{P}(e \in E_0', \max_{i,j} d_G(i, j) \leq l_1) \\
&\qquad = \int_{l_0}^{l_1} e^{-\xi} \int_0^\infty \mathbb{P}\left(\bigcap_{w \sim v} \{l_{w,v} \geq \xi - x\}\right) dx d\xi \\
&\qquad \leq \int_{l_0}^{l_1} \int_0^\infty \mathbb{P}\left(\bigcap_{w \sim v} \{l_{w,v} \geq \xi - x\}\right) dx d\xi.
\end{aligned}$$

Mặt khác, vì $\deg(v) \geq (1-\theta)dn$ nên có nhiều hơn $dn/2$ đỉnh kề với v nên biểu thức trên bị chặn bởi:

$$I := \int_{l_0}^{l_1} \int_0^\infty \min\{1, e^{-dn(\xi - x)/2}\} dx d\xi.$$

Xét $l_2 = l_0 - \frac{(\log(\log n))^2}{dn} < l_0 \leq \xi$, ta có:

$$\begin{aligned}
I_1 &:= \int_{l_0}^{l_1} \int_0^{l_2} \min\{1, e^{-dn(\xi - x)/2}\} dx d\xi \\
&= \int_{l_0}^{l_1} \int_0^{l_2} e^{-dn(\xi - x)/2} dx d\xi \\
&= \frac{4}{d^2 n^2}\left(e^{dnl_2/2} - 1\right)\left(e^{-dnl_0/2} - e^{-dnl_1/2}\right) \\
&\leq \frac{4}{d^2 n^2}\left(e^{dn(l_2 - l_0)/2} - e^{-dnl_0/2}\right).
\end{aligned}$$

Để ý rằng $l_0 = \frac{(1+\sqrt{\theta})\log n}{dn}, d \gg \theta \log(\log n)$, và $\theta = \frac{1}{\sqrt{\log n}}$ nên

$$I_1 \leq \frac{4}{d^2 n^2}\left(e^{-(\log(\log n))^2/2} - e^{-(1+\theta)\log n/2}\right)$$

$$= \underbrace{\frac{4}{d^2n^2}\left(e^{-(\log(\log n))^2/2}-1\right)}_{I_{1,1}} + \underbrace{\frac{4}{d^2n^2}\left(1-e^{-(1+\theta)\log n/2}\right)}_{I_{1,2}},$$

trong đó

$$\begin{aligned}
I_{1,1} &\leq \frac{4}{d^2n^2}\left((\log(\log n))^2/2 + (\log(\log n))^4/8\right) \\
&= \frac{\log n}{n}\left(\frac{4}{d^2\log n}\cdot\frac{(\log(\log n))^2}{2}\cdot\frac{(\log(\log n))^2+4}{4n}\right) \\
&= o\left(\frac{\log n}{n}\right),
\end{aligned}$$

và

$$I_{1,2} \leq \frac{4}{d^2n^2}\cdot\frac{(1+\theta)\log n}{2} \leq \frac{4\log n}{d^2n^2} = o\left(\frac{\log n}{n}\right).$$

Ta suy ra: $I_1 = o\left(\frac{\log n}{n}\right)$.
Bây giờ, ta xét:

$$I_2 = \int_{\xi=l_0}^{l_1}\int_{x=l_2}^{\infty}\min\{1, e^{-dn(\xi-x)/2}\}dxd\xi.$$

Theo chứng minh tại Bổ đề 7 trong [3], ta có $I_2 = o\left(\frac{\log n}{n}\right)$.
Vậy:

$$\mathbb{P}(e\in E_0', \max_{i,j} d_G(i,j) \leq l_1) =\leq I = I_1 + I_2 = o\left(\frac{\log n}{n}\right).$$

Mà theo Bổ đề 2.7 thì:

$$\mathbb{P}(\max_{i,j} d_G(i,j) \leq l_1) = 1 - o(1)$$

. Kết hợp hai ước lượng trên ta thu được điều phải chứng minh. □

Sử dụng Bổ đề 2.7, và Bổ đề 2.8, dễ dàng ta thu được kết quả với xác suất cao thì $|E_0'| = o(n\log n)$. Mà ta đã chứng minh được $w.h.p$ $|E_0| \approx \frac{1}{2}n\log n$ (theo Bổ đề 2.6) nên ta suy ra:

$$|E'| = |E_0 \cup E_0'| = |E_0| + |E_0'| \approx \frac{1}{2}n\log n \quad w.h.p.$$

Do đó, $G' = (V, E')$ là $1 - bao$ của G với xác suất cao có kích thước xấp xỉ $\left(\frac{1}{2} + o(1)\right) n \log n$. Như vậy, với xác suất cao thì kích thước tối thiểu $1 - bao$ của G không vượt quá $\left(\frac{1}{2} + o(1)\right) n \log n$. Định lý 2.1 được chứng minh xong.

Chương 3

Bao của đồ thị ngẫu nhiên có trọng hình học

Như chúng ta đã thấy, tại chương 2, mô hình đồ thị ngẫu nhiên ta nghiên cứu có tập đỉnh và tập cạnh là tất định, còn hàm trọng là ngẫu nhiên. Trong chương này, ta sẽ nghiên cứu kích thước của bao trên một mô hình đồ thị ngẫu nhiên khác, hiểu đơn giản là đồ thị với tập đỉnh và tập cạnh là ngẫu nhiên, còn hàm trọng là khoảng cách Euclid.

3.1 Đồ thị ngẫu nhiên có trọng hình học

Trước khi tới với mô hình đồ thị ngẫu nhiên có trọng hình học - đối tượng nghiên cứu chính của chúng ta trong chương này, ta cùng điểm qua một số khái niệm liên quan.

3.1.1 Đồ thị hình học ngẫu nhiên

Đồ thị hình học ngẫu nhiên là một khái niệm mở rộng từ khái niệm đồ thị hình học - một lĩnh vực lớn và được quan tâm tương tối nhiều trong các nghiên cứu về mạng nói chung. Tại đây, ta nêu lại định nghĩa về đồ thị hình học ngẫu nhiên trong [9].

Định nghĩa 3.1. [*9*] *Xét* d *là một số nguyên dương, và* $\|.\|$ *là một chuẩn* *trong* $\mathbb{R}^d$. *Khi đó, với tham số* $r > 0$ *và tập* $V \subset \mathbb{R}^d$ *cho trước, ta định nghĩa một* ***đồ thị hình học vô hướng*** $G = G(V, r)$ *như sau:*

- *tập đỉnh là* V*;*
- *hai đỉnh* X, Y *được nối với nhau nếu và chỉ nếu* $\|X - Y\| \leq r$.

Ta thường xét $d \in \{1, 2\}$ *và chuẩn là khoảng cách Euclid. Đồng thời, đồ thị hình học được trang bị hàm trọng chính là chuẩn của các cạnh, tức là nếu* $X \sim Y$ *thì trọng* $l_{X,Y} = \|X - Y\|$.

$G(V, r)$ ***là một đồ thị hình học ngẫu nhiên*** *nếu tập đỉnh* V *được chọn ngẫu nhiên (thường theo phân phối đều) trong* $\mathbb{R}^d$.

3.1.2 Đồ thị ngẫu nhiên có trọng hình học

Quay trở lại với chủ đề chính trong chương, ta sẽ định nghĩa $\mathcal{X}_p-$ một đồ thị ngẫu nhiên có trọng hình học.

Định nghĩa 3.2. *Cho trước* $0 \leq p \leq 1$, *xét* $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ *là tập* n *điểm được chọn ngẫu nhiên theo phân phối đều trong hình vuông đơn vị* $[0, 1]^2$. *Khi đó* $\mathcal{X}_p = (\mathcal{X}, p)$ *là một đồ thị vô hướng có hàm trọng* l, *với*

- *tập đỉnh là* $\mathcal{X}$*;*
- *hai đỉnh được nối với nhau một cách độc lập theo xác suất* p*;*
- *nếu hai đỉnh* X_i *và* X_j *được nối với nhau thì định nghĩa* $l(X_i, X_j) = |X_i - X_j|$: *khoảng cách Euclid của* X_i *và* X_j *trong* $\mathbb{R}^2$.

Từ định nghĩa trên, ta có thể thấy $\mathcal{X}_p$ là một sự kết hợp thú vị giữa đỉnh và hàm trọng của đồ thị hình học ngẫu nhiên, với xác suất nối cạnh của $G_{n,p}$, được nhúng trong hình vuông đơn vị.

Thực tế, khái niệm này đã được đề cập trong [10], nhưng chỉ xét trường hợp đặc biệt $p = 1$. Ngoài ra, trong một phiên thảo luận về các vấn đề mở tại Canadian Conference on Computational Geometry 2009 [11], J. O'Rourke đã đặt ra câu hỏi: *Với giá trị nào của p thì w.h.p* $\mathcal{X}_p$ *sẽ là* $t-$*bao của* $\mathcal{X}_1$. Khoảng 4 năm sau, Mehrabian và Wormald [12] chỉ ra rằng không có lựa chọn nào cho p thỏa mãn mệnh đề trên. Gần đây, Frieze và Pegden [13] chứng minh một kết quả phủ định, và đồng thời xem xét sự gia tăng của độ dài đường đi ngắn nhất khi chuyển từ $\mathcal{X}_1$ sang $\mathcal{X}_p$.

Từ cách xây dựng $\mathcal{X}_p$, dễ thấy với xác suất cao thì $1-$*bao* của $\mathcal{X}_p$ chứa $\approx \binom{n}{2}p$ cạnh, tức là chứa tất cả các cạnh. Lý do là nếu $X \sim Y$ thì cạnh nối giữa X và Y chính là đường đi ngắn nhất. Trong khuôn khổ chương này, ta quan tâm tới cỡ của $(1+\varepsilon)-$*bao* của $\mathcal{X}_p$.

3.2 Kết quả chính

Định lý 3.1. *[4] Giả sử* $np^{1+\theta} \to \infty$. *Xét đồ thị* $\mathcal{X}_p$. *Với* $\varepsilon, \theta > 0$ *là những hằng số tùy ý cố định, khi đó tồn tại một cách xây dựng tập cạnh* E_ε *thỏa mãn đồng thời:*

(i.) w.h.p E_ε *là* $(1+5\varepsilon)-$*bao của* $\mathcal{X}_p$.

(ii.) $\mathbb{E}(|E_\varepsilon|) \leq O_{\varepsilon,\theta}(p^{-\theta}n)$.

Với điều kiện $np^{1+\theta} \to \infty$ thì xác suất cao đồ thị $\mathcal{X}_p$ liên thông. Điều này đảm bảo việc chúng ta làm là có ý nghĩa, vì khi đó thì tồn tại bao của $\mathcal{X}_p$.

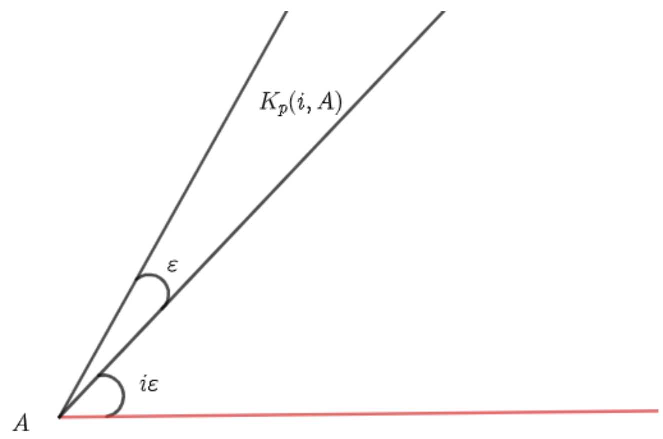
3.3 Chứng minh Định lý 3.1

Định lý 3.1 sẽ được chứng minh bằng cách chỉ ra thuật toán xây dựng E_ε của đồ thị $\mathcal{X}_p$, và từ đó ta ước lượng kích thước của $|E_\varepsilon|$.

3.3.1 Một số kí hiệu

Trong toàn bộ chứng minh, ta xét $0 < \varepsilon \ll 1$ và sử dụng một số kí hiệu sau:

- Với P là một đường đi $v_1 \to v_2 \to \cdots \to v_n$ nào đó, ta kí hiệu $l(P)$ là độ dài của P, tức:
$$l(P) = \sum_{i=2}^{n} |v_{i+1} - v_i|.$$
- Với $A, B \in \mathcal{X}$, ta kí hiệu $P_{A,B}$ là đường đi ngắn nhất từ A tới B trong $\mathcal{X}_p$, kí hiệu $d_{A,B} = l\left(P_{A,B}\right)$, và kí hiệu $E(P_{A,B})$ là tập cạnh chứa trong đường đi $P_{A,B}$.
- Với $A \in \mathcal{X}$, ta xét $\tau \le \frac{2\pi}{\varepsilon}$ nón, kí hiệu là $K_p(i, A)$ với $0 \le i < \tau$. Chúng theo thứ tự là nón đỉnh A, với hai tia tạo với trục hoành hai góc lần lượt là $i\varepsilon$ và $(i+1)\varepsilon$.



Hình 3.1: Mô tả $K_p(i, A)$.

Khi đó, ta định nghĩa $Y(i, A)$ là điểm có khoảng cách Euclid tới A là bé nhất, đồng thời kề với A, nằm trong nón $K_p(i, A)$.
Trường hợp nón không chứa điểm nào kề với A, ta kí hiệu $Y(i, A) = \perp$ và đặt $d_{A,\perp} = \infty$. Thêm vào đó, nếu $B \in \mathcal{X} \cap K_p(i, A)$ thì ta định nghĩa $i_{A,B} = i$.

- Đặt

$$r_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}}{np^{1+\theta}}} \quad \text{và} \quad R_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}\log n}{np^{1+\theta}}},$$

  trong đó $M_{\theta,\varepsilon}$ là hằng số đủ lớn, phụ thuộc vào θ và ε.

3.3.2 Trường hợp $p = 1$

Trước tiên, có lẽ ta nên quan tâm tới trường hợp $p = 1$, hay $\mathcal{X}_1$. Trong trường hợp này ta đã biết một thuật toán đơn giản để tìm $(1+\varepsilon)-$ *bao* của $\mathcal{X}_1$. Cụ thể trong [14], Yao xây dựng đường đi $\widetilde{P}_{A,B} = (A = Z_0, Z_1, Z_2, \ldots, Z_m = B)$ với $Z_{i+1} = Y(i_{Zi,B}, Z_i)$, và chứng minh được $\widetilde{P}_{A,B}$ có độ dài không quá $(\cos\varepsilon - \sin\varepsilon)^{-1}|A-B|$.

Khi đó rõ ràng đồ thị con chứa tất cả các cạnh $(A, Y(i,A)), 0 \leq i < \tau, \forall A \in \mathcal{X}$ sẽ chính là là $t-$ *bao* của $\mathcal{X}_1$ với $t = (\cos\varepsilon - \sin\varepsilon)^{-1} \leq 1 + \varepsilon + O(\varepsilon^2)$. Đồ thị con này được mang tên ông - Yao Graph. Thêm nữa, dễ thấy số cạnh của Yao Graph không quá $n \cdot \frac{2\pi}{\varepsilon} = O(n/\varepsilon)$. Do vậy, ta chỉ cần chọn E_ε chính bằng Yao Graph thì Định lý 3.1 được chứng minh xong trong trường hợp $p = 1$.

3.3.3 Trường hợp $p < 1$

Như vậy, ta chỉ còn quan tâm trường hợp $p < 1$, khi đó nhìn chung $\widetilde{P}_{A,B}$ được định nghĩa phía trên không tồn tại trong $\mathcal{X}_p$.

Phần 1: Một số kết quả chuẩn bị

Xét $A, B \in \mathcal{X}$, ta có hai bổ đề mô tả tính chất của $\mathcal{X}_p$ khi khoảng cách Euclid $|A - B| \geq R_\varepsilon$.

Bổ đề 3.1. *Nếu $|A-B| \geq R_\varepsilon$ thì với xác suất $1 - o(n^{-10})$ thì*

$$|A - Y| \leq \varepsilon|A - B|,$$

trong đó $Y = Y(i_{A,B}, A)$.

Chứng minh. Đặt $r = |A - B| \Rightarrow r \geq R_\varepsilon$. Ta gọi biến cố trong bổ đề là $\mathcal{E}$ và ước lượng $\mathbb{P}(\mathcal{E})$. Rõ ràng:

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}(\exists Y \in X \cap K_p(i_{A,B}, A), |A - Y| \leq \varepsilon r, Y \sim A).$$

Chọn ngẫu nhiên một điểm $Y \in [0,1]^2$, ta có:

$$\begin{aligned}\mathbb{P}(\mathcal{E}') &:= \mathbb{P}(Y \in X \cap K_p(i_{A,B}, A), |A - Y| \leq \varepsilon r, Y \sim A)\\ &= \frac{\varepsilon}{2\pi}.\pi(\varepsilon r)^2.p \geq \frac{\varepsilon^3 R_\varepsilon^2 p}{2}.\end{aligned}$$

Suy ra $\mathbb{P}(\bar{\mathcal{E}}') \leq 1 - \dfrac{\varepsilon^3 R_\varepsilon^2 p}{2}$, nên

$$\mathbb{P}(\bar{\mathcal{E}}) \leq \left(1 - \frac{\varepsilon^3 R_\varepsilon^2 p}{2}\right)^{n-2} \text{ hay } \mathbb{P}(\mathcal{E}) \geq 1 - \left(1 - \frac{\varepsilon^3 R_\varepsilon^2 p}{2}\right)^{n-2}.$$

Đến đây, sử dụng giả thiết $R_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}\log n}{np^{1+\theta}}}$ và bất đẳng thức $e^{-x} \geq 1 - x$, ta thu được:

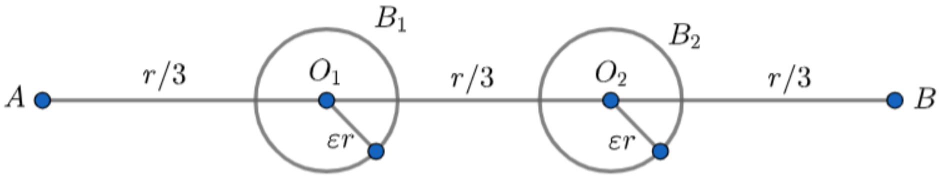
$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\geq 1 - \left(1 - \frac{\varepsilon^3 R_\varepsilon^2 p}{2}\right)^{n-2}\\ &\geq 1 - \left(1 - \frac{\varepsilon^3 M_{\theta,\varepsilon}\log n}{2p^\varepsilon n}\right)^{n-2}\\ &\geq 1 - \exp\left(-(n-2)\frac{\varepsilon^3 M_{\theta,\varepsilon}\log n}{2p^\varepsilon n}\right)\\ &\geq 1 - n^{-(\varepsilon^3 M_{\theta,\varepsilon})/3p^\varepsilon} = 1 - o(n^{-10}).\end{aligned}$$

Vậy ta có điều phải chứng minh. □

Bổ đề 3.2. *Nếu* $|A - B| \geq R_\varepsilon$ *thì với xác suất cao*

$$d_{A,B} \leq (1 + 4\varepsilon)|A - B|.$$

Chứng minh. Đặt $r = |A - B| \Rightarrow r \geq R_\varepsilon$.
Trên đoạn thẳng AB lấy hai điểm O_1, O_2 chia AB thành 3 phần bằng nhau. Đặt $B_i, (i = 1, 2)$, là hình cầu tâm O_i, bán kính εr.
Gọi



Hình 3.2: Mô tả hình cầu B_1, B_2.

- A_1 là tập các điểm thuộc $\mathcal{X} \cap B_1$ mà kề với A;
- A_2 là tập các điểm thuộc $\mathcal{X} \cap B_2$ mà kề với B;
- Biến cố $\mathcal{E} = \{d_{A,B} \leq (1+4\varepsilon)|A-B|\}$;
- Biến cố $\mathcal{E}_1 = \left\{|A_1| \geq \frac{\pi r^2 np}{10}\right\}$;
- Biến cố $\mathcal{E}_2 = \left\{|A_2| \geq \frac{\pi r^2 np}{10}\right\}$;
- $\mathcal{E}_3$ là biến cố tồn tại một cạnh của $\mathcal{X}_p$ nối giữa B_1 và B_2, tức chúng nối hai đỉnh u, v nào đó, với $u \in B_1$ và $v \in B_2$.

Sử dụng bất đẳng thức tam giác, ta dễ dàng chứng minh được nếu $\mathcal{E}_1, \mathcal{E}_2$, và $\mathcal{E}_3$ xảy ra đồng thời thì $\mathcal{E}$ sẽ xảy ra. Vậy:

$$\mathbb{P}(\mathcal{E}) \geq \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3) = \mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_1 \cap \mathcal{E}_2)\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2). \tag{3.3.1}$$

Đến đây, ta sẽ đi ước lượng các thừa số ở vế phải.

- Ước lượng $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2)$:
  Lấy ngẫu nhiên một điểm $Y \in [0,1]^2$, ta có với $\forall i = 1, 2$ thì

  $$\mathbb{P}(Y \in B_i) = \pi(\varepsilon r)^2 p.$$

  Ta suy ra $|A_i| \sim Bin(n, \pi(\varepsilon r)^2 p)$, xong ta áp dụng Bất đẳng thức Chernoff 1.3:

  $$\mathbb{P}\left(|A_i| \geq (1-\alpha)n\pi(\varepsilon r)^2 p\right) \geq 1 - \exp(-\alpha^2 n\pi(\varepsilon r)^2 p/2).$$

Chọn $(1-\alpha)\varepsilon^2 = \frac{1}{10}$ hay $\alpha = 1 - \frac{1}{10\varepsilon^2}$, và $\alpha^2\varepsilon^2 \geq \frac{1}{500}$, ta thu được

$$\mathbb{P}(\mathcal{E}_i) = \mathbb{P}\left(|A_1| \geq \frac{\pi r^2 np}{10}\right) \geq 1 - e^{\frac{-n\pi r^2 p}{1000}},\ \forall i = 1, 2.$$

Mặt khác, vì các điểm được lấy độc lập theo phân phối đều trong hình vuông đơn vị nên $\mathcal{E}_1$ và $\mathcal{E}_2$ độc lập. Do đó

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq \left(1 - e^{\frac{-n\pi r^2 p}{1000}}\right)^2 \geq 1 - 2e^{\frac{-n\pi r^2 p}{1000}}.$$

Cuối cùng, sử dụng giả thiết $r \geq R_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}\log n}{np^{1+\theta}}}$, ta có ước lượng

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - 2n^{\frac{-\pi}{1000p^\theta}}. \tag{3.3.2}$$

- Ước lượng $\mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_1 \cap \mathcal{E}_2)$:
  Vì $\mathbb{P}(\mathcal{E}_3 \mid |A_1|, |A_2|) = 1 - (1-p)^{|A_1|\cdot|A_2|}$ nên

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_1 \cap \mathcal{E}_2) &\geq 1 - (1-p)^{\frac{\pi^2 r^4 n^2 p^2}{100}} \\
&\geq 1 - (1-p)^{\frac{\pi^2 R_\varepsilon^4 n^2 p^2}{100}} \\
&\geq 1 - (1-p)^{\frac{\pi^2 M_{\theta,\varepsilon}^2 \log^2 n}{100p^{2\theta}}} \\
&\geq 1 - e^{\frac{-p\pi^2 M_{\theta,\varepsilon}^2 \log^2 n}{100p^{2\theta}}} = 1 - n^{\frac{-\pi^2 M_{\theta,\varepsilon}^2 \log n}{100p^{2\theta-1}}}.
\end{aligned} \tag{3.3.3}$$

Sau cùng, ta thế (3.3.2) và (3.3.3) vào (3.3.1) và đi đến kết luận:

$$\mathbb{P}(\mathcal{E}) \geq \left(1 - 2n^{\frac{-\pi}{1000p^\theta}}\right)\left(1 - n^{\frac{-\pi^2 M_{\theta,\varepsilon}^2 \log n}{100p^{2\theta-1}}}\right) \to 1.$$

Ta thu được điều phải chứng minh. □

Phần 2: Chứng minh (i): Thuật toán xây dựng E_ε

Xét $A, B \in \mathcal{X}$, kí hiệu $r = |A - B|$. Ta đặt

- $E_1 = \bigcup_{|A-B| \leq r_\varepsilon} E\left(P_{A,B}\right)$.
- $E_2 = \{\{A, Y(i, A)\} \in E(\mathcal{X}_p), A \in \mathcal{X}, i = 0, 1, \ldots, \tau - 1\}$.

- $E_3 = \bigcup_{\{A,B\}\in\mathcal{B}_\varepsilon} E\left(P_{A,B}\right)$, trong đó
  $\mathcal{B}_\varepsilon = \{\{A,B\} : d_{A,B} \geq (1+\varepsilon)r, r \geq r_\varepsilon\}$.

- $E_4 = \bigcup_{\{A,B\}\in\mathcal{C}_\varepsilon} E\left(P_{A,B}\right)$, trong đó
  $\mathcal{C}_\varepsilon = \{\{A,B\} : d_{A,B} \leq (1+\varepsilon)r, |A-Y| \geq \varepsilon r, r \in [r_\varepsilon, R_\varepsilon]\}$.
  Ở đây, $Y = Y(i_{A,B}, A)$.

Đặt

$$E_\varepsilon = \bigcup_{i=1}^{4} E_i.$$

Mệnh đề 3.1. *Với xác suất cao E_ε là $(1+5\varepsilon)-$bao của $\mathcal{X}_p$.*

Để chứng minh được Mệnh đề 3.1, trước tiên ta xây dựng Thuật toán 3.1.

Thuật toán 3.1. *Cho trước $A, B \in \mathcal{X}$, ta sẽ xây dựng một đường đi từ A đến B, kí hiệu $\hat{P}_{A,B} = (A = Z_0, Z_1, Z_2, \ldots, Z_k = B)$ trong $\mathcal{X}_p$ như sau: Tại bước thứ $j, (j \geq 0)$, sau khi xác định được Z_j, kí hiệu $Y_j = Y(i_{Z_j,B}, Z_j)$, ta thực hiện lần lượt 4 khâu:*

1. *Nếu $|Z_j - B| \leq r_\varepsilon$, ta sử dụng $P_{Z_j,B}$ để hoàn thiện $\hat{P}_{A,B}$.*
   Kết thúc thuật toán.

2. *Nếu $|Z_j - Y_j > \varepsilon|Z_j - B|$, ta sử dụng $P_{Z_j,B}$ để hoàn thiện $\hat{P}_{A,B}$.*
   Kết thúc thuật toán.

3. *Nếu $d_{Y_j,B} \geq (1+5\varepsilon)|Y_j - B|$, ta sử dụng $P_{Z_j,B}$ để hoàn thiện $\hat{P}_{A,B}$.*
   Kết thúc thuật toán.

4. *$Z_{j+1} := Y_j$.*
   Tiếp tục vòng lặp với bước thứ $j+1$.

Áp dụng Thuật toán 3.1, ta thu được một kết quả quan trọng về sự gia tăng của đường đi ngắn nhất như sau.

Bổ đề 3.3. *Với $A, B \in \mathcal{X}$, ta kí hiệu $\hat{d}_{A,B}$ là độ dài của $\hat{P}_{A,B}$. Khi đó:*

$$\hat{d}_{A,B} \leq (1+5\varepsilon)d_{A,B}, \ \forall A, B \in \mathcal{X}.$$

Chứng minh. Giả sử đường đi $\hat{P}_{A,B} = (A = Z_0, Z_1, Z_2, \ldots, Z_k = B)$ được tạo ra bởi Thuật toán 3.1 sau đúng l bước ($l \leq k-1$). Điều này có nghĩa là:

$$A = \underbrace{Z_0 \to Z_1 \to Z_2 \to \cdots \to Z_l}_{Z_{j+1}=Y_j \ \forall 0\leq j<l} \xrightarrow{P_{Z_l,B} \text{ trong } \mathcal{X}_p} Z_k = B. \tag{3.3.4}$$

Ta nhắc lại rằng $Y_j = Y(i_{Z_j,B}, Z_j)$. Như vậy:

$$\hat{d}_{A,B} = \sum_{j=0}^{l-1} |Z_{j+1} - Z_j| + d_{Z_l,B}. \tag{3.3.5}$$

Trường hợp $l = 0$ thì hiển nhiên $\hat{d}_{A,B} = d_{A,B} \leq (1+5\varepsilon)d_{A,B}$.
Ta xét $l > 0$. Theo giả thiết về số bước của thuật toán mô tả tại (3.3.4), ta suy ra các điều kiện của khâu 1, khâu 2, và khâu 3 của Thuật toán 3.1 đồng thời không xảy ra từ bước 0 tới bước thứ $l-1$. Do đó:

$$\forall\ 0 \leq j \leq l-1: \quad \begin{cases} |Z_j - B| > r_\varepsilon; \\ |Z_j - Y_j| \leq \varepsilon |Z_j - B|; \\ d(Y_j, B) < (1+5\varepsilon)|Y_j - B|. \end{cases} \tag{3.3.6}$$

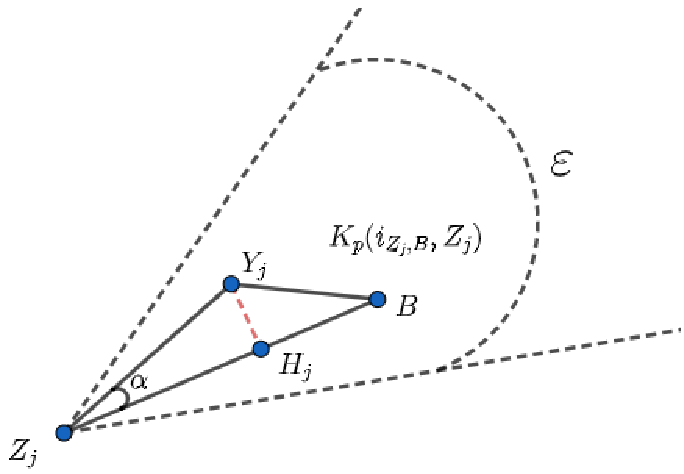
Đặt $d_j = |Z_j - B|$ với $0 \leq j \leq l$. Ta sẽ chứng minh

$$|Z_{j+1} - Z_j| \leq (1+5\varepsilon)(d_j - d_{j+1}) \ \forall\ 0 \leq j < l. \tag{3.3.7}$$

Thật vậy, theo mô tả (3.3.4) và tính chất (3.3.6), ta có:

$$|Z_{j+1} - Z_j| = |Y_j - Z_j| \leq \varepsilon |Z_j - B| < |Z_j - B|, \ \forall\ 0 \leq j < l.$$

Do đó ta thu được hình vẽ dưới đây, với $Y_jH_j \perp Z_jB$ và $\alpha = \widehat{Y_jZ_jB} \leq \varepsilon$.



Ta có

$$\begin{aligned}
(1+5\varepsilon)(d_j - d_{j+1}) &= (1+5\varepsilon)(|Z_j - B| - |Z_{j+1} - B|) \\
&= (1+5\varepsilon)(|Z_j - B| - |Y_j - B|) \\
&\geq (1+5\varepsilon)\left[|Z_j - B| - (|Y_j - H_j| + |H_j - B|)\right] \\
&= (1+5\varepsilon)\left(|Z_j - H_j| - |Y_j - H_j|\right) \\
&= (1+5\varepsilon)(\cos\alpha - \sin\alpha)|Z_j - Y_j| \\
&\geq \underbrace{(1+5\varepsilon)(\cos\varepsilon - \sin\varepsilon)}_{\geq 1 \text{ khi } \varepsilon \text{ đủ nhỏ}} |Z_j - Y_j| \\
&\geq |Z_j - Y_j| = |Z_{j+1} - Z_j|.
\end{aligned}$$

Như vậy ta đã chứng minh xong bất đẳng thức (3.3.7). Thêm nữa, rõ ràng $Z_l = Y_{l-1}$, sử dụng tính chất (3.3.6) với $j = l-1$ ta có:

$$d_{Z_l,B} = d_{Y_{l-1},B} \leq (1+5\varepsilon)|Y_{l-1} - B| = (1+5\varepsilon)|Z_l - B| = (1+5\varepsilon)d_l.$$

Áp dụng (3.3.7) và nhận xét trên vào (3.3.5) ta thu được:

$$\begin{aligned}
\hat{d}_{A,B} &= \sum_{j=0}^{l-1} |Z_{j+1} - Z_j| + d_{Z_l,B} \\
&\leq \sum_{j=0}^{l-1} (1+5\varepsilon)(d_j - d_{j+1}) + d_l = d_0 = |A - B| \leq d_{A,B}.
\end{aligned}$$

Vậy ta có điều phải chứng minh. $\square$

Theo Bổ đề 3.3 rõ ràng $\bigcup_{\{A,B\}\in\mathcal{X}} E(\hat{P}_{A,B})$ là $(1+5\varepsilon)-bao$ của $\mathcal{X}_p$.

Bổ đề 3.4. *Với xác suất cao thì*

$$\bigcup_{\{A,B\}\in\mathcal{X}} E(\hat{P}_{A,B}) \subset E_\varepsilon.$$

Chứng minh. Ta sẽ chứng minh $\bigcup_{\{A,B\}\in\mathcal{X}} E(\hat{P}_{A,B}) \subset \bigcup_{i=1}^{4} E_i$. Thật vậy, hiển nhiên tất cả các cạnh sinh ra bởi khâu 1 đều thuộc E_1.

Xét các cạnh sinh ra với khâu 2, khi đó điều kiện ở khâu 1 không xảy ra, tức là:

$$\begin{cases} |Z_j - B| \geq r_\varepsilon; \\ |Z_j - Y_j| > \varepsilon|Z_j - B| \Rightarrow w.h.p\ |Z_j - B| \leq R_\varepsilon \text{ (Bổ đề 3.1).} \end{cases}$$

Khi đó:

- Nếu $d_{Z_j,B} \geq (1+\varepsilon)|Z_j - B|$ thì các cạnh đó thuộc E_3.
- Nếu $d_{Z_j,B} \leq (1+\varepsilon)|Z_j - B|$ thì w.h.p các cạnh đó thuộc E_4.

Xét các cạnh sinh ra ở khâu 3, khi đó điều kiện ở khâu 1 và khâu 2 đồng thời không xảy ra, tức là:

$$\begin{cases} |Z_j - B| \geq r_\varepsilon; \\ |Z_j - Y_j| \leq \varepsilon|Z_j - B|; \\ d_{Y_j,B} \geq (1+5\varepsilon)|Y_j - B|. \end{cases}$$

Khi đó:

- Nếu $d_{Z_j,B} \geq (1+\varepsilon)|Z_j - B|$ thì các cạnh đó thuộc E_3.
- Nếu $d_{Z_j,B} \leq (1+\varepsilon)|Z_j - B|$ thì

$$\begin{aligned} (1+5\varepsilon)|Y_j - B| \leq d_{Y_j,B} &\leq |Y_j - Z_j| + d_{Z_j,B} \\ &\leq \varepsilon|Z_j - B| + (1+\varepsilon)|Z_j - B| \end{aligned}$$

$$
\begin{aligned}
&= (1+2\varepsilon)|Z_j - B| \\
&\le (1+2\varepsilon)\left(|Y_j - Z_j| + |Y_j - B|\right),
\end{aligned}
$$

suy ra $|Y_j - Z_j| \ge \frac{3\varepsilon}{1+2\varepsilon}|Y_j - B|$. Mà $|Z_j - Y_j| \le \varepsilon|Z_j - B|$ nên

$$|Z_j - B| \ge \frac{3}{1+2\varepsilon}|Y_j - B|.$$

Lúc này, ta có đánh giá

$$
\begin{aligned}
|Z_j - Y_j| &\ge |Z_j - B| - |Y_j - B| \text{ (bđt tam giác)} \\
&\ge \left(1 - \frac{1+2\varepsilon}{3}\right)|Z_j - B| \\
&\ge \left(1 - \frac{1+2\varepsilon}{3}\right)\frac{|Z_j - Y_j|}{\varepsilon}
\end{aligned}
$$

tương đương $5\epsilon \ge 2$ - mâu thuẫn với giả thiết $0 < \varepsilon \ll 1$. Vậy trường hợp này không xảy ra.

Xét các cạnh sinh ra ở khâu 4, hiển nhiên chúng thuộc E_2.
Kết hợp tất cả các khẳng định trên, ta có điều phải chứng minh. $\square$

Vậy với xác suất cao E_ε cũng là $(1+5\varepsilon)-$*bao* của $\mathcal{X}_p$, ta chứng minh xong Mệnh đề 3.1.

Phần 3: Chứng minh (ii): Ước lượng $|E_\varepsilon|$

Ta sẽ ước lượng $|E_\varepsilon|$ qua việc ước lượng các $|E_i|$, $i = \overline{1,4}$

1. **Ước lượng $|E_2|$**
   Nhắc lại $E_2 = \{\{A, Y(i,A)\} \in E(\mathcal{X}_p), A \in \mathcal{X}, i = 0, 1, \ldots, \tau - 1\}$.
   Do đó:
   $$|E_2| \le \tau n \le \frac{2\pi}{\varepsilon}n = O(n/\varepsilon). \qquad (3.3.8)$$

2. **Ước lượng $|E_3|$**
   Nhắc lại rằng với $A, B \in \mathcal{X}, r = |A - B|$ thì
   $$E_3 = \bigcup_{\{A,B\}\in\mathcal{B}_\varepsilon} E\left(P_{A,B}\right), \text{ với } \mathcal{B}_\varepsilon = \{\{A,B\} : d_{A,B} \ge (1+\varepsilon)r, r \ge r_\varepsilon\}.$$

Nhận xét rằng nếu $\{A,B\} \in \mathcal{B}_\varepsilon$ thì $A \nsim B$ nên đường đi ngắn nhất từ $A \to B$ là $P_{A,B}$ có tối thiểu 2 cạnh.
Với $K, k \in \mathbb{Z}_+, A, B \in \mathcal{X}$ ta định nghĩa $\mathcal{L}_{K,k,A,B}$ là tập hợp tất cả các đường đi P từ $A \to B$ trong $\mathcal{X}_p$ có đúng $k+1$ cạnh thỏa mãn đồng thời các điều kiện sau:

(i). P là một *đường đi rút gọn (induced path[1])*,

(ii). $l(P) \in [(1+K\varepsilon)r, (1+(K+1)\varepsilon)r]$,

(iii). $l(P^*) \geq l(P) \, \forall$ đường đi $P^* : A \to B$ không sử dụng đỉnh của P.

Khi đó, rõ ràng với $\forall \{A,B\} \in \mathcal{B}_\varepsilon$ thì $\exists K, k \in \mathbb{Z}_+ : P_{A,B} \in \mathcal{L}_{K,k,A,B}$.
Ta suy ra:

$$|E_3| \leq \sum_{\{A,B\} \in \mathcal{B}_\varepsilon} |E(P_{A,B})| \leq \sum_{\substack{A,B \in \mathcal{X} \\ |A-B| \geq r_\varepsilon}} \sum_{K,k \in \mathbb{Z}_+} (k+1)|\mathcal{L}_{K,k,A,B}|. \quad (3.3.9)$$

Do đó, để ước lượng được $|E_3|$, ta đi ước lượng $|\mathcal{L}_{K,k,A,B}|$.
Đầu tiên, ta thấy $k \leq n-2$, và khi k cố định thì vì các điểm đều thuộc $[0,1]$ nên khoảng cách lớn nhất giữa hai điểm là $\sqrt{2}$, do đó:

$$(1+K\varepsilon)r \leq l(P) \leq (k+1)\sqrt{2} \Rightarrow K \leq \frac{(k+1)\sqrt{2}}{r\varepsilon}.$$

Thêm nữa, xét $k \geq k_n$ với $k_n = \frac{20\log(np)}{p} = o(n)$ thì:

$$\begin{aligned}
\mathbb{E}(|\mathcal{L}_{K,k,A,B}|) &\leq \binom{n}{k} \mathbb{P}(P : \text{đường rút gọn}) \\
&\leq n^k p^{k+1} (1-p)^{k(k+1)/2} \\
&\leq p \cdot (npe^{-p(k+1)/2})^k \leq p \cdot (npe^{-pk_n/4})^k \\
&\leq p \cdot \left(\frac{1}{(np)^4}\right)^k = p \cdot O(n^{-4}).
\end{aligned}$$

[1]Induced path là một đường đi không có đường tắt. Tức nó là một dãy có thứ tự các đỉnh sao cho hai đỉnh kề nhau trong dãy thì nối với nhau, hai đỉnh không kề nhau thì không nối với nhau.

Tiếp theo, ta xét $k \leq k_n = o(n)$.

Cố định $A, B \in \mathcal{X}, |A - B| = r \geq r_\varepsilon$, và $K, k \in \mathbb{Z}_+$. Xét một đường đi $P = (A = P_0, P_1, P_2, \dots, P_k, P_{k+1} = B)$, và $\mathcal{E} = \{P \in \mathcal{L}_{K,k,A,B}\}$. Khi đó:

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &:= \mathbb{P}\left(P \in \mathcal{L}_{K,k,A,B}\right) \\
&= \mathbb{P}(P : \text{đường rút gọn},\ l(P) \in [(1 + K\varepsilon)r, (1 + (K+1)\varepsilon)r], \\
&\qquad l(P^*) \geq l(P)\ \forall P^* : A \to B, P^* \subset \mathcal{X} \setminus P) \\
&\leq \mathbb{P}(P : \text{đường rút gọn},\ l(P) \leq (1 + (K+1)\varepsilon)r, \\
&\qquad l(P^*) \geq (1 + K\varepsilon)r\ \forall P^*).
\end{aligned}$$

Đặt

$$\begin{aligned}
\mathcal{E}_1 &= \{P : \text{đường rút gọn}\}, \\
\mathcal{E}_2 &= \{l(P) \leq (1 + (K+1)\varepsilon)r\}, \\
\text{và } \mathcal{E}_3 &= \{l(P^*) \geq (1 + K\varepsilon)r\ \forall P^* : A \to B, P^* \subset \mathcal{X} \setminus P\}.
\end{aligned}$$

Ta thấy chúng độc lập với nhau nên:

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3) = \mathbb{P}(\mathcal{E}_1) \cdot \mathbb{P}(\mathcal{E}_2) \cdot \mathbb{P}(\mathcal{E}_3) \tag{3.3.10}$$

- Tính $\mathbb{P}(\mathcal{E}_1)$:

  Ta có:

  $$\mathbb{P}(\mathcal{E}_1) = \mathbb{P}(P : \text{đường rút gọn}) = p^{k+1}(1-p)^{k(k+1)/2}.$$

- Tính $\mathbb{P}(\mathcal{E}_2)$:

  Đặt $a = (1 + (K+1)\varepsilon)r, Z_i = |P_i - P_{i-1}|$. Ta có:

  $$\begin{aligned}
  \mathbb{P}(\mathcal{E}_2) &= \mathbb{P}\left(l(P) \leq (1 + (K+1)\varepsilon)r\right) \\
  &\leq \mathbb{P}\left(Z_1 + Z_2 + \dots + Z_k \leq a\right).
  \end{aligned}$$

  Tới đây, ta sẽ chứng minh bằng phương pháp quy nạp rằng:

  $$\forall 1 \leq l \leq k : \mathbb{P}\left(\sum_{i=1}^{l} Z_i \leq a\right) \leq \mathbb{P}\left(\sum_{i=1}^{l} \sqrt{U_i} \leq a\sqrt{\pi}\right) \quad (*),$$

trong đó, $U_1, U_2, \ldots$ là các các biến ngẫu nhiên độc lập, cùng phân phối $U[0,1]$.

– Với $l = 1$:

Kí hiệu $\mathcal{B}(I,R)$ là hình tròn tâm I, bán kính R. Ta có:

$$\begin{aligned}\mathbb{P}(Z_1 \le a) &= \mathbb{P}\left(|P_1 - P_0| \le a\right)\\ &\le \max_{P_0 \in [0,1]^2} \mathbb{P}\left(P_1 \in \mathcal{B}(P_0, a) \cap [0,1]^2\right)\\ &\le \min\{\pi a^2, 1\} = \mathbb{P}\left(\sqrt{U_1} \le a\sqrt{\pi}\right).\end{aligned}$$

Suy ra mệnh đề (*) đúng với $l = 1$.

– Giả sử mệnh đề (*) đúng tới $l \ge 1$, ta chứng minh nó cũng đúng với $l+1$.

Thật vậy: Đặt $S_l = Z_1 + Z_2 + \cdots + Z_l$, ta có

$$\mathbb{P}(Z_{l+1} \le v \mid S_l) \le \min\{\pi v^2 \mathbf{1}_{\{v>0\}}, 1\} \quad h.c.c\,.$$

Giải thích: bất đẳng thức trên không phụ thuộc vào các vị trí của các điểm $P_0, P_1, \ldots, P_l$ (giống cách chứng minh ở trường hợp $l = 1$) nên nó đúng với mọi khả năng của S_l. Do đó:

$$\begin{aligned}\mathbb{P}(S_l + Z_{l+1} \le a) &= \mathbb{E}\left[\mathbb{P}(Z_{l+1} \le a - S_l \mid S_l)\right]\\ &\le \mathbb{E}\left[\min\{\pi(a-S_l)^2 \mathbf{1}_{\{a>S_l\}}, 1\}\right].\end{aligned}$$

Mặt khác, ta cũng có:

$$\begin{aligned}\mathbb{P}\left(S_l + \frac{\sqrt{U_{l+1}}}{\sqrt{\pi}} \le a\right) &= \mathbb{E}\left(\mathbb{P}\left(\sqrt{U_{l+1}} \le \sqrt{\pi}(a - S_l)\right)\right)\\ &= \mathbb{E}\left[\min\{\pi(a-S_l)^2 \mathbf{1}_{\{a>S_l\}}, 1\}\right].\end{aligned}$$

Kết hợp cùng với giả thiết quy nạp thì ta sẽ thu được:

$$\mathbb{P}\left(\sum_{i=1}^{l+1} Z_i \le a\right) = \mathbb{P}(S_l + Z_{l+1} \le a)$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[\min\{\pi(a-S_l)^2\mathbf{1}_{\{a>S_l\}},1\}\right]\\
&= \mathbb{P}\left(S_l + \frac{\sqrt{U_{l+1}}}{\sqrt{\pi}} \leq a\right)\\
&= \mathbb{P}\left(\sum_{i=1}^{l} Z_i \leq a - \frac{\sqrt{U_{l+1}}}{\sqrt{\pi}}\right)\\
&\leq \mathbb{P}\left(\sum_{i=1}^{l} \sqrt{U_i} \leq \left(a - \frac{\sqrt{U_{l+1}}}{\sqrt{\pi}}\right)\sqrt{\pi}\right)\\
&= \mathbb{P}\left(\sum_{i=1}^{l+1} \sqrt{U_i} \leq a\sqrt{\pi}\right)
\end{aligned}
$$

Vậy mệnh đề (*) cũng đúng với $l+1$.

Theo nguyên lý quy nạp, ta đã chứng minh được tính đúng đắn của mệnh đề (*). Sử dụng Định lý 1.4 với $\gamma = 1/2$ và $u = a\sqrt{\pi}$, và Hệ quả xấp xỉ Stirling 1.2, ta kết luận:

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_2) \leq \mathbb{P}\left(\sum_{i=1}^{k} Z_i \leq a\right) &\leq \mathbb{P}\left(\sum_{i=1}^{k} \sqrt{U_i} \leq a\sqrt{\pi}\right)\\
&\leq \frac{(a\sqrt{\pi})^{2k}\cdot 2^k}{(2k)!} \leq \frac{a^{2k}\pi^k e^{2k}}{2^k k^{2k}}.
\end{aligned}
$$

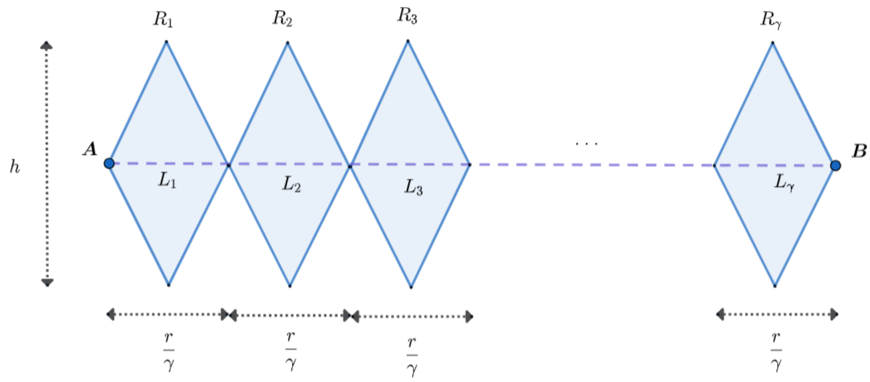
- Tính $\mathbb{P}(\mathcal{E}_3)$:

  Với γ là một hằng số nguyên dương bất kỳ, ta có:

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_3) &= \mathbb{P}\left(l(P^*) \geq (1+K\varepsilon)r \ \forall P^*\right)\\
&\leq \mathbb{P}\left(l(P^*) \geq (1+K\varepsilon)r \ \forall P^* \text{ có } \gamma+1 \text{ cạnh}\right),
\end{aligned}
$$

  với P^* : path $A \to B, P* \subset \mathcal{X} \setminus P$.

  Ta chia đoạn AB thành γ đoạn $L_1, L_2, \ldots, L_\gamma$ bằng nhau, độ dài bằng $\frac{|A-B|}{\gamma} = \frac{r}{\gamma}$. Với mỗi $i = 1, 2, \ldots, \gamma$, ta vẽ một hình thoi R_i có một đường chéo là L_i, đường chéo còn lại có độ dài $h = \frac{rF(K,\varepsilon)}{\gamma} = \frac{r\sqrt{2K\varepsilon + K^2\varepsilon^2}}{\gamma}$. Gọi diện tích của hình thoi như vậy là α, tức $\alpha = \frac{rh}{2\gamma}$. Đặt $\hat{R}_i = R_i \cap [0,1]^2$ và α_i là diện tích của $\hat{R}_i$,



ta suy ra $\forall i : \alpha \geq \alpha_i \geq \frac{1}{2}\alpha$.

Xét một đường đi $Q : A \to S_1 \to S_2 \to \cdots \to S_\gamma \to B$, với $S_i \in \hat{R}_i \; \forall i = 1, \ldots, \gamma$, ta có:

$$l(Q) \leq 2\gamma\sqrt{\left(\frac{1}{2} \cdot \frac{r}{\gamma}\right)^2 + \left(\frac{1}{2} \cdot h\right)^2} \leq (1 + K\varepsilon)r.$$

Khi đó, nếu kí hiệu Z là số đường đi Q như vậy trong $\mathcal{X}_p \setminus P$, thì:

$$\mathbb{P}(\mathcal{E}_3) \leq \mathbb{P}(Z = 0).$$

Mặt khác, đặt $\nu = |\mathcal{X} \setminus P| = n - k = n - o(n)$, ta có:

$$\mathbb{E}(Z) = \nu(\nu - 1) \ldots (\nu - \gamma + 1)p^{\gamma+1} \prod_{i=1}^{\gamma} \alpha_i \geq \left(\frac{n}{2}\right)^\gamma p^{\gamma+1} \left(\frac{\alpha}{2}\right)^\gamma.$$

Đồng thời, với hai đường đi Q, Q', đặt $\rho(Q, Q')$ và $\sigma(Q, Q')$ lần lượt là số điểm chung (không tính A, B) và số cạnh chung của hai đường đi. Kí hiệu $Q \sim Q'$ nếu $\sigma(Q, Q') > 0$. Khi đó do $\nu(\nu - 1) \ldots (\nu - (2\gamma - \rho) + 1) < \nu^{2\gamma - \rho}$ và $\alpha_i \leq \alpha \; \forall i$ nên:

$$\begin{aligned}\bar{\Delta} := \sum_{Q \sim Q'} \mathbb{P}(Q, Q') &\leq \sum_{\substack{1 \leq \sigma \leq \gamma+1 \\ \sigma \leq \rho \leq 2\sigma}} \nu^{2\gamma - \rho} \binom{2\gamma - \rho}{\rho} \alpha^{2\gamma - \rho} p^{2\gamma + 2 - \sigma} \\ &\leq 2^{2\gamma - 1}(\gamma + 2)^2 (n\alpha)^{2\gamma - 1} p^{2\gamma + 1}.\end{aligned}$$

Áp dụng công thức ước lượng 1.7 trong [15], ta thu được:

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_3) \le \mathbb{P}(Z=0) &\le \exp\left(-\frac{\mathbb{E}^2(Z)}{2\bar{\Delta}}\right) \\
&\le \exp\left(-\frac{n\alpha p}{2^{6\gamma}(\gamma+2)^2}\right) \\
&\le \exp\left(-\frac{nr^2F(K,\varepsilon)p}{2^{6\gamma+1}(\gamma+2)^2\gamma^2}\right).
\end{aligned}
$$

Vậy ta đã ước lượng xong $\mathbb{P}(\mathcal{E}_i)$ với $i=1,2,3$. Lắp các thừa số này vào (3.3.10), ta được:

$$
\begin{aligned}
\mathbb{P}(P \in \mathcal{L}_{K,k,A,B}) &= \mathbb{P}(\mathcal{E}) \le \mathbb{P}(\mathcal{E}_1)\cdot\mathbb{P}(\mathcal{E}_2)\cdot\mathbb{P}(\mathcal{E}_3) \\
&\le p^{k+1}(1-p)^{k(k+1)/2}\cdot\frac{a^{2k}\pi^k e^{2k}}{2^k k^{2k}}\cdot\exp\left(-\frac{nr^2F(K,\varepsilon)p}{2^{6\gamma+1}(\gamma+2)^2\gamma^2}\right) \\
&= p\cdot\left(\frac{C_0a^2p(p-1)^{(k+1)/2}}{k^2}\right)^k\cdot\exp\left(-C_2npr^2F(K,\varepsilon)\right),
\end{aligned}
$$

trong đó $a=(1+(K+1)\varepsilon)r$, $F(K,\varepsilon)=\sqrt{2K\varepsilon+K^2\varepsilon^2}$, và C_0, C_2 là các hằng số dương. Ta suy ra với $k<k_n$:

$$
\begin{aligned}
\mathbb{E}(|\mathcal{L}_{K,k,A,B}|) &= \binom{n-2}{k}\mathbb{P}(P\in\mathcal{L}_{K,k,A,B}) \\
&\le p\cdot\left(\frac{C_1a^2np(p-1)^{(k+1)/2}}{k^3}\right)^k\cdot\exp\left(-C_2npr^2F(K,\varepsilon)\right),
\end{aligned}
$$

vì $\binom{n-2}{k}\le\frac{n^k}{k!}\le\frac{n^k\cdot e^k}{k^k}$.

Lúc này, với giả thiết $\sqrt{2}\ge|A-B|=r\ge r_\varepsilon$, để thuận tiện ta sẽ thay thế r bởi $\frac{\rho}{\sqrt{np}}$, và chia các khoảng $J_\rho=\left[\frac{\rho}{\sqrt{np}},\frac{\rho+1}{\sqrt{np}}\right]$. Ta suy ra $r_\varepsilon\sqrt{np}\le\rho\le\sqrt{2np}$.

Sử dụng đánh giá (3.3.9) và giả thiết $r_\varepsilon=\sqrt{\frac{M_{\theta,\varepsilon}}{np^{1+\theta}}}$, ta có:

$$
\mathbb{E}(|E_3|)\le\mathbb{E}\left(\sum_{\substack{A,B\in\mathcal{X}\\|A-B|\ge r_\varepsilon}}\sum_{K,k\in\mathbb{Z}_+}(k+1)|\mathcal{L}_{K,k,A,B}|\right)
$$

$$
\begin{aligned}
&\le \sum_{\rho=r_\varepsilon\sqrt{np}}^{\sqrt{2np}} \binom{n}{2} \sum_{K,k\in\mathbb{Z}_+} (k+1)\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right)\cdot\mathbb{P}\left(r\in J_\rho\right)\\
&\le n^2\cdot \sum_{\rho=r_\varepsilon\sqrt{np}}^{\sqrt{2np}} \frac{\pi(2\rho+1)}{np} \sum_{K,k\in\mathbb{Z}_+} (k+1)\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right)\\
&\le 6\pi n\cdot \sum_{\rho=r_\varepsilon\sqrt{np}}^{\sqrt{2np}} \sum_{K,k\in\mathbb{Z}_+} \frac{k\rho}{p}\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right)\\
&= 6\pi n(H_1+H_2),
\end{aligned}
$$

trong đó:

$$
H_1 = \sum_{\substack{\rho,K\\ k\ge k_n}} \frac{k\rho}{p}\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right) \text{ và } H_2 = \sum_{\substack{\rho,K\\ k\le k_n}} \frac{k\rho}{p}\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right).
$$

Nhắc lại rằng: $k_n = \frac{20\log(np)}{p} = o(n)$ và $K \le \frac{(k+1)\sqrt{2}}{r\varepsilon} \le \frac{(k+1)\sqrt{2np}}{\rho\varepsilon}$.

- Tính H_1.

  Ta có:

$$
\begin{aligned}
H_1 &= \sum_{\rho=r_\varepsilon\sqrt{np}}^{\sqrt{2np}} \sum_{K=1}^{\frac{(k+1)\sqrt{2np}}{\rho\varepsilon}} \sum_{k=k_n}^{n-2} \frac{k\rho}{p}\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right)\\
&\le \sqrt{2np}\frac{(n-1)\sqrt{2np}}{\rho\varepsilon}(n-2)^2\rho\cdot O(n^{-4}) = O(1).
\end{aligned}
$$

- Tính H_2.

  Với $a = (1+(K+1)\varepsilon)\rho/\sqrt{np}$, $F(K,\varepsilon) = \sqrt{2K\varepsilon+K^2\varepsilon^2} > K\varepsilon := h$, $k_n = \frac{20\log(np)}{p}$, và $r_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}}{np^{1+\theta}}}$, ta có:

$$
\begin{aligned}
H_2 &= \sum_{\substack{\rho,K\\ k\le k_n}} \frac{k\rho}{p}\mathbb{E}\left(|\mathcal{L}_{K,k,A,B}|\right)\\
&\le \sum_{\substack{\rho,K\\ k\le k_n}} k\rho\left(\frac{C_1a^2np(p-1)^{(k+1)/2}}{k^3}\right)^k \exp\left(-C_2\rho^2F(K,\varepsilon)\right)
\end{aligned}
$$

$$\leq \sum_{\substack{\rho,h \\ k\leq k_n}} \rho \left(\frac{C_3h^2}{k^3}\right)^k \exp\left(-C_2h\rho^2\right) \text{ (với } C_3 = const)$$

$$\text{(Vì } k(1-p)^{k(k+1)/2} \leq 1 \ \forall\ k \in \mathbb{Z}_+, \forall\ 0 < p < 1)$$

$$\leq \sum_{\rho=r_\varepsilon\sqrt{np}}^{\sqrt{2np}} \sum_{h=\varepsilon}^{\frac{(k+1)\sqrt{2np}}{\rho}} \sum_{k=1}^{k_n} \rho \left(\frac{C_3h^2}{k^3}\right)^k \exp\left(-C_2h\rho^2\right)$$

$$\leq 2np(k_n+1) \sum_{k=1}^{k_n} \left(\frac{C_4np^{1-\theta}}{k}\right)^k \exp\left(-C_5k\sqrt{np^{1-\theta}}\right)$$

$$= O(1) \quad \text{(với } C_4, C_5 = const).$$

Vậy ta kết luận:

$$\mathbb{E}(|E_3|) \leq 6\pi n(H_1 + H_2) = O(n). \tag{3.3.11}$$

3. **Ước lượng** $|E_4|$

Nhắc lại $E_4 = \bigcup_{\{A,B\}\in\mathcal{C}_\varepsilon} E\left(P_{A,B}\right)$, trong đó

$$\mathcal{C}_\varepsilon = \left\{\{A,B\} : d_{A,B} \leq (1+\varepsilon)r, |A-Y| \geq \varepsilon r, r \in [r_\varepsilon, R_\varepsilon]\right\}.$$

Ở đây $Y = Y(i_{A,B}, A), r_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}}{np^{1+\theta}}}, R_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}\log n}{np^{1+\theta}}}$.

Đầu tiên, ta tách E_4 thành hai phần:

$$E_4 = \underbrace{\left(\bigcup_{\substack{\{A,B\}\in\mathcal{C}_\varepsilon \\ A\sim B}} E\left(P_{A,B}\right)\right)}_{E_{4,1}} \cup \underbrace{\left(\bigcup_{\substack{\{A,B\}\in\mathcal{C}_\varepsilon \\ A\nsim B}} E\left(P_{A,B}\right)\right)}_{E_{4,2}}.$$

Với $E_{4,1}$, để ý rằng khi $A \sim B$ thì $d_{A,B} = |A-B| = r \leq (1+\varepsilon)r$. Còn điều kiệu $|A-Y| \geq \varepsilon r$ tương đương với việc $\nexists\ Y \in K(i_{A,B}, A) \cap (A, \varepsilon r)^2, Y \sim A$ nên xác suất của nó bị chặn trên bởi $\left(1 - \frac{\varepsilon}{2\pi}\cdot\pi(\varepsilon r)^2\cdot p\right)^{n-2} = \left(1-\varepsilon^3r^2p\right)^{n-2}$. Do đó:

$$\mathbb{E}(|E_{4,1}|) \leq \int_{r=r_\varepsilon}^{R_\varepsilon} \binom{n}{2}(2\pi r)p\left(1-\varepsilon^3r^2p\right)^{n-2} dr$$

[2] Hình cầu tâm A, bán kính εr.

$$\leq 2\pi p\binom{n}{2}\int_{r=r_\varepsilon}^{R_\varepsilon}\exp(-\varepsilon^3 p(n-2)r^2)rdr$$
$$= \frac{\pi n(n-1)}{2\varepsilon^3(n-2)}\int_{-\varepsilon^3 p(n-2)r_\varepsilon^2}^{-\varepsilon^3 p(n-2)R_\varepsilon^2} e^{-t}dt$$
$$= O(n).$$

Với $E_{4,2}$, ta sẽ ước lượng $\mathbb{E}\left(|E_{4,2}|\right)$ tương tự như cách ước lượng $\mathbb{E}\left(|E_3|\right)$ đã nêu ở phần trên. Cụ thể: Với $k \in \mathbb{Z}_+, A, B \in \mathcal{X}$ ta định nghĩa $\mathcal{H}_{k,A,B}$ là tập hợp tất cả các đường đi P từ $A \to B$ trong $\mathcal{X}_p$ có đúng $k+1$ cạnh thỏa mãn đồng thời các điều kiện sau:

(i). P là một đường đi rút gọn.

(ii). $l(P) \leq (1+\varepsilon)r$.

(iii). $\nexists X \in \mathcal{X} \setminus P : X \in K(i_{A,B}, A), X \sim A, |A - X| < \varepsilon r$.

Khi đó rõ ràng với $\forall A, B \in \mathcal{C}_\varepsilon, A \nsim B$ thì tồn tại $k \in \mathbb{Z}_+ : P_{A,B} \in \mathcal{H}_{k,A,B}$. Với điều kiện (i) và (ii) thì cách xử lý giống hệt với $\mathcal{L}_{K,k,A,B}$ ở phần E_3, còn với (iii) ta để ý rằng xác suất để nó xảy ra sẽ bị chặn trên bởi $\left(1 - \frac{\varepsilon}{2\pi}\cdot\pi(\varepsilon r)^2\cdot p\right)^{n-2-k} = \left(1-\varepsilon^3 r^2 p\right)^{n-2-k}$. Do đó nếu đặt $\bar{a} = (1+\varepsilon)r$ thì:

$$\mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) \leq \binom{n-2}{k}\cdot\left(p^{k+1}(1-p)^{\frac{k(k+1)}{2}}\right) \cdot\left(\frac{\bar{a}^{2k}\pi^k e^{2k}}{2^k k^{2k}}\right)\cdot\left(1-\varepsilon^3 r^2 p\right)^{n-2-k}.$$

Lại có $\binom{n-2}{k} \leq \frac{n^k}{k!}$ nên với $C = \frac{e^3\pi(1+\varepsilon)^2}{2}$ là hằng số dương:

$$\mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) \leq \frac{p}{k!}\cdot\left(\frac{Cnp(1-p)^{(k+1)/2}r^2}{ek^2}\right)^k\cdot\left(1-\varepsilon^3 r^2 p\right)^{n-2-k}$$
$$\leq \frac{p}{k!}\cdot\left(\frac{Cnp(1-p)^{(k+1)/2}r^2}{ek^2}\right)^k\cdot e^{-\varepsilon^3 r^2 p(n-2-k)}$$
$$\left(\text{Vì } r \leq \sqrt{2} \Rightarrow e^{\varepsilon^3 r^2 p(k+2)} \leq e^{6\varepsilon^3 k} \leq e^k\right)$$

$$\leq \frac{p}{k!} \cdot \left(\frac{Cnp(1-p)^{(k+1)/2} r^2}{k^2} \right)^k \cdot e^{-\varepsilon^3 r^2 pn}$$
$$:= C_k r^{2k} e^{-\varepsilon^3 r^2 pn}.$$

Lúc này ta có đánh giá:

$$\mathbb{E}(|E_{4,2}|) \leq \int_{r=r_\varepsilon}^{R_\varepsilon} \binom{n}{2} (2\pi r) \left(\sum_{k=0}^{n-2} (k+1) \mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) \right) dr$$
$$\leq \pi n \sum_{k=1}^{n-2} kn \int_{r=r_\varepsilon}^{R_\varepsilon} \mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) 2r dr.$$

Đặt $S = \varepsilon^3 r_\varepsilon^2 pn = \frac{\varepsilon^3 M_{\theta,\varepsilon}}{p^\theta}$ thì:

$$\int_{r=r_\varepsilon}^{R_\varepsilon} \mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) 2r dr \leq C_k \int_{r=r_\varepsilon}^{R_\varepsilon} r^{2k} e^{-\varepsilon^3 r^2 pn} 2r dr$$
$$\leq \frac{C_k}{(\varepsilon^3 pn)^{k+1}} \int_S^\infty s^k e^{-s} ds.$$

Bây giờ, ta đặt $I_k = \int_S^\infty s^k e^{-s} ds$ thì rõ ràng $I_k = kS^{k-1}e^{-S} + kI_{k-1}$, do vậy:

$$I_k = e^{-S} \cdot k! \cdot \sum_{l=0}^{k-1} \frac{S^l}{l!} \leq 2e^{-S} \cdot k! \cdot S^k.$$

Để ý rằng $C_k = \frac{p}{k!} \cdot \left(\frac{Cnp(1-p)^{(k+1)/2}}{k^2} \right)^k = \frac{p}{k!} \cdot \left(\frac{(1+\varepsilon)^2 \pi e^3 np(1-p)^{(k+1)/2}}{2k^2} \right)^k$ và $\forall k \in \mathbb{Z}_+, \forall 0 < p < 1 : k(1-p)^{k(k+1)/2} \leq 1$, ta suy ra:

$$\sum_{k=1}^{n-2} kn \int_{r=r_\varepsilon}^{R_\varepsilon} \mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) 2r dr \leq \sum_{k=1}^{n-2} kn \frac{C_k}{(\varepsilon^3 pn)^{k+1}} I_k$$
$$\leq \sum_{k=1}^{n-2} 2k \left(\frac{(1+\varepsilon)^2 \pi e^3 (1-p)^{(k+1)/2} S}{2\varepsilon^3 k^2} \right)^k e^{-S}$$
$$\leq \sum_{k=1}^{n-2} 2e^{-S} \left(\frac{(1+\varepsilon)^2 \pi e^3 S}{2\varepsilon^3 k^2} \right)^k$$
$$\leq 2e^{-S} \sum_{k=1}^{n-2} \left(\frac{\tilde{C}}{k^2} \right)^k = O(1),$$

với $\tilde{C} = \frac{M_{\theta,\varepsilon}(1+\varepsilon)^2\pi e^3}{2p^\theta} = const.$

Do đó:

$$\mathbb{E}(|E_{4,2}|) \leq \pi n \sum_{k=1}^{n-2} kn \int_{r=r_\varepsilon}^{R_\varepsilon} \mathbb{E}\left(|\mathcal{H}_{k,A,B}|\right) 2rdr = O(n).$$

Vậy ta thu được kết luận:

$$\mathbb{E}(|E_4|) \leq \mathbb{E}(|E_{4,1}|) + \mathbb{E}(|E_{4,2}|) = O(n). \tag{3.3.12}$$

4. **Ước lượng** $|E_1|$

Ta có $E_1 = \bigcup_{|A-B|\leq r_\varepsilon} E\left(P_{A,B}\right) := E_{1,1} \cup E_{1,2} \cup E_{1,3}$ với:

$$E_{1,1} = \bigcup_{\substack{|A-B|\leq r_\varepsilon \\ A\sim B}} E\left(P_{A,B}\right),$$

$$E_{1,2} = \bigcup_{\substack{|A-B|\leq r_\varepsilon \\ A\nsim B \\ d_{A,B}\geq(1+\varepsilon)|A-B|}} E\left(P_{A,B}\right),$$

$$E_{1,3} = \bigcup_{\substack{|A-B|\leq r_\varepsilon \\ A\nsim B \\ d_{A,B}\leq(1+\varepsilon)|A-B|}} E\left(P_{A,B}\right).$$

Đặt $|A-B| = r$. Ta sẽ đi ước lượng từng thành phần.

- Tính $|E_{1,1}|$.

  Rõ ràng khi $A \sim B$ thì cạnh AB chính là đường đi ngắn nhất từ $A \to B$, tức $E\left(P_{A,B}\right) = \{$cạnh $AB\}$. Do đó:

  $$E_{1,1} = \{\{A, B\} \in \mathcal{X}_p : |A-B| \leq r_\varepsilon\}.$$

  Ta suy ra:

  $$\mathbb{E}(|E_{1,1}|) \leq \binom{n}{2}\pi r_\varepsilon^2 p \leq \frac{M_{\theta,\varepsilon}n}{2p^\theta} = O(np^{-\theta}).$$

- Tính $|E_{1,2}|$.

  Hoàn toàn tương tự như cách tính $|E_3|$, chỉ thay $r \geq r_\varepsilon$ bởi $r \leq r_\varepsilon$. Vì vậy ta cũng thu được $\mathbb{E}(|E_{1,2}|) = O(n)$.

- Tính $|E_{1,3}|$.

  Hoàn toàn tương tự như cách tính $|E_{4,2}|$, chỉ thay $r_\varepsilon \le r \le R_\varepsilon$ bởi $r \le r_\varepsilon$, và không có điều kiện (iii). Cụ thể với $C'_k = \frac{p}{k!} \cdot \left(\frac{(1+\varepsilon)^2\pi e^2 np(1-p)^{(k+1)/2}}{2k^2}\right)^k \le p \cdot \left(\frac{(1+\varepsilon)^2\pi e^3 np(1-p)^{(k+1)/2}}{2k^3}\right)^k$ thì

$$\begin{aligned}
\mathbb{E}(|E_{1,3}|) &\le \pi n \sum_{k=1}^{n-2} kn \int_{r=0}^{r_\varepsilon} C'_k r^{2k} 2r dr \\
&\le \pi n \sum_{k=1}^{n-2} knC'_k \frac{r_\varepsilon^{2k+2}}{k+1} \qquad \left(\text{có } r_\varepsilon = \sqrt{\frac{M_{\theta,\varepsilon}}{np^{1+\theta}}}\right) \\
&\le \pi n \sum_{k=1}^{n-2} \left(\frac{\pi e^3(1+\varepsilon)^2}{k^3}\right)^k \left(\frac{M_{\theta,\varepsilon}}{p^\theta}\right)^{k+1} \\
&\le \frac{\pi M_{\theta,\varepsilon} n}{p^\theta} \sum_{k=1}^{n-2} \left(\frac{\pi e^3(1+\varepsilon)^2 M_{\theta,\varepsilon}}{k^3 p^\theta}\right)^k \\
&= O(np^{-\theta}).
\end{aligned}$$

Lấy tổng các thành phần nêu trên, ta thu được:

$$\begin{aligned}
\mathbb{E}(|E_1|) &\le \mathbb{E}(|E_{1,1}|) + \mathbb{E}(|E_{1,2}|) + \mathbb{E}(|E_{1,3}|) \\
&= O(np^{-\theta}). \qquad (3.3.13)
\end{aligned}$$

Vậy, sau tất cả các bước tính toán, sử dụng các kết quả (3.3.13), (3.3.8), (3.3.11), (3.3.12), ta kết luận:

$$\mathbb{E}(|E_\varepsilon|) = \mathbb{E}\left(\bigcup_{i=1}^{4} E_i\right) \le \sum_{i=1}^{4} \mathbb{E}(|E_i|) = O_{\varepsilon,\theta}(np^{-\theta}),$$

và kết thúc việc chứng minh định lý 3.1.

Kết luận

Các kết quả nghiên cứu chính của luận văn bao gồm:

- Tìm hiểu về bài toán kích thước của $t - bao$ trên đồ thị có trọng ngẫu nhiên. Cụ thể là xấp xỉ kích thước của $1 - bao$ của một lớp các đồ thị ngẫu nhiên có trọng là các biến ngẫu nhiên phân phối mũ.
- Tìm hiểu về bài toán kích thước của $t - bao$ trên đồ thị ngẫu nhiên có trọng hình học. Cụ thể là đưa ra thuật toán xây dựng E_ε - w.h.p là $(1 + 5\varepsilon) - bao$ của $\mathcal{X}_p$, và ước lượng kích thước của nó. Ở đây $\mathcal{X}_p$ là một mô hình đồ thị ngẫu nhiên với trọng là độ dài *Euclid*, hiểu đơn giản là nhúng $G_{n,p}$ vào $[0, 1]^2$.

Một số câu hỏi mở:

- Liệu có thể làm giảm nhẹ các điều kiện về chính quy của lớp các đồ thị ngẫu nhiên trong Chương 2? Khi đó kích cỡ của bao của đồ thị sẽ như thế nào? Các câu hỏi tương tự với phân phối khác phân phối mũ?
- Liệu có thể trả lời các câu hỏi tương tự của Chương 3 trong không gian nhiều chiều, cụ thể là với $\mathcal{X}_p^n$: tương tự $\mathcal{X}_p$ nhưng tập đỉnh được chọn ngẫu nhiên trong $[0, 1]^n$?
- Liệu có tồn tại mối liên hệ giữa p, q sao cho w.h.p $\mathcal{X}_q$ là $t - bao$ của $\mathcal{X}_p$?

Tài liệu tham khảo

[1] R. Ahmed, G. Bodwin, F. Sahneh, K. Hamm, S. Kobourov and R. Spence. *Graph Spanners: A Tutorial Review.* Computer Science Review **37**, 100–253 (2020).

[2] D. Peleg and A. A Schäffer. *Graph Spanners.* Journal of Graph Theory **13**, 99–116 (1989).

[3] A.M. Frieze and W. Pegden. *Spanners in randomly weighted graphs: independent edge lengths.* Discrete Applied Mathematics **309**, 68–74 (2022).

[4] A.M. Frieze and W. Pegden. *Spanners in randomly weighted graphs: Euclidean case.* Journal of Graph Theory **104**, 87–103 (2023).

[5] E. Dijkstra. *A note on two problems in connexion with graphs.* Numerische Mathematik **1**, 269–271 (1959).

[6] A.M. Frieze and T. Tkocz. *Shortest paths with a cost constraint: a probabilistic analysis.* Discrete Applied Mathematics **302**, 46–53 (2021).

[7] A.M. Frieze and M. Karoński. *Introduction to Random Graphs.* Cambridge University Press (2016).

[8] S. Jason. *One, two and three times* $\log n/n$ *for paths in a complete graph with random weights.* Combinatorics, Probability and Computing **8**, 347–361 (1999).

[9] M. Penrose. *Random Geometric Graphs*. Oxford University Press (2003).

[10] G.Narasimhan and M. Smid. *Geometric Spanner Networks*. Cambridge University Press (2007).

[11] E. D. Demaine and J. O'Rourke. *Open Problems from CCCG 2009.* In Proceedings of the 22nd Canadian Conference on Computational Geometry (CCCG 2010), 83-86 (2010).

[12] A. Mehrabian and N. Wormald. *On the Stretch Factor of Randomly Embedded Random Graphs.* Discrete & Computational Geometry **49**, 647-658 (2013).

[13] A.M. Frieze and W. Pegden. *Travelling in Randomly Embedded Random Graphs.* Random Structures and Algorithms **55**, 649-676 (2019).

[14] A.C. Yao. *On constructing minimum spanning trees in k-dimensional spaces and related problems.* SIAM Journal on Computing **11**, 721-736 (1982).

[15] S. Jason. *Poisson approximation for large deviations.* Random Structures and Algorithms **1**, 221-230 (1990).