

Chủ đề 07. Đồ thị ngẫu nhiên

Mặc dù lý thuyết về các đồ thị ngẫu nhiên là một trong những nhánh trẻ nhất của lý thuyết đồ thị hiện đại nhưng về tầm quan trọng thì không có bất kì một nhánh nào khác có thể sánh kịp. Nó được bắt đầu với một số bài báo lẻ tẻ của Erdős vào những năm 1940 và 1950, trong đó Erdős đã sử dụng các phương pháp ngẫu nhiên để chỉ ra sự tồn tại của các đồ thị với các tính chất dường như trái ngược nhau. Trong số các kết quả này, Erdős đã đưa ra một chặn dưới có dạng hàm mũ cho số Ramsey $R(s, s)$, nghĩa là ông ấy đã chỉ ra sự tồn tại của những đồ thị với bậc lớn sao cho cả đồ thị này cũng như phần bù của nó đều không có chứa bất kì một K_s nào. Ông ấy cũng đã chỉ ra rằng, với mọi số tự nhiên k và g , chúng ta luôn có thể tìm được các đồ thị k – chromatic với chỉ số girth $\geq g$. Như chúng ta đã thấy trong **Chủ đề 05. Phép tô màu đồ thị** và Chủ đề 06. Lý thuyết Ramsey, việc thực hiện các phép xây dựng thỏa mãn các yêu cầu của những khẳng định này không phải lúc nào cũng có thể thực hiện một cách dễ dàng. Khám phá tuyệt vời nhất của Erdős là chỉ ra rằng chúng ta có thể sử dụng các phương pháp xác suất để chứng minh sự tồn tại của các đồ thị mong muốn mà không cần thực sự xây dựng chúng. Hiện tượng này không chỉ đúng trong lý thuyết đồ thị và toán tổ hợp: các phương pháp xác suất đã được sử dụng rất thành công trong các lĩnh vực như hình học của các không gian Banach, giải tích Fourier, lý thuyết số, khoa học máy tính – đặc biệt là trong lý thuyết về các thuật toán và trong nhiều lĩnh vực khác. Tuy nhiên, không có lĩnh vực nào mà các kết quả của phương pháp xác suất xuất hiện một cách tự nhiên và nổi bật hơn các kết quả trong toán tổ hợp.

Trên thực tế, các đồ thị ngẫu nhiên cũng rất được quan tâm, không chỉ như những công cụ để giải quyết các vấn đề không liên quan gì đến lý thuyết xác suất hoặc lý thuyết về các quá trình ngẫu nhiên.

Một vấn đề được đặt ra ở đây đó là: “Các đồ thị ngẫu nhiên trong các họ khác nhau sẽ trông như thế nào?”. Thay vì quan tâm đến các **giá trị cực trị** của các tham số đặc trưng của một đồ thị, chúng ta thường chỉ quan tâm đến những gì sẽ xảy ra đối với các **giá trị trung bình**. Ngoài ra, điều làm cho lĩnh vực này trở nên hấp dẫn và quan trọng đó là những hiện tượng mà chúng ta khám phá ra thường rất bất ngờ và tinh tế.

Những nghiên cứu có hệ thống về các đồ thị ngẫu nhiên vì lợi ích của bản thân chúng được bắt đầu nghiên cứu bởi Erdős và Rényi vào những năm 1959: trong một loạt các bài báo của mình, họ đã đặt nền móng vững chắc cho một lý thuyết phong phú về các đồ thị ngẫu nhiên và chứng minh được nhiều kết quả cơ bản về chúng. Nói một cách đại khái, Erdős và Rényi đã phát hiện ra rằng trong những không gian mà họ nghiên cứu luôn có một số “loại” đồ thị ngẫu nhiên điển hình nào đó: với một xác suất đủ lớn, các đồ thị ngẫu nhiên sẽ có một số tính chất nào đó được phác họa một cách rõ nét. Một khám phá tuyệt vời khác của Erdős và Rényi là đã chỉ ra được rằng tất cả các tính chất tiêu chuẩn của các đồ thị (tính liên thông, có đường kính không vượt quá 5, có chứa một đồ thị đầy đủ bậc 4, Hamiltonian, v.v...) hóa ra lại phát sinh một cách khá đột ngột: trong khi một đồ thị ngẫu nhiên với n đỉnh và một số cạnh nhất định nào đó có thể khó có một tính chất cho trước thì một đồ thị ngẫu nhiên với n đỉnh và có thêm một vài cạnh khác rất có khả năng sẽ có tính chất cho trước đó. Hiện tượng này được gọi bằng một cụm từ được mượn từ vật lý: **hiện tượng chuyển pha**. Hầu hết những ví dụ ẩn tượng về hiện tượng chuyển pha được phát hiện bởi Erdős và Rényi đều có liên quan đến bậc của thành phần liên thông lớn nhất của một đồ thị ngẫu nhiên.

Nội dung của chủ đề này sẽ phản ánh cả hai khía cạnh của lý thuyết này: chúng ta sẽ chứng minh một số kết quả cơ bản liên quan đến các mô hình đồ thị ngẫu nhiên được nghiên cứu thường xuyên nhất và chúng ta cũng sẽ sử dụng các phương pháp xác suất để trả lời một số câu hỏi quan trọng về lý thuyết đồ thị mà không có liên quan gì đến tính ngẫu nhiên.

Với một số vấn đề, chúng ta thường sử dụng các mô hình ngẫu nhiên được thiết kế riêng. Chẳng hạn, chúng ta đã đạt những thành công đáng chú ý bằng cách sử dụng các lập luận dựa trên việc tô màu ngẫu nhiên các đồ thị. Lý thuyết thẩm thấu không gì khác hơn là các nghiên cứu về các đồ thị con ngẫu nhiên của các mạng khác nhau. Ngoài ra, nhiều thuật toán dựa trên việc sử dụng một số đồ thị nhất định, sự tồn tại của chúng sẽ được chứng minh dễ dàng nhất bằng cách sử dụng các kỹ thuật ngẫu nhiên. Tầm quan trọng của các đồ thị ngẫu nhiên và các phương pháp ngẫu nhiên chính là do các ứng dụng thuộc loại này.

Để thuận tiện, chúng ta sẽ phát biểu một số bất đẳng thức đơn giản sẽ được sử dụng trong các tính toán của chúng ta. Đối với việc gần đúng các giai thừa, chúng ta thường sẽ không bao giờ cần nhiều hơn phiên bản sau của **công thức Stirling**:

$$\sqrt{2\pi s}(s/e)^s \leq s! \leq e^{1/12s}\sqrt{2\pi s}(s/e)^s. \quad (1)$$

Trên thực tế, trong hầu hết các trường hợp, chúng ta chỉ cần $s! \geq 2\sqrt{s}(s/e)^s \geq (s/e)^s$, nghĩa là

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq \frac{1}{2\sqrt{k}} \left(\frac{en}{k}\right)^k \leq \left(\frac{en}{k}\right)^k. \quad (2)$$

Chúng ta cũng sẽ sử dụng bất đẳng thức $1 - x \leq e^{-x}$, nghĩa là

$$(1 - x)^k \leq e^{-kx}, \quad (3)$$

với mọi $x < 1$ và $k \geq 0$.

7.1. Các mô hình cơ bản

Nhiệm vụ đầu tiên của chúng ta ở đây đó là làm rõ khái niệm về một “đồ thị ngẫu nhiên”. Một cách khá tầm thường, mọi không gian xác suất mà các điểm là các đồ thị đều cung cấp cho chúng ta khái niệm về một đồ thị ngẫu nhiên. Chúng ta sẽ tập trung vào khảo sát những không gian hoặc những mô hình xác suất phát sinh một cách tự nhiên và đã được chứng minh là hữu ích nhất. Ba mô hình liên quan một cách chặt chẽ và nổi bật nhất bao gồm: $\mathcal{G}(n, M)$, $\mathcal{G}(n, p)$, $\tilde{G}^n$. Trong mỗi trường hợp, không gian xác suất đều gồm các đồ thị xác định trên một tập cố định gồm n đỉnh phân biệt: như thường lệ, chúng ta sẽ chọn tập này là $V = [n] = \{1, 2, \dots, n\}$. Lưu ý rằng đồ thị đầy đủ K_n xác định trên $[n]$ sẽ có $N = \binom{n}{2}$ cạnh và 2^N đồ thị con.

Với $0 \leq M \leq N$, không gian $\mathcal{G}(n, M)$ bao gồm tất cả $\binom{N}{M}$ đồ thị con của K_n với M cạnh: chúng ta sẽ biến $\mathcal{G}(n, M)$ thành một không gian xác suất bằng cách coi các phần tử của nó là các biến cố sơ cấp. Khi đó, bằng cách dùng $G_M = G_{n, M}$ để kí hiệu cho một đồ thị ngẫu nhiên trong không gian $\mathcal{G}(n, M)$, xác suất để G_M là một đồ thị cố định H cho trước xác định trên $[n]$ với M cạnh là $1/\binom{N}{M}$:

$$P_M(G_M = H) = \binom{N}{M}^{-1}.$$

Không gian $\mathcal{G}(n, p)$, hoặc không gian $\mathcal{G}(n, P(\text{cạnh}) = p)$ được định nghĩa với mọi $0 \leq p \leq 1$. Để có thể thu được một phần tử ngẫu nhiên của không gian này, chúng ta chỉ cần chọn ra các cạnh một cách độc lập với xác suất p . Nói cách khác, tập nền của $\mathcal{G}(n, p)$ chính là tập tất cả 2^N đồ thị xác định trên $[n]$ và xác suất của một đồ thị H xác định trên $[n]$ với m cạnh là $p^m(1 - p)^{N-m}$: mỗi một trong m cạnh của H đều được chọn và không có $N - m$ “cạnh không” nào của H được chọn. Chúng ta thường kí hiệu q thay cho $1 - p$ để

chỉ xác suất mà một cạnh của K_n không được chọn. Khi đó, bằng cách dùng $G_p = G_{n,p}$ để kí hiệu cho một phần tử ngẫu nhiên của $\mathcal{G}(n, p)$, $P_p(G_p = H) = p^{\|H\|} q^{N-\|H\|}$.

Không gian $\tilde{\mathcal{G}}^n$ không phải là một không gian của các đồ thị ngẫu nhiên mà là một không gian gồm các dãy đồ thị ngẫu nhiên, mỗi đồ thị ngẫu nhiên này đều là một phần tử nào đó của $\mathcal{G}(n, M)$. Mỗi phần tử của $\tilde{\mathcal{G}}^n$ đều là một **quá trình đồ thị**, một dãy các đồ thị lồng nhau $G_0 \subset G_1 \subset \dots \subset G_N$, trong đó G_t có chính xác t cạnh. Hiển nhiên là chúng ta sẽ có $N!$ quá trình đồ thị $\tilde{G} = (G_t)_{t=0}^N$ xác định trên $[n]$, bởi vì mọi quá trình đồ thị $\tilde{G}$ đều có thể được xác định tầm thường bằng cách đồng nhất nó với một hoán vị $(e_i)_{i=1}^N$ của N cạnh của đồ thị đầy đủ K_n xác định trên $[n]$: phép đồng nhất này được đưa ra bởi $\{e_t\} = E(G_t) - E(G_{t-1})$. Trở lại với $\tilde{\mathcal{G}}^n$, chúng ta sẽ biến tập gồm tất cả $N!$ quá trình đồ thị thành một không gian xác suất bằng cách xem tất cả các quá trình này là các biến cố sơ cấp.

Có một cách khá đơn giản để chúng ta có thể giải thích về một quá trình đồ thị ngẫu nhiên $\tilde{G} = (G_t)_{t=0}^N \in \tilde{\mathcal{G}}^n$: nó là một sinh vật sống bắt đầu cuộc đời của nó dưới dạng đồ thị rỗng $G_0 = E_n$ và tiến hóa dần bằng cách bổ sung thêm ngày càng nhiều cạnh, cụ thể, tại thời điểm t nó có thể được bổ sung thêm ngẫu nhiên một cạnh bất kì từ trong số $N - t$ khả năng.

Trong tất cả các ví dụ nêu ở trên, chúng ta đều có xu hướng quan tâm đến những gì sẽ xảy ra khi $n \rightarrow \infty$. Cần lưu ý rằng cả $M = M(n)$ và $p = p(n)$ đều là các hàm của n . Không gian $\mathcal{G}(n, p)$ cũng rất được quan tâm nghiên cứu đối với các giá trị cố định của p ; đặc biệt, $\mathcal{G}(n, 1/2)$ có thể được xem như là không gian của các đồ thị ngẫu nhiên bậc n : nó bao gồm tất cả 2^N đồ thị xác định trên $[n]$ và xem tất cả các đồ thị là các biến cố sơ cấp. Như vậy, $G_{n,1/2}$ có thể thu được bằng cách chọn ra ngẫu nhiên một đồ thị bất kì từ trong 2^N đồ thị xác định trên $[n]$. Tuy nhiên, $\mathcal{G}(n, M)$ không quá thú vị đối với một giá trị cố định của M khi $n \rightarrow \infty$, bởi vì khi đó, với xác suất tiến dần tới 1, $G_{n,M}$ chỉ đơn thuần là một tập gồm M cạnh độc lập và $n - 2M$ đỉnh cô lập.

Các không gian $\mathcal{G}(n, M)$, $\mathcal{G}(n, p)$, $\tilde{\mathcal{G}}^n$ có quan hệ mật thiết với nhau. Chẳng hạn, ánh xạ $\tilde{\mathcal{G}}^n \rightarrow \mathcal{G}(n, M)$ xác định bởi $\tilde{G} = (G_t)_{t=0}^N \mapsto G_M$ là một ánh xạ bảo toàn độ đo, nghĩa là các không gian $\tilde{\mathcal{G}}^n$ được ghép cặp với các không gian $\mathcal{G}(n, M)$, $M = 0, 1, \dots, N$. Ngoài ra, nếu, trong $\mathcal{G}(n, p)$, chúng ta đặt thêm điều kiện $\|G_p\| = M$ thì chúng ta sẽ thu được không gian $\mathcal{G}(n, M)$. Để có thể thu được $\mathcal{G}(n, p)$ từ $\tilde{\mathcal{G}}^n$, chúng ta chỉ cần chọn ra ngẫu nhiên một phần tử $\tilde{G} = (G_t)_{t=0}^N$ và lấy ra một phần tử G_t bất kì, trong đó t là một biến ngẫu nhiên nhị thức ứng với các tham số N và p , nghĩa là $P(t = M) = \binom{N}{M} p^M q^{N-M}$. Như chúng ta sẽ thấy sau đây, với $M \sim pN$, các không gian $\mathcal{G}(n, M)$ và $\mathcal{G}(n, p)$ là gần như nhau.

Như vậy, chúng ta đã có được một không gian các đồ thị ngẫu nhiên và mọi đại lượng bất biến của đồ thị đều sẽ trở thành một biến ngẫu nhiên tương ứng; bản chất của một đại lượng ngẫu nhiên như vậy sẽ chỉ phụ thuộc chủ yếu vào không gian xác suất tương ứng. Chẳng hạn, số $X_s(G)$ các đồ thị đầy đủ bậc s trong một đồ thị G là một biến ngẫu nhiên trên không gian các đồ thị ngẫu nhiên tương ứng cho dù đó là $\mathcal{G}(n, M)$, $\mathcal{G}(n, p)$ hoặc bất kì một không gian nào khác.

Trong phần này của bài viết, chúng ta sẽ giới hạn bản thân trong việc chỉ sử dụng các kỳ vọng của một số biến ngẫu nhiên cơ bản: thật đáng ngạc nhiên là ngay cả khi chỉ sử dụng một số tối thiểu các kết quả của lý thuyết xác suất như vậy vẫn cho phép chúng ta chứng minh được một số đáng kể các kết quả về lý thuyết đồ thị. Chẳng hạn, chúng ta sẽ tính toán kỳ vọng của X_s . Điều này sẽ giúp chúng ta thu được một kết quả về

chặn dưới của Erdős về các số Ramsey. Như trước đây, chúng ta cũng sẽ sử dụng các chỉ số M và p để xác định không gian mà chúng ta đang làm việc. Khi đó, $E_M(X)$ sẽ được dùng để kí hiệu kỳ vọng của biến ngẫu nhiên X trong không gian $\mathcal{G}(n, M)$.

Tương tự với $X_s(G)$, kí hiệu $X'_s(G)$ là số các tập độc lập bậc s trong một đồ thị G . Chúng ta sẽ trình bày cách tính toán các kỳ vọng của X_s và X'_s trong $\mathcal{G}(n, M)$ và $\mathcal{G}(n, p)$. Kí hiệu $\mathcal{S} = [n]^{(s)}$ là tập tất cả các s – tập con của $[n]$ và, với mọi $\alpha \in \mathcal{S}$, kí hiệu Y_α là hàm đặc trưng của đồ thị đầy đủ K_α với α là tập đỉnh:

$$Y_\alpha(G) = \begin{cases} 1 & \text{nếu } G[\alpha] = K_\alpha, \\ 0 & \text{trong trường hợp ngược lại.} \end{cases}$$

Khi đó, $X_s(G) = \sum_{\alpha \in \mathcal{S}} Y_\alpha(G)$, với mọi đồ thị G xác định trên $[n]$. Tương tự, bằng cách sử dụng Y'_α để kí hiệu hàm đặc trưng của đồ thị rỗng E_α với α là tập đỉnh, chúng ta cũng sẽ thu được $X'_s(G) = \sum_{\alpha \in \mathcal{S}} Y'_\alpha(G)$.

Như vậy, bất kể chúng ta đang sử dụng không gian xác suất nào thì, từ tính cộng tính của kỳ vọng,

$$E(X_s) = \sum_{\alpha \in \mathcal{S}} E(Y_\alpha) = \sum_{\alpha \in \mathcal{S}} P(G[\alpha] = K_\alpha),$$

và một khẳng định tương tự cũng đúng cho $E(X'_s)$.

Chúng ta sẽ sử dụng các công thức này trong các không gian $\mathcal{G}(n, M)$ và $\mathcal{G}(n, p)$:

Bắt đầu với $\mathcal{G}(n, M)$, $E_M(Y_\alpha) = P_M(G_M[\alpha] = K_\alpha) = \binom{N-S}{M-S} \binom{N}{M}^{-1}$, trong đó $S = \binom{s}{2}$, bởi vì chúng ta có $\binom{N-S}{M-S}$ đồ thị với kích thước M xác định trên $[n]$ và chứa tất cả S cạnh nối giữa các đỉnh của α . Tương tự, $E_M(Y'_\alpha) = P_M(G_M[\alpha] = E_\alpha) = \binom{N-S}{M} \binom{N}{M}^{-1}$.

Các công thức trên thậm chí còn đơn giản hơn trong $\mathcal{G}(n, p)$:

$$E_p(Y_\alpha) = P_p(G_p[\alpha] = K_\alpha) = p^S,$$

$$E_p(Y'_\alpha) = P_p(G_p[\alpha] = E_\alpha) = q^S.$$

Do $|\mathcal{S}| = \binom{s}{2}$ nên chúng ta dễ dàng sau đây kết quả sau đây:

Kết quả 7.1.1.

Kí hiệu $X_s = X_s(G)$ là số các đồ thị con đầy đủ bậc s chứa trong G và kí hiệu $X'_s = X'_s(G) = X_s(\bar{G})$.

Khi đó,

$$E_M(X_s) = \binom{n}{s} \binom{N-S}{M-S} \binom{N}{M}^{-1},$$

$$E_M(X'_s) = \binom{n}{s} \binom{N-S}{M} \binom{N}{M}^{-1},$$

$$E_p(X_s) = \binom{n}{s} p^S,$$

$$E_p(X'_s) = \binom{n}{s} q^S,$$

trong đó $S = \binom{s}{2}$ và $q = 1 - p$.

Nhận thấy rằng nếu thay vì xem xét các đồ thị con đầy đủ chúng ta đi xem xét các đồ thị con đẳng cấu với một đồ thị F cho trước nào đó thì các lập luận hầu như không có bất kì thay đổi gì. Như vậy, bằng cách sử dụng $X_F = X_F(G_p)$ để kí hiệu số các đồ thị con của G_p đẳng cấu với F , $E_p(X_F) = N_F p^{\|F\|}$, trong đó N_F là số các đồ thị con của K_n đẳng cấu với F .

Lưu ý rằng chúng ta cũng có thể thu được đẳng thức $E_p(X_F) = N_F p^{\|F\|}$ như trên bằng cách liệt kê ra dãy các đồ thị con $F_1, F_2, \dots, F_{N_F}$ của K_n đẳng cấu với F và dùng Y_i để kí hiệu cho hàm đặc trưng của F_i , nghĩa là $Y_i(G_p) = 1$ nếu $F_i = G_p$ và $Y_i(G_p) = 0$ nếu $F_i \neq G_p$. Khi đó, $X_F = \sum_{i=1}^{N_F} Y_i$ và phép lấy tổng của chúng ta một lần nữa lại được phân phối giống hệt các biến ngẫu nhiên 0 – 1. Như vậy, từ tính cộng tính của kỳ vọng,

$$E_p(X_F) = \sum_{i=1}^{N_F} E(Y_i) = \sum_{i=1}^{N_F} P_p(Y_i = 1) = N_F p^{\|F\|},$$

như đã tuyên bố ở trên.

Số N_F có quan hệ mật thiết với nhóm các tự đẳng cấu của F , nghĩa là nhóm các hoán vị của các đỉnh của F bảo toàn tính liên hợp giữa các đỉnh. Thật vậy, nếu F có k đỉnh và nhóm các tự đẳng cấu của nó có bậc a , nghĩa là K_k có $k!/a$ đồ thị con đẳng cấu với F , thì $N_F = \binom{n}{k} \frac{k!}{a} = \frac{(n)_k}{a}$, trong đó $(n)_k$ là **giai thừa lùi bậc k** : $(n)_k = n(n-1) \dots (n-k+1)$.

Như vậy, trong trường hợp này,

$$E_p(X_F) = \frac{(n)_k}{a} p^{\|F\|}.$$

Ngoài ra, nếu F là một k – chu trình C_k thì

$$E_p(X_{C_k}) = \frac{(n)_k}{2k} p^k.$$

Các công thức tương tự cũng đúng cho số các đồ thị con cảm sinh: chẳng hạn, bằng cách sử dụng Y_{C_k} để kí hiệu cho số các k – chu trình cảm sinh,

$$E_p(Y_{C_k}) = \frac{(n)_k}{2k} p^k q^{\binom{k}{2}-k} = \frac{(n)_k}{2k} p^k q^{k(k-3)/2}.$$

Chúng ta sẽ quay trở lại với nội dung chính của chủ đề này. Kết quả 7.1.1) đơn giản là tất cả những gì Erdős cần để có thể thu được một chặn dưới có dạng hàm mũ cho các số Ramsey $R(s, t)$.

Kết quả 7.1.2. (Erdős)

a) Nếu $3 \leq s \leq n$ sao cho $\binom{n}{s} < 2^{\binom{s}{2}-1}$ thì $R(s, s) \geq n + 1$. Hơn nữa, $R(s, s) > \frac{1}{e\sqrt{2}} s 2^{s/2}$.

b) Giả sử cho trước $3 \leq s \leq t \leq n$ và $0 < p < 1$ sao cho $\binom{n}{s} p^{\binom{s}{2}} + \binom{n}{t} p^{\binom{t}{2}} < 1$, trong đó $q = 1 - p$.

Khi đó, $R(s, t) \geq n + 1$.

Chứng minh kết quả 7.1.2.

i) Xét không gian $\mathcal{G}(n, 1/2)$. Với các ký hiệu nêu ở trên, $E_{1/2}(X_s + X'_s) = 2\binom{n}{s}2^{-(s)} < 1$, nghĩa là chúng ta luôn có thể tìm được một $G \in \mathcal{G}(n, 1/2)$ sao cho $(X_s + X'_s)(G) = X_s(G) + X'_s(G) = 0$. Điều này chỉ ra rằng cả G và phần bù $\bar{G}$ của nó đều không có chứa bất kì một đồ thị đầy đủ bậc s nào.

Như vậy, $R(s, s) \geq n + 1$.

Bất đẳng thức $R(s, s) > \frac{1}{e\sqrt{2}}s2^{s/2}$ là hệ quả tức thì của bất đẳng thức $R(s, s) \geq n + 1$ và (1).

Thật vậy, với $n = \left\lfloor \frac{s2^{s/2}}{e\sqrt{2}} \right\rfloor$, từ (1), chúng ta dễ dàng suy ra rằng

$$\binom{n}{s}2^{-(s)}+1 < \frac{n^s}{s!}2^{-(s)}+1 < \frac{(e\sqrt{2})^{-s}s^s2^{s^2/2}}{\sqrt{2\pi s}(s/e)^s}2^{-(s)}+1 = \frac{2}{\sqrt{2\pi s}} < 1,$$

nghĩa là $R(s, s) \geq n + 1$.

ii) Khẳng định này chỉ là một dạng biến thể nhỏ của khẳng định đầu tiên: từ giả thiết của bài toán, chúng ta sẽ có $E_p(X_s + X'_t) < 1$, nghĩa là $X_s(G) = X'_t(G) = 0$ với một đồ thị $G \in \mathcal{G}(n, p)$ nào đó. Điều này có nghĩa là đồ thị G không có chứa bất kì một đồ thị đầy đủ bậc s nào và phần bù $\bar{G}$ của nó không có chứa bất kì một đồ thị bậc t nào. Do G có n đỉnh nên $R(s, t) \geq n + 1$. $\square$

Lập luận trên cũng có thể được áp dụng cho không gian $\mathcal{G}(n, M)$ thay vì $\mathcal{G}(n, p)$: nếu $E_M(X_s) + E_M(X'_t) < 1$ thì $R(s, t) \geq n + 1$. Trên thực tế, theo cách này, chúng ta thu được một kết quả tốt hơn đôi chút. Chẳng hạn, giả sử N là một số chẵn, chúng ta dễ dàng nhận thấy rằng, với $M = N/2$, $E_M(X_s) \leq E_{1/2}(X_s)$. Tuy nhiên, sự cải thiện này là không đáng kể và các tính toán trong $\mathcal{G}(n, p)$ đẹp hơn đáng kể so với trong $\mathcal{G}(n, M)$.

Sau khi chứng kiến tính đơn giản đáng kinh ngạc của phép chứng minh của Kết quả 7.1.2), chúng ta hoàn toàn có thể tin rằng, với việc thực hiện thêm một số công việc nữa, chúng ta có thể cải thiện lại chặn dưới $cs2^{s/2}$, trong đó c là một hằng số. Trên thực tế, có vẻ như không phải là như vậy: mặc dù hằng số $c = 1/(e\sqrt{2})$ có thể được cải thiện lại thành hằng số $c = \sqrt{2}/e$ bằng cách sử dụng một lập luận ứng dụng đơn giản của bổ đề địa phương của Lovász sẽ không được thảo luận trong bài viết này nhưng thậm chí chúng ta cũng còn chưa biết được liệu số mũ 1 của s có thể được cải thiện hay không. Như vậy, từ chặn trên Erdős – Szekeres (Kết quả 6.1.1) và chặn dưới Erdős (Kết quả 7.1.2), chúng ta nhận thấy rằng

$$2^{s/2} \leq R(s, s) \leq 2^{2s},$$

và, tại thời điểm này, $1/2$ và 2 là các hằng số tốt nhất trong bất đẳng thức trên. Rất có khả năng là trên thực tế chúng ta sẽ có $R(s, s) = 2^{(c+o(1))s}$ với một hằng số c nào đó và cũng rất có thể là $c = 1$, nhưng một bằng chứng về điều này dường như là rất xa trong tương lai.

Liên quan đến các số Ramsey $R(s, t)$, chúng ta đặc biệt quan tâm đến việc xác định bậc của $R_s(t) = R(s, t)$ khi s được giữ cố định và $t \rightarrow \infty$. Sau nhiều thập kỷ cải tiến, ngày nay chúng ta biết rằng

$$\frac{c_1 t^2}{\log t} \leq R(3, t) \leq \frac{c_2 t^2}{\log t}, \text{ với } c_1 \text{ và } c_2 \text{ là các hằng số dương nào đó.}$$

Chặn trên đã được chứng minh bởi Shearer vào năm 1983 bằng cách cải tiến lại một phương pháp chứng minh trước đó của Ajtai, Komlós, Szemerédi trong khi chặn dưới đã được chứng minh bởi Kim vào năm 1995 bởi một lập luận xác suất có phần khéo léo và phức tạp. Như là một trong những ứng dụng nổi bật đầu tiên của phương pháp xác suất, Erdős đã chỉ ra hơn 30 năm trước đó rằng $\frac{c_1 t^2}{\log^2 t}$ là một chặn dưới của $R(3, t)$.

Nhận thấy rằng chỉ cần sử dụng một lập luận phức tạp hơn một chút chúng ta có thể đưa ra một chặn dưới cho bài toán Zarankiewicz. Chúng ta sẽ sử dụng một dạng tương tự của mô hình $\mathcal{G}(n, M)$ chứ không phải một dạng tương tự của $\mathcal{G}(n, p)$ một phần vì để đảm bảo tính đa dạng của nội dung và cũng vì nó làm cho phần thứ hai của lập luận trở nên đơn giản hơn một chút.

Kết quả 7.1.3.

Giả sử cho trước $2 \leq s \leq n_1$, $2 \leq t \leq n_2$, $\alpha = (s-1)/(st-1)$, $\beta = (t-1)/(st-1)$.

Khi đó, chúng ta luôn có thể tìm được một đồ thị lưỡng phân $G_2(n_1, n_2)$ với kích thước

$$\left\lfloor \left(1 - \frac{1}{s! t!}\right) n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor$$

mà không có chứa bất kì một đồ thị $K(s, t)$ nào (với s đỉnh nằm trong lớp đầu tiên và t đỉnh nằm trong lớp thứ hai).

Chứng minh kết quả 7.1.3.

Kí hiệu:

$$n = n_1 + n_2,$$

$$V_1 = \{1, 2, \dots, n_1\},$$

$$V_2 = \{n_1 + 1, n_1 + 2, \dots, n_1 + n_2\},$$

$$E = \{ij : i \in V_1, j \in V_2\},$$

$$M = \left\lfloor n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor.$$

Chúng ta sẽ xem xét không gian xác suất $\mathcal{G}(K_{n_1, n_2}, M)$ gồm $\binom{|E|}{M}$ đồ thị với tập đỉnh $V = V_1 \cup V_2$, có chính xác M cạnh nằm trong E và không có bất kì một cạnh nào nằm ngoài E . (Lưu ý rằng đây không phải là không gian xác suất được giới thiệu trong các kết quả trước đó.)

Khi đó, kỳ vọng của số các đồ thị con $K_{s,t}$ được chứa trong một đồ thị $G \in \mathcal{G}(K_{n_1, n_2}, M)$ là

$$E_{s,t} = \binom{n_1}{s} \binom{n_2}{t} \binom{|E| - st}{M - st} \binom{|E|}{M}^{-1},$$

trong đó nhân tử đầu tiên là số cách chọn lớp đầu tiên của $K_{s,t}$, nhân tử thứ hai là số cách chọn lớp thứ hai của $K_{s,t}$, nhân tử thứ ba là số cách chọn $M - st$ cạnh nằm ngoài $K_{s,t}$.

Kế tiếp, do

$$\binom{|E| - st}{M - st} \binom{|E|}{M}^{-1} = \prod_{i=0}^{st-1} \frac{M - i}{n_1 n_2 - i} < \left(\frac{M}{n_1 n_2} \right)^{st},$$

nên

$$E_{s,t} < \frac{1}{s! t!} n_1^s n_2^t \left(\frac{M}{n_1 n_2} \right)^{st} \leq \frac{1}{s! t!} n_1^s n_2^t (n_1^{-\alpha} n_2^{-\beta})^{st} = \frac{1}{s! t!} n_1^{1-\alpha} n_2^{1-\beta}.$$

Như vậy, chúng ta luôn có thể tìm được một đồ thị $G_0 \in \mathcal{G}(K_{n_1, n_2}, M)$ chứa ít hơn $n_1^{1-\alpha} n_2^{1-\beta} / s! t!$ đồ thị lưỡng phân đầy đủ $K_{s,t}$. Loại bỏ một cạnh từ mỗi đồ thị $K_{s,t}$ trong G_0 . Đồ thị thu được $G = G_2(n_1, n_2)$ chứa ít nhất

$$\left\lfloor n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor - \left\lfloor \frac{1}{s! t!} n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor \geq \left\lfloor \left(1 - \frac{1}{s! t!} \right) n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor \text{ cạnh}$$

và không có chứa bất kì một đồ thị $K(s, t)$ nào. □

Bằng cách sử dụng phương pháp tương tự, chúng ta cũng có thể xây dựng được một đồ thị với bậc n và kích thước

$$\left\lfloor \frac{1}{2} \left(1 - \frac{1}{s! t!} \right) n^{2-(s+t-2)/(st-1)} \right\rfloor$$

và không có chứa bất kì một đồ thị $K_{s,t}$ nào.

Rất có khả năng là chặn dưới nêu trong Kết quả 7.1.3) khác khá xa so với sự thật: chẳng hạn, với $s = t$, nó chỉ ra rằng $z(n, n; t, t) \geq cn^{2-2/(t+1)}$ trong khi chúng ta luôn kỳ vọng rằng nó sẽ trở thành một chặn trên trong Kết quả 4.2.6), cụ thể là $cn^{2-\frac{1}{t}}$ sẽ là bậc đúng của $z(n, n; t, t)$. Tuy nhiên, với một giá trị cố định đủ lớn của $s = t$, Kết quả 7.1.3) về cơ bản là chặn trên tốt nhất tại thời điểm hiện tại.

Như là một ví dụ cuối cùng của chúng ta về việc ứng dụng các đồ thị ngẫu nhiên vào các bài toán trọng tâm của lý thuyết đồ thị, chúng ta sẽ trình bày một định lí của Erdős về sự tồn tại của các đồ thị với chỉ số girth và sắc số lớn.

Kết quả 7.1.4. (Erdős)

Với các số tự nhiên $g \geq 3$ và $k \geq 2$, chúng ta luôn có thể tìm được một đồ thị với bậc k^{3g} , chỉ số girth $\geq g$ và sắc số $\geq k$.

Chứng minh kết quả 7.1.4.

Không giảm tính tổng quát của vấn đề, chúng ta có thể giả định rằng $g \geq 4$ và $k \geq 4$, bởi vì trong trường hợp này tính đúng đắn của khẳng định là hiển nhiên. Kí hiệu $n = k^{3g}$, $p = 2k^{2-3g} = 2k^2/n$. Chúng ta sẽ xem xét không gian $\mathcal{G}(n, p)$. Bằng cách dùng $Z_l = Z_l(G_p)$ để kí hiệu cho số các l – chu trình trong đồ thị ngẫu nhiên G_p của chúng ta, từ đẳng thức $E_p(X_{C_k}) = \frac{(n)_k}{2k} p^k$, chúng ta nhận thấy rằng

$$E_p(Z_l) = \frac{(n)_l}{2l} p^l < \frac{(np)^l}{2l} \frac{2^l k^{2l}}{2l}.$$

Như vậy, kỳ vọng của số các chu trình với chiều dài $\leq g-1$ là

$$\sum_{l=3}^{g-1} E_p(Z_l) < \sum_{l=3}^{g-1} 2^l \frac{k^{2l}}{2l} < \frac{2^{g-1} k^{2g-2}}{3},$$

trong đó bất đẳng thức cuối cùng là khá thô.

Kí hiệu Ω_1 là tập các đồ thị trong $\mathcal{G}(n, p)$ mà chứa $\leq f = 2^{g-1} k^{2g-2}$ chu trình có chiều dài $< g$. Do $\sum_{l=3}^{g-1} E_p(Z_l) \geq P_p(\overline{\Omega_1})f = (1 - P_p(\Omega_1))f$, trong đó $\overline{\Omega_1} = \mathcal{G}(n, p) \setminus \Omega_1$, nên chúng ta nhận thấy rằng

$$P_p(\Omega_1) > \frac{2}{3}.$$

Kí hiệu $s = n/k = k^{3g-1}$ và Ω_2 là tập các đồ thị trong $\mathcal{G}(n, p)$ không có chứa bất kì một tập gồm s đỉnh nào mà cảm sinh ra $\leq f$ cạnh. Lưu ý rằng khẳng định trên sẽ được chứng minh là đúng nếu chúng ta có thể chỉ ra rằng $\Omega_1 \cap \Omega_2 \neq \emptyset$. Thật vậy, giả sử tồn tại một đồ thị $G_0 \in \Omega_1 \cap \Omega_2$. Trong G_0 , bằng cách xóa đi một cạnh bất kỳ khỏi mỗi chu trình có chiều dài $< g$, chúng ta sẽ nhận được một đồ thị G với chỉ số girth $> g$. Do $G_0 \in \Omega_1$ nên tối đa không vượt quá l cạnh đã bị chúng ta xóa đi. Đồng thời, do $G_0 \in \Omega_2$ nên mọi tập gồm s đỉnh đều sẽ cảm sinh ra tối thiểu là $f+1$ cạnh. Như vậy, trong G , mọi tập gồm s đỉnh đều sẽ cảm sinh ra tối thiểu một cạnh, nghĩa là mọi tập đỉnh cô lập đều chứa tối đa không vượt quá $s-1$ đỉnh: $\alpha(G) \leq s-1$.

Từ bất đẳng thức $\chi(G) \geq \max\{\omega(G), |G|/\alpha(G)\}$ của **Chủ đề 05. Phép tô màu đồ thị**, chúng ta có $\alpha(G)\chi(G) \geq n$. Điều này chỉ ra rằng $\chi(G) > k$.

Kế tiếp, khẳng định $\Omega_1 \cap \Omega_2 \neq \emptyset$ sẽ được chứng minh nếu chúng ta có thể chỉ ra rằng $P_p(\Omega_1 \cap \Omega_2) > \frac{1}{3}$.

Trên thực tế, chúng ta sẽ hào phóng hơn và chúng ta sẽ chỉ ra rằng $P_p(\Omega_2)$ là rất gần 1. Cụ thể, chúng ta sẽ chỉ ra rằng $P_p(\Omega_2) > \frac{2}{3}$, nên $P_p(\Omega_1 \cap \Omega_2) \geq P_p(\Omega_1) + P_p(\Omega_2) - 1 > \frac{1}{3}$.

Với $G_p \in \mathcal{G}(n, p)$ và $l \geq 0$, kí hiệu $I_l(G_p)$ là số các s -tập đỉnh cảm sinh ra chính xác l cạnh. Khi đó, $I_l(G_p)$ là số các s -tập đỉnh cô lập nhưng có chính xác l cạnh. Bằng cách kí hiệu $I(G_p) = \sum_{l=0}^f I_l(G_p)$, chúng ta sẽ có

$$\Omega_2 = \{G_p \in \mathcal{G}(n, p) : I(G_p) = 0\}.$$

Như vậy,

$$\begin{aligned} E_p(I) &= \sum_{m=0}^{\infty} P_p(I = m)m \geq \sum_{m=1}^{\infty} P_p(I = m) = P_p(I \geq 1) \\ &= 1 - P_p(I = 0) = 1 - P(\Omega_2), \end{aligned}$$

nên chúng ta chỉ cần chứng minh $E_p(I) < 1/3$ là đủ.

Đây chỉ là một bài toán ước tính đơn giản.

Thật vậy, với $S = \binom{s}{2}$,

$$E_p(I_l) = \binom{n}{s} \binom{S}{l} p^l (1-p)^{S-l},$$

bởi vì chúng ta có $\binom{n}{s}$ cách chọn các s -tập và $\binom{S}{l}$ cách chọn cho l cạnh được cảm sinh bởi s -tập này; bằng cách chọn l cạnh này và giả định rằng không có cạnh nào khác nối giữa các đỉnh trong s -tập này, chúng ta sẽ nhận được nhân tử $p^l (1-p)^{S-l}$. Lưu ý rằng, từ (3), $(1-p)^{S-l}$ tối đa không vượt quá $e^{-p(S-l)}$.

Đễ dàng nhận thấy rằng $E_p(I_l)/E_p(I_{l+1}) < 1/2$ với mọi $0 \leq l < f$. Nhắc lại rằng $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$, $s = n/k = k^{3g-1}$, $S = s^2/2 - s/2 \leq s^2/2$, $p = 2k^2/n$, $ps = 2k$, $ps^2 = 2n = 2k^{3g}$, $f = 2^{g-1}k^{2g-2}$, $pf = 2^g k^{-g}$, chúng ta dễ dàng tìm được rằng

$$\begin{aligned} E_p(I) &= \sum_{l=0}^f E_p(I_l) \leq 2E_p(I_f) \\ &\leq \left(\frac{en}{s}\right)^s \left(\frac{eS}{f}\right)^f p^f e^{-pS+pf} \\ &< (ek)^{n/k} \left(\frac{eps^2}{2f}\right)^f e^{-ps^2/2+ps/2+pf} \\ &< (ek)^{n/k} k^{(g+2)f} e^{-n} \\ &= (ek)^{n/k} \left(\frac{ek^{g+2}}{2^{g-1}}\right)^f e^{-n+k+2^g k^{-g}}. \end{aligned}$$

Trong bước cuối cùng ở trên, chúng ta sẽ sử dụng đánh giá $f \geq 8k^6$ và nhận thấy rằng

$$(e/2^{g-1})^f e^{k+2^g/k^{-g}} < e^{k+1-f} < 1.$$

Như vậy,

$$\log E_p(I) < -n \left\{ 1 - \frac{1 + \log k}{k} - \frac{(g+2)2^{g-1} \log k}{k^{g+2}} \right\} < -n/4,$$

nên

$$E_p(I) < e^{-n/4} < 1/3. \quad \square$$

Kết quả 7.1.4) đặt ra cho chúng ta một câu hỏi hóc búa mà tính đến thời điểm hiện tại vẫn chưa có câu trả lời: “Độ lớn của giá trị $n_1(g, k)$, bậc nhỏ nhất có thể có của một đồ thị với chỉ số girth $\geq g$ và sắc số $\geq k$, vào khoảng bao nhiêu?”. Do một đồ thị với sắc số $\geq k$ đều có chứa một đồ thị con với bậc nhỏ nhất $\geq k-1$ nên chúng ta dễ dàng suy ra rằng $n_1(g, k) \geq n_0(g, k-1)$, trong đó $n_0(g, \delta)$ là hàm nêu trong Kết quả 4.1.1). Như vậy, $n_1(g, k)$ nằm trong khoảng giữa $k^{g/2}$ và k^{3g} . Với việc thực hiện thêm một số công việc nữa, chúng ta hoàn toàn có thể làm giảm chặn trên đi một chút nữa, nhưng có vẻ như việc xác định $\lim_{g, k \rightarrow \infty} \frac{\log n_1(g, k)}{g \log k}$, nếu giới hạn này tồn tại, là một việc rất khó có thể thực hiện được.

7.2. Các tính chất cơ bản của hầu hết tất cả các đồ thị - Kỹ thuật sử dụng kỳ vọng

Trong mục 7.1. Các mô hình cơ bản, chúng ta đã nhận thấy được những lợi ích của việc biết được hầu hết các đồ thị trong một mô hình xác suất đều có một tính chất nhất định nào đó. Kế tiếp, chúng ta sẽ tiến thêm một bước nữa; cụ thể, chúng ta sẽ thảo luận về các tính chất được chia sẻ bởi hầu hết các đồ thị. Giả sử cho trước một tính chất Q , chúng ta sẽ nói rằng ***hầu hết các đồ thị trong một không gian xác suất Ω_n gồm các đồ thị bậc n có tính chất Q*** nếu $P(G \in \Omega_n: G \text{ có tính chất } Q) \rightarrow 1$ khi $n \rightarrow \infty$. Trong phần này của bài viết, chúng ta sẽ luôn chọn $\Omega_n = \mathcal{G}(n, p)$, trong đó $0 < p < 1$ có thể phụ thuộc vào n .

Đầu tiên, chúng ta sẽ giả định rằng $0 < p < 1$ là cố định, nghĩa là p độc lập với n .

Nhận thấy rằng chúng ta có nhiều tính chất đơn giản đúng với hầu hết mọi đồ thị trong $\mathcal{G}(n, p)$. Chẳng hạn, nếu H là một đồ thị cố định cho trước bất kỳ thì hầu hết mọi đồ thị $G_p \in \mathcal{G}(n, p)$ đều có chứa H như một đồ thị con cảm sinh. Thật vậy, nếu $|H| = h$ thì xác suất để một đồ thị con của G được cảm sinh bởi một tập đỉnh gồm h đỉnh cho trước đẳng cấu với H là một số thực dương, tạm giả sử là $r > 0$. Do $V(G)$ có chứa $\lfloor n/h \rfloor$ tập con phân biệt của mỗi h đỉnh này nên xác suất để không có bất kỳ một đồ thị con cảm sinh nào của G đẳng cấu với H tối đa không vượt quá $(1 - r)^{\lfloor n/h \rfloor}$, giá trị này có xu hướng tiến tới 0 khi $n \rightarrow \infty$. Kết quả sau đây là một phiên bản mạnh hơn của nhận định này:

Kết quả 7.2.1.

Giả sử cho trước các số tự nhiên cố định $1 \leq h \leq k$ và một số thực dương cố định $0 < p < 1$.

Khi đó, trong $\mathcal{G}(n, p)$, với mọi dãy k đỉnh $x_1, x_2, \dots, x_k$, hầu hết các đồ thị G_p đều có chứa một đỉnh x sao cho $xx_i \in E(G_p)$ nếu $1 \leq i \leq h$ và $xx_i \notin E(G_p)$ nếu $h < i \leq k$.

Chứng minh kết quả 7.2.1.

Kí hiệu $x_1, x_2, \dots, x_k$ là một dãy gồm k đỉnh nào đó.

Khi đó, xác suất để tồn tại một đỉnh $x \in W = V(G) - \{x_1, x_2, \dots, x_k\}$ sẽ thỏa mãn các tính chất cần thiết là $p^h q^{k-h}$. Do, với mọi $x, y \in W, x \neq y$, các cạnh xx_i được chọn độc lập với các cạnh yx_i nên xác suất không thể tìm thấy bất kỳ một đỉnh x nào phù hợp cho dãy các đỉnh cụ thể này là $(1 - p^h q^{k-h})^{n-k}$. Hơn nữa, do có $(n)_k = n(n-1) \dots (n-k+1)$ cách chọn cho dãy các đỉnh $x_1, x_2, \dots, x_k$ nên xác suất để chúng ta có thể tìm được một dãy các đỉnh $x_1, x_2, \dots, x_k$ sao cho không thể tìm được bất kỳ một đỉnh x nào phù hợp cho dãy các đỉnh này tối đa không vượt quá

$$\varepsilon = n^k (1 - p^h q^{k-h})^{n-k}.$$

Như vậy, $\varepsilon \rightarrow 0$ khi $n \rightarrow \infty$. □

Từ một kết quả của Gaifman về các mệnh đề logic bậc nhất, Kết quả 7.2.1) ngụ ý rằng: ***với mọi xác suất $0 < p < 1$ cố định cho trước, mọi mệnh đề logic bậc nhất về các đồ thị sẽ là đúng với hầu hết các đồ thị $G \in \mathcal{G}(n, p)$ hoặc là sai với hầu hết các đồ thị $G \in \mathcal{G}(n, p)$*** . Mặc dù kết quả này thoạt nhìn có vẻ khá phức tạp nhưng trên thực tế là nó yếu hơn so với Kết quả 7.2.1). Thật vậy, với mọi mệnh đề logic bậc nhất, Kết quả 7.2.1) cho phép chúng ta đưa ra kết luận mà không cần phải tốn bất kỳ một nỗ lực nào xem xét liệu mệnh đề đó có đúng với hầu hết các đồ thị hoặc nó có sai với hầu hết các đồ thị hay không. Hơn nữa, các khẳng

định sau đây liên quan đến mô hình $\mathcal{G}(n, p)$, với mọi xác suất $p \in (0, 1)$ cho trước, sau đây đều là hệ quả trực tiếp của Kết quả 7.2.1):

1. Với mọi số nguyên k cố định cho trước, hầu hết các đồ thị $G_{n,p}$ đều có bậc nhỏ nhất $\geq k$.
2. Hầu hết mọi đồ thị $G_{n,p}$ đều có đường kính bằng 2.
3. Với mọi đồ thị H cố định cho trước, hầu hết các đồ thị $G_{n,p}$, bất cứ khi nào $F_0 \subset G_{n,p}$ đẳng cấu với một đồ thị con F của H , đều có chứa một đồ thị con H_0 đẳng cấu với H sao cho $F_0 \subset H_0 \subset G_{n,p}$.

Một điều khá tự nhiên, hầu hết các mệnh đề mà chúng ta quan tâm sẽ không phải là các mệnh đề bậc nhất: “**Với mọi số thực dương $\varepsilon > 0$, hầu hết mọi đồ thị $G_{n,p}$ đều chứa $\geq \frac{1}{2}(p - \varepsilon)n^2$ cạnh và $\leq \frac{1}{2}(p + \varepsilon)n^2$ cạnh**”. “**Hầu hết không có bất kì một đồ thị $G_{n,p}$ nào có thể được tô bởi $n^{\frac{1}{2}}$ màu**”. “**Hầu hết mọi đồ thị $G_{n,p}$ đều có chứa một đồ thị đầy đủ với bậc $\log n / \log(1/p)$** ”. “Với mọi số thực dương $\varepsilon > 0$, hầu hết mọi đồ thị $G_{n,p}$ đều là $\frac{1}{2}(p - \varepsilon)n -$ liên thông”. Những khẳng định này đều đúng với mọi p cố định và đều có thể được chứng minh một cách dễ dàng. Tuy nhiên, không có bất kì một khẳng định nào trong số chúng là một mệnh đề bậc nhất, bởi vì chúng đều có liên quan đến các tập con lớn của các đỉnh.

Kế tiếp, chúng ta sẽ xem xét mô hình $\Omega = \mathcal{G}(n, p)$ với giả định $0 < p < 1$ phụ thuộc vào n , nhưng $pn^2 \rightarrow \infty$, $(1 - p)n^2 \rightarrow \infty$ khi $n \rightarrow \infty$. Trong trường hợp này, với mọi số nguyên m cố định cho trước, hầu hết mọi đồ thị G_p đều có $\|G_p\| \geq m$ và $\|\overline{G_p}\| \geq m$. Như trước đây, chúng ta sẽ kí hiệu $N = \binom{n}{2}$ và, với mọi $M = 0, 1, \dots, N$, chúng ta sẽ dùng Ω_M để kí hiệu cho tập các đồ thị trong $\mathcal{G}(n, M)$. Hiển nhiên, $\Omega = \bigcup_{M=0}^N \Omega_M$ và các phần tử của Ω_M đều có cùng một xác suất trong cả $\mathcal{G}(n, M)$ và $\mathcal{G}(n, p)$.

Chúng ta sẽ chỉ ra rằng các mô hình $\Omega = \mathcal{G}(n, p)$ và $\mathcal{G}(n, M)$ là rất gần nhau, với điều kiện là M xấp xỉ với pN – kỳ vọng của số các cạnh của một đồ thị trong Ω .

Hiển nhiên,

$$P_p(\Omega_M) = P_p(\|G_p\| = M) = \binom{N}{M} p^M q^{N-M}.$$

Như vậy,

$$\frac{P_p(\Omega_M)}{P_p(\Omega_{M+1})} = \frac{M+1}{N-M} \frac{q}{p}.$$

Điều này cho thấy rằng $P_p(\Omega_M)/P_p(\Omega_{M+1})$ là một hàm tăng theo M và $P_p(\Omega_M)$ sẽ đạt cực đại với một số M nào đó thỏa mãn $pN - p \leq M \leq pN + q$. Hơn nữa, nếu $0 < \varepsilon < 1$ và n đủ lớn thì do $pn^2 \rightarrow \infty$ khi $n \rightarrow \infty$ nên,

$$\frac{P_p(\Omega_M)}{P_p(\Omega_{M+1})} < 1 - \varepsilon,$$

với điều kiện là $M < (1 - \varepsilon)pN$; đồng thời, do $(1 - p)n^2 \rightarrow \infty$ khi $n \rightarrow \infty$ nên,

$$\frac{P_p(\Omega_{M+1})}{P_p(\Omega_M)} < (1 + \varepsilon)^{-1},$$

với điều kiện là $M > (1 + \varepsilon)pN$.

Kí hiệu $N_\varepsilon = \lfloor (1 + \varepsilon)pN \rfloor$ và $N_{-\varepsilon} = \lceil (1 - \varepsilon)pN \rceil$, chúng ta cũng nhận thấy từ các bất đẳng thức này rằng hầu hết mọi đồ thị G_p đều thỏa mãn $N_{-\varepsilon} \leq \|G_p\| \leq N_\varepsilon$; nghĩa là,

$$P_p \left(\bigcup_{M=N_{-\varepsilon}}^{N_\varepsilon} \Omega_M \right) \rightarrow 1 \text{ khi } n \rightarrow \infty.$$

Một hệ quả khác của hệ thức $\frac{P_p(\Omega_M)}{P_p(\Omega_{M+1})} = \frac{M+1}{N-M} \frac{q}{p}$ là chúng ta luôn có thể tìm được một số thực dương $\eta > 0$ (trên thực tế, bất kỳ giá trị $0 < \eta < \frac{1}{2}$ nào cũng được) sao cho

$$P_p \left(\bigcup_{M=0}^{\lfloor pN \rfloor} \Omega_M \right) > \eta,$$

nếu n đủ lớn.

Kế tiếp, các khẳng định trên ngụ ý rằng nếu tồn tại $\Omega^* \subset \Omega$ sao cho $P_p(\Omega^*) \rightarrow 1$ khi $n \rightarrow \infty$ thì, với mọi $\varepsilon > 0$, chúng ta luôn có thể tìm được các chỉ số M_1 và M_2 sao cho $(1 - \varepsilon)pN \leq M_1 \leq pN \leq M_2 \leq (1 + \varepsilon)pN$ và

$$\frac{|\Omega_{M_i} \cap \Omega^*|}{|\Omega_{M_i}|} \rightarrow 1 \text{ khi } n \rightarrow \infty \text{ (} i = 1, 2 \text{)}.$$

Chúng ta sẽ gọi một tập $\Omega^* \subset \Omega$ là **lỗi** nếu $G \in \Omega^*$ bất cứ khi nào $G_1 \subset G \subset G_2$ và $G_1, G_2 \in \Omega^*$; một **tính chất lỗi** của các đồ thị cũng được định nghĩa tương tự. Dễ dàng nhận thấy rằng, với mọi tập lỗi Ω^* , hệ thức trên ngụ ý rằng

$$\frac{|\Omega_M \cap \Omega^*|}{|\Omega_M|} \rightarrow 1 \text{ khi } n \rightarrow \infty,$$

bất cứ khi nào $M_1 \leq M \leq M_2$ và, đặc biệt, nếu $M = \lfloor pN \rfloor$.

Chúng ta cũng sẽ trình bày lại khẳng định trên như một định lý về mối liên hệ giữa các mô hình $\mathcal{G}(n, p)$ và $\mathcal{G}(n, M)$:

Kết quả 7.2.2.

Kí hiệu $0 < p = p(n) < 1$ sao cho $pn^2 \rightarrow \infty$ và $(1 - p)n^2 \rightarrow \infty$ khi $n \rightarrow \infty$ và Q là một tính chất của các đồ thị.

(i) Giả sử cho trước một số thực dương cố định $\varepsilon > 0$ và nếu $(1 - \varepsilon)pN < M < (1 + \varepsilon)pN$ thì hầu hết mọi đồ thị trong $\mathcal{G}(n, M)$ đều có tính chất Q .

Khi đó, hầu hết mọi đồ thị trong $\mathcal{G}(n, p)$ cũng đều có tính chất Q .

(ii) Nếu Q là một tính chất lỗi và hầu hết mọi đồ thị trong $\mathcal{G}(n, P(\text{cạnh}) = p)$ đều có tính chất Q thì hầu hết mọi đồ thị trong $\mathcal{G}(n, \lfloor pN \rfloor)$ cũng đều có tính chất Q .

Tất cả những khẳng định này khá là đơn giản và chúng hoàn toàn có thể được chứng minh ở dạng sắc nét hơn nhiều nhưng, ngay khi ở trong phiên bản yếu này, chúng cũng cho thấy rằng, trên thực tế, $\mathcal{G}(n, p)$ và $\mathcal{G}(n, M)$ hoàn toàn có thể thay thế cho nhau trong nhiều tình huống khác nhau, miễn là $p = M/N$, $M \rightarrow \infty$, $N - M \rightarrow \infty$.

7.3. Các biến gần như được xác định – Kỹ thuật sử dụng phương sai

Nhận thấy rằng nếu $X = X(G)$ là một biến không âm xác định trên $\Omega = \mathcal{G}(n, M)$ hoặc $\Omega = \mathcal{G}(n, p)$ và kỳ vọng của X là μ , nghĩa là $E(X) = \mu$, thì, với mọi $c > 1$,

$$P(X \geq c\mu) \leq \frac{1}{c} \text{ và } P(X \leq c\mu) \geq \frac{c-1}{c},$$

bởi vì $\mu = E(X) \geq P(X \geq c\mu)c\mu$.

Như vậy, nếu kỳ vọng của X là rất nhỏ thì X là nhỏ đối với hầu hết các đồ thị. Trên thực tế, các đánh giá đơn giản này, các bất đẳng thức Markov, đã được sử dụng lặp đi lặp lại nhiều lần trong mục **7.1. Các mô hình cơ bản**. Tuy nhiên, nếu chúng ta muốn chỉ ra rằng X lớn hoặc khác không đối với hầu hết các đồ thị trong Ω thì giá trị kỳ vọng của nó rất hiếm khi giúp ích cho chúng ta, nên chúng ta sẽ phải thử một hướng tiếp cận ít tầm thường hơn một chút.

Trong các trường hợp này, chúng ta sẽ nhờ đến sự giúp đỡ của khái niệm phương sai: Nhắc lại rằng nếu $\mu = E(X)$ là giá trị kỳ vọng của X thì giá trị $Var(X) = \sigma^2(X) = E((X - \mu)^2) = E(X^2) - \mu^2$ được gọi là **phương sai của X** và $\sigma = \sigma(X) \geq 0$ được gọi là **độ lệch chuẩn của X** . Bất đẳng thức Chebyshev, một phiên bản ứng dụng của bất đẳng thức Markov cho $(X - \mu)^2$, phát biểu rằng: nếu $a > 0$ thì

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Đặc biệt,

$$P(X = 0) \leq P(|X - \mu| \geq \mu) \leq \frac{\sigma^2}{\mu^2}.$$

Trong nhiều ví dụ khác nhau, chúng ta thường quan tâm đến số $X = X(G)$ các đồ thị con của G được chứa trong một họ $\mathcal{F} = \{F_1, F_2, \dots\}$ nào đó. Khi đó, $\mathcal{F}$ sẽ phụ thuộc vào n và $V(F_i) \subset V(G) = [n]$. Chẳng hạn, $\mathcal{F}$ có thể là tập gồm $\binom{n}{s}$ đồ thị con đầy đủ bậc s , tập các chu trình Hamilton hoặc tập các cặp ghép đầy đủ. Như trong mục **7.1. Các mô hình cơ bản**, chúng ta có thể biểu diễn lại X dưới dạng $X = \sum_i Y_i$, trong đó $Y_i = Y_{F_i} = Y_{F_i}(G)$ là hàm đặc trưng của F_i : nó sẽ bằng 1 nếu $F_i \subset G$ hoặc 0 trong trường hợp ngược lại.

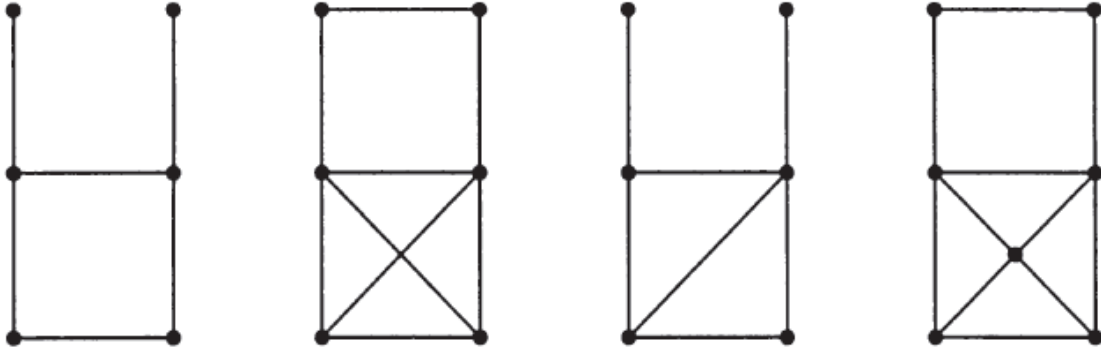
Khi đó,

$$E(X^2) = E\left(\left(\sum_i Y_i\right)^2\right) = \sum_i \sum_j E(Y_i Y_j) = \sum_{(F_i, F_j)} P(G \text{ có chứa } F_i \cup F_j),$$

trong đó tổng được tính trên tất cả các cặp (F_i, F_j) với $F_i, F_j \in \mathcal{F}$.

Chúng ta sẽ sử dụng những ý tưởng này để xác định xác suất $p(n)$ để $G_{n,p}$ có chứa một đồ thị con F . Theo Erdős và Rényi, chúng ta sẽ gọi một đồ thị là **cân bằng** nếu không có bất kỳ một đồ thị con nào của nó có bậc trung bình thật sự lớn hơn bậc trung bình của chính bản thân nó. Như vậy, nếu $F = G(k, l)$, nghĩa là,

F có k đỉnh và l cạnh, thì nó sẽ cân bằng nếu mọi đồ thị con với k' đỉnh sẽ chỉ có thể chứa tối đa không quá $k'l/k$ cạnh. Lưu ý rằng các đồ thị đầy đủ, các chu trình và các cấu trúc cây đều là các đồ thị cân bằng.



Hình vẽ minh họa: Hai đồ thị đầu tiên là các đồ thị cân bằng còn hai đồ thị còn lại thì không.

Kết quả sau đây của Erdős và Rényi chỉ ra rằng, khi $p(n)$ tăng, các đồ thị con cân bằng của $G_{n,p}$ xuất hiện một cách khá đột ngột:

Kết quả 7.3.1. (Erdős – Rényi)

Giả sử cho trước các số nguyên $k \geq 2, k - 1 \leq l \leq \binom{k}{2}$ và một đồ thị cân bằng $F = G(k, l)$ (với k đỉnh và l cạnh).

Khi đó, nếu $p(n)n^{k/l} \rightarrow 0$ thì hầu như không có bất kì đồ thị $G_{n,p}$ có chứa F và nếu $p(n)n^{k/l} \rightarrow \infty$ thì hầu hết mọi đồ thị $G_{n,p}$ đều có chứa F .

Chứng minh kết quả 7.3.1.

Kí hiệu $p = \gamma n^{-k/l}, 0 < \gamma < n^{k/l}$, và kí hiệu $X = X(G)$ là số các bản sao của F chứa trong $G_{n,p}$. Kí hiệu k_F là số các đồ thị xác định trên một tập gồm k đỉnh cố định cho trước đẳng cấu với F . Hiển nhiên, $k_F \leq k!$.

Khi đó,

$$\mu = E_p(X) = \binom{n}{k} k_F p^l (1 - p)^{\binom{k}{2} - l} \leq n^k (\gamma^l n^{-k}) = \gamma^l,$$

nên $E_p(X) \rightarrow \infty$ khi $\gamma \rightarrow 0$, điều này chứng minh tính đúng đắn của khẳng định thứ nhất.

Kế tiếp, chúng ta sẽ ước tính giá trị phương sai của X khi γ lớn. Lưu ý rằng chúng ta luôn có thể tìm được một hằng số $c_1 > 0$ sao cho

$$\mu \geq c_1 \gamma^l,$$

với mọi γ .

Từ hệ thức $E(X^2) = \sum_{(F_i, F_j)} P(G \text{ có chứa } F_i \cup F_j)$, chúng ta sẽ phải ước tính xác suất để đồ thị G có chứa hai bản sao cố định của F , tạm giả sử là F' và F'' .

Kí hiệu

$$A_s = \sum_s P_p(G_{n,p} \supset F' \cup F''),$$

trong đó tổng được tính trên tất cả các cặp (F', F'') với s đỉnh chung.

Hiển nhiên,

$$A_0 < \mu^2.$$

Hơn nữa, mọi tập đỉnh F' gồm s đỉnh đều chứa $t \leq (l/k)s$ cạnh. Như vậy, đầu tiên đếm số cách chọn cho F' và sau đó cho F'' với điều kiện F'' có $s \geq 1$ đỉnh chung với F' , chúng ta luôn có thể tìm được các hằng số c_2 và c_3 sao cho

$$\begin{aligned} \frac{A_s}{\mu} &\leq \sum_{t \leq ls/k} \binom{k}{s} \binom{n-k}{k-s} k! p^{l-t} q^{\binom{k}{2} - \binom{s}{2} - k+t} \\ &\leq \sum_{t \leq ls/k} c_2 n^{k-s} (\gamma n^{-k/l})^{l-t} \\ &\leq c_2 n^{-s} \gamma^l + c_3 \gamma^{l-1}, \end{aligned}$$

trong đó, tại bước cuối cùng, chúng ta đã tách số hạng tương ứng với $t = 0$ ra khỏi phần còn lại của biểu thức.

Do đó, bằng cách sử dụng $E(X^2) = \sum_{(F_i, F_j)} P(G \text{ có chứa } F_i \cup F_j)$, chúng ta có thể nhận thấy rằng

$$\frac{E(X^2)}{\mu^2} = \frac{\sum_{s=0}^k A_s}{\mu^2} \leq 1 + c_4 \gamma^{-1},$$

trong đó c_4 là một hằng số nào đó.

Như vậy, $P(X = 0) \leq \frac{\sigma^2}{\mu^2}$, $P(X = 0) \leq c_4 \gamma^{-1}$, nên $P(X = 0) \rightarrow 0$ khi $\gamma \rightarrow \infty$. □

Một trong những ví dụ nổi bật nhất về một bất biến đồ thị gần như được xác định trong một đồ thị ngẫu nhiên là chỉ số clique – bậc lớn nhất có thể có của một đồ thị con đầy đủ. Chúng ta sẽ chỉ ra rằng với một giá trị p cố định cho trước, $0 < p < 1$, chỉ số clique của hầu hết mọi đồ thị trong $\mathcal{G}(n, p)$ đều sẽ nhận một trong hai giá trị có thể có nào đó. Trên thực tế, vào năm 1976, Bollobás và Erdős đã chứng minh được rằng, với mọi số tự nhiên n , chỉ số clique của hầu hết mọi đồ thị thật ra chỉ là một hàm của p và n . Tuy nhiên, chúng ta sẽ hạn chế chính chúng ta ở việc chỉ chứng minh một kết quả rất đơn giản của Erdős và Rényi theo hướng này. Như trong Kết quả 7.1.1), bằng cách dùng $X_r = X_r(G_{n,p})$ để kí hiệu cho số các đồ thị con K_r của $G_{n,p}$,

$$E(X_r) = \binom{n}{r} p^{\binom{r}{2}}.$$

Kí hiệu $d = d(n, p)$ là số tự nhiên lớn nhất sao cho

$$E(X_d) = \binom{n}{d} p^{\binom{d}{2}} \geq \log n.$$

(do $E(X_1) = n$ và $E(X_n) = p^{\binom{n}{2}} < 1$ nên chúng ta luôn có thể tìm được một chỉ số nguyên $d, 1 \leq d \leq n - 1$ như vậy.)

Bằng cách sử dụng các công thức Stirling, chúng ta dễ dàng kiểm tra được rằng

$$\frac{n}{\log_b n} < \frac{n}{d} < p^{-d/2} < n,$$

trong đó $b = 1/p$ và $\log_b n = \log n / \log b$.

Hơn nữa,

$$d = 2\log_b n + O(\log \log n).$$

Kết quả 7.3.2. (Erdős – Rényi)

Giả sử cho trước một số thực dương cố định $0 < p < 1$.

Khi đó, chỉ số clique của hầu hết mọi đồ thị $G_{n,p}$ sẽ là d hoặc $d + 1$, trong đó $d = d(n)$.

Chứng minh kết quả 7.3.2.

Nhận thấy rằng khẳng định trên tương đương với việc:

$$P(X_{d+2} > 0) \rightarrow 0,$$

$$P(X_d > 0) \rightarrow 1.$$

Từ cách định nghĩa của d , $E_p(X_{d+1}) < \log n$, chúng ta suy ra rằng

$$E(X_{d+2}) = \frac{n - d - 1}{d + 2} p^{d+1} E(X_{d+1}) < p^{d/2} \log 2 < n^{-1/4} \rightarrow 0,$$

điều này ngụ ý tính đúng đắn của khẳng định đầu tiên.

Kế tiếp, chúng ta sẽ chứng minh rằng $P(X_d > 0) \rightarrow 1$. Đầu tiên, chúng ta nhận thấy rằng $\mu_d = E(X_d) \geq \log n \rightarrow \infty$, nên, từ $P(X = 0) \leq \frac{\sigma^2}{\mu^2}$, để chứng minh tính đúng đắn của khẳng định trên, chúng ta chỉ cần chỉ ra $\sigma_d / \mu_d \rightarrow 0$, trong đó $\sigma_d = \sigma_d(X_d)$, là đủ.

Chúng ta sẽ sử dụng hệ thức $E(X^2) = \sum_{(F_i, F_j)} P(G \text{ có chứa } F_i \cup F_j)$ để xác định moment bậc hai của X_d – tổng riêng tính trên tất cả các cặp đồ thị con của K_d với đúng l đỉnh chung:

$$E(X_d^2) = \sum_{l=0}^d \binom{n}{d} \binom{d}{l} \binom{n-d}{d-l} p^{2\binom{d}{2} - \binom{l}{2}} = \binom{n}{d} p^{2\binom{d}{2}} \sum_{l=0}^d \binom{d}{l} \binom{n-d}{d-l} p^{-\binom{l}{2}}.$$

Do

$$\mu_d^2 = E(X_d)^2 = \binom{n}{d}^2 p^{2\binom{d}{2}} = \binom{n}{d} p^{2\binom{d}{2}} \sum_{l=0}^d \binom{d}{l} \binom{n-d}{d-l},$$

với $\sigma_d = \sigma(X_d)$, nên chúng ta sẽ có

$$\begin{aligned} \frac{\sigma_d^2}{\mu_d^2} &= \frac{E(X_d^2) - \mu_d^2}{\mu_d^2} \leq \sum_{l=0}^d \binom{d}{l} \binom{n-d}{d-l} (p^{-\binom{l}{2}} - 1) \binom{n}{d}^{-1}, \\ &\leq \sum_{l=2}^d \binom{d}{l} \binom{n-d}{d-l} p^{-\binom{l}{2}} \binom{n}{d}^{-1}, \\ &\leq 2 \sum_{l=2}^d \frac{d!^2}{l! (d-l)!^2} n^{-l} p^{-\binom{l}{2}}, \\ &= 2 \sum_{l=2}^d \varepsilon_d, \end{aligned}$$

trong đó các số hạng $\varepsilon_2, \varepsilon_3, \dots, \varepsilon_l$ giảm dần còn các số hạng $\varepsilon_{l+1}, \varepsilon_{l+2}, \dots, \varepsilon_d$ tăng dần. Trên thực tế, chúng ta còn có thể kiểm tra được rằng

$$\varepsilon_l \leq \varepsilon_3 + \varepsilon_{d-1},$$

với mọi $3 \leq l \leq d-1$.

Như vậy,

$$\frac{\sigma_d^2}{\mu_d^2} \leq 2(\varepsilon_2 + \varepsilon_d) + 2d(\varepsilon_3 + \varepsilon_{d-1}).$$

Kế tiếp, chúng ta nhận thấy rằng:

$$\begin{aligned} 2\varepsilon_2 &\leq d^4 n^{-2} p^{-1} < n^{-1}, \\ 2d\varepsilon_3 &\leq 2d^7 n^{-3} p^{-3} < n^{-2}, \\ 2\varepsilon_d &= 2d! n^{-d} p^{-\binom{d}{2}} \leq 2/\mu_d, \\ 2d\varepsilon_{d-1} &\leq d^2 n p^{l-1} \varepsilon_d < n^{-1/2}. \end{aligned}$$

Bằng cách đặt các bất đẳng thức này vào đánh giá $\frac{\sigma_d^2}{\mu_d^2} \leq 2(\varepsilon_2 + \varepsilon_d) + 2d(\varepsilon_3 + \varepsilon_{d-1})$, chúng ta nhận thấy rằng

$$\frac{\sigma_d^2}{\mu_d^2} \leq 2n^{-1/2} + 2/\mu_d = o(1).$$

Điều này hoàn tất phép chứng minh của khẳng định thứ hai. □

Sau tất cả những ví dụ cụ thể nêu ở trên, chúng ta cần nói vài lời về bức tranh rộng lớn này. Như trước đây, một tính chất của các đồ thị được hiểu là một lớp các đồ thị đóng kín dưới một phép đẳng cấu nào đó. Đặc biệt, một tính chất Q_n của các đồ thị bậc n có thể được xem như là một tập con của tập các đồ thị với tập đỉnh $[n]$: tất cả những gì chúng ta yêu cầu chỉ là tập này là bất biến dưới các hoán vị của $[n]$. Một tính chất Q của các đồ thị được gọi là **đơn điệu tăng** nếu Q là bất biến dưới phép bổ sung các cạnh: nếu $G \in Q$, $G \subset H$ và $V(G) = V(H)$ thì $H \in Q$. (Tương tự, một tính chất được gọi là **đơn điệu giảm** nếu nó là bất biến dưới phép xóa các cạnh.) Như vậy, tính chất liên thông hoặc Hamiltonian là một tính chất đơn điệu tăng,

tính chất k – liên thông ($0 \leq k \leq 3$) là một tính chất đơn điệu giảm, nhưng tính chất có chứa một 6 – chu trình cảm sinh không tăng cũng không giảm.

Giả sử cho trước một tính chất Q , chúng ta sẽ dùng $P_p(Q) = P_p(G_{n,p} \text{ có tính chất } Q) = P_p(G_{n,p} \in Q)$ để kí hiệu xác suất để đồ thị $G_{n,p} \in \mathcal{G}(n, p)$ có tính chất Q ; kí hiệu $P_M(Q)$ cũng được định nghĩa tương tự trong $\mathcal{G}(n, M)$. Nghe có vẻ giống như một sự lặp lại, nhưng, trên thực tế, chúng ta sẽ cần một phép chứng minh thực sự cho khẳng định sau: với mọi tính chất đơn điệu tăng Q , xác suất $P_p(Q)$ là một hàm tăng của p và xác suất $P_M(Q)$ là một hàm tăng của M .

Kết quả 7.3.1) và Kết quả 7.3.2) minh họa cho một nguyên lí tổng quát được khám phá bởi Erdős và Rényi: nhiều tính chất đơn điệu tăng của các đồ thị phát sinh một cách khá đột ngột. Để tạo sự thuận tiện khi phát biểu khẳng định này một cách chính xác, chúng ta sẽ cần giới thiệu về các hàm ngưỡng. Một hàm $p_l(n)$ được gọi là một **hàm ngưỡng dưới của tính chất Q** nếu hầu như không có bất kì một đồ thị $G_{n,p_l(n)}$ nào có tính chất Q và $p_u(n)$ được gọi là một hàm ngưỡng trên của tính chất Q nếu hầu hết mọi đồ thị $G_{n,p_u(n)}$ đều có tính chất Q . Các hàm ngưỡng cũng được định nghĩa tương tự cho không gian $\mathcal{G}(n, M)$.

Theo cách định nghĩa của thuật ngữ hàm ngưỡng, Kết quả 7.3.1) chỉ ra rằng: “**Nếu $\omega(n) \rightarrow \infty$ và F là một đồ thị cân bằng với bậc trung bình $2l/k$ thì $n^{-k/l}/\omega$ là một hàm ngưỡng dưới và $\omega n^{-k/l}$ là một hàm ngưỡng trên của tính chất có chứa F như một đồ thị con.**” Mặc dù chúng ta sẽ không chứng minh những khẳng định này ở đây nhưng điều ngược lại của những khẳng định này cũng đúng: “ **$p_l(n)$ là một hàm ngưỡng dưới của tính chất có chứa F như một đồ thị con nếu và chỉ nếu $p_l(n)n^{k/l} \rightarrow 0$ và $p_u(n)$ là một hàm ngưỡng trên nếu và chỉ nếu $p_u(n)n^{k/l} \rightarrow \infty$.**” Trên thực tế, trong nhiều trường hợp các hàm ngưỡng dưới và các hàm ngưỡng trên sẽ gần nhau hơn rất nhiều so với trong ví dụ vừa nêu ở trên. Để minh họa nhận định này, chúng ta sẽ trình bày một kết quả cổ điển của Erdős và Rényi:

Kết quả 7.3.3. (Erdős – Rényi)

Giả sử cho trước một hàm $\omega(n) \rightarrow \infty$. Kí hiệu $p_l = (\log n - \omega(n))/n$ và $p_u = (\log n + \omega(n))/n$.

Khi đó, hầu hết mọi đồ thị G_{p_l} đều không liên thông và hầu hết mọi đồ thị G_{p_u} đều liên thông. Như vậy, trong mô hình $\mathcal{G}(n, p)$, p_l là một hàm ngưỡng dưới và p_u là một hàm ngưỡng trên của tính liên thông.

Chứng minh kết quả 7.3.3.

Không làm giảm tính tổng quát khi chứng minh tính đúng đắn của khẳng định trên, chúng ta có thể giả định rằng $\omega(n)$ là không quá lớn, tạm giả sử $\omega(n) \leq \log \log \log n$, và n đủ lớn để đảm bảo $\omega(n) \geq 10$. Với mọi $k \in \mathbb{N}$, kí hiệu $X_k = X_k(G)$ là số các thành phần liên thông của một đồ thị $G \in \mathcal{G}(n, p)$ với chính xác k đỉnh.

i) Kí hiệu $p = p_l = (\log n - \omega(n))/n$ và μ là kỳ vọng của số các đỉnh cô lập của đồ thị G_p .

Khi đó,

$$\mu = E(X_1) = n(1 - p)^{n-1} \sim ne^{-\log n + \omega(n)} = e^{\omega(n)} \rightarrow \infty.$$

Hơn nữa, kỳ vọng của số các cặp đỉnh cô lập sẽ được xác định bởi

$$E[X_1(X_1 - 1)] = n(n - 1)(1 - p)^{2n-3},$$

điều này suy ra từ việc chúng ta có $n(n - 1)$ cách để chọn một cặp sắp thứ tự các đỉnh và hai đỉnh cho trước bất kỳ là độc lập với nhau nếu và chỉ nếu không có bất kì một cặp đỉnh nào trong số $2n - 3$ cặp đỉnh liên thuộc với ít nhất một trong hai đỉnh trên là một cạnh của G_p .

Như vậy,

$$E(X_1^2) = n(n - 1)(1 - p)^{2n-3} + n(1 - p)^{n-1},$$

nên phương sai $\sigma^2 = \sigma^2(X_1)$ được xác định bởi

$$\begin{aligned} E[(X_1 - \mu)^2] &= E(X_1^2) - \mu^2 \\ &= n(n - 1)(1 - p)^{2n-3} + n(1 - p)^{n-1} - n^2(1 - p)^{2n-2} \\ &\leq n(1 - p)^{n-1} + pn^2(1 - p)^{2n-3} \\ &\leq \mu + (\omega(n) + \log n)ne^{-2\log n + 2\omega(n)}(1 - p)^{-3} \\ &\leq \mu + \frac{2\log n}{n}e^{2\omega(n)} \\ &\leq \mu + 1, \end{aligned}$$

trong đó, tại bất đẳng thức áp chót, chúng ta đã viện dẫn một thực tế là p nhỏ nên $(1 - p)^3 \geq 1/2$.

Như vậy, chúng ta dễ dàng nhận thấy rằng

$$P(G_p \text{ liên thông}) \leq P(X_1 = 0) \leq \frac{1}{\mu^2}E[(X_1 - \mu)^2] \leq \frac{\mu + 1}{\mu^2} = \mu^{-1} + \mu^{-2}.$$

Điều này cho thấy rằng hầu hết mọi đồ thị G_p đều không liên thông.

ii) Kí hiệu $p = p_u = (\log n + \omega(n))/n$.

Khi đó,

$$\begin{aligned} P(G_p \text{ không liên thông}) &= P\left[\sum_{k=1}^{\lfloor n/2 \rfloor} X_k \geq 1\right] \\ &\leq E\left[\sum_{k=1}^{\lfloor n/2 \rfloor} X_k\right] \\ &= \sum_{k=1}^{\lfloor n/2 \rfloor} E[X_k] \\ &\leq \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1 - p)^{k(n-k)}, \end{aligned}$$

điều này suy ra từ việc chúng ta có $\binom{n}{k}$ cách để lựa chọn tập đỉnh cho một thành phần liên thông với k đỉnh và chúng ta phải đảm bảo rằng không có bất kì một cạnh nào nối tập này với phần còn lại của đồ thị (ngoài việc, chúng ta sẽ phải có một số cạnh để đảm bảo thành phần đang xét là liên thông nhưng chúng ta lại không cần sử dụng điều kiện này ở đây).

Chúng ta sẽ chia tổng trên thành hai phần như sau:

Đầu tiên, chúng ta sẽ xét trường hợp khi k nhỏ:

$$\begin{aligned} \sum_{1 \leq k \leq n^{3/4}} \binom{n}{k} (1-p)^{k(n-k)} &\leq \sum_{1 \leq k \leq n^{3/4}} \left(\frac{en}{k}\right)^k e^{-knp} e^{k^2 p} \\ &\leq \sum_{1 \leq k \leq n^{3/4}} e^{(1-\omega(n))k} k^{-k} e^{2k^2(\log n)/n} \\ &\leq 3e^{-\omega(n)}, \end{aligned}$$

nếu n đủ lớn.

Với k lớn, chúng ta sẽ lập luận hơi khác hơn một chút:

$$\begin{aligned} \sum_{n^{3/4} < k \leq n/2} \binom{n}{k} (1-p)^{k(n-k)} &\leq \sum_{n^{3/4} < k \leq n/2} \left(\frac{en}{k}\right)^k e^{-knp/2} \\ &\leq \sum_{n^{3/4} < k \leq n/2} (en^{1/4})^k n^{-k/2} \\ &\leq \sum_{n^{3/4} < k \leq n/2} (e/n^{1/4})^k \\ &\leq n^{-n^{3/4}/5}. \end{aligned}$$

Bằng cách kết hợp các đánh giá nêu ở trên, chúng ta nhận thấy rằng

$$P(G_p \text{ không liên thông}) \leq 4e^{-\omega(n)},$$

nếu n đủ lớn.

Điều này cho thấy rằng hầu hết mọi đồ thị G_p đều liên thông. □

Một vấn đề được đặt ra ở đây đó là: “Liệu chúng ta có thể nói được điều gì về sắc số của một đồ thị ngẫu nhiên $G_{n,p}$ với một giá trị xác suất p cố định cho trước?”. Ngay lập tức, Kết quả 7.3.2) chỉ ra cho chúng ta một chặn dưới của $\chi(G)$, bởi vì $\chi(G) \geq \max\{\omega(G), |G|/\alpha(G)\}$ với mọi đồ thị G . Hơn nữa, do phần bù của một đồ thị ngẫu nhiên $G_{n,p}$ là một đồ thị ngẫu nhiên $G_{n,q}$, với $q = 1 - p$, nên phân phối của chỉ số độc lập $\alpha(G_{n,p})$ cũng chính là phân phối của chỉ số clique $\omega(G_{n,p})$.

Như vậy, từ Kết quả 7.3.2), $\omega(G_{n,p}) = \left(\frac{1}{2} + o(1)\right) \log n / \log(1/q)$, chúng ta nhận thấy rằng

$$\chi(G_{n,p}) \geq \left(\frac{1}{2} + o(1)\right) \frac{\log n}{\log(1/q)},$$

với hầu hết mọi đồ thị $G_{n,p}$.

Chặn dưới tầm thường mà chúng ta vừa tìm ra ở trên cách chặn trên của $\chi(G_{n,p})$ bao xa? Một cách tự nhiên để chúng ta có thể thu được một chặn trên của $\chi(G_{n,p})$ là phân tích kết quả của một thuật toán tô màu nào đó khi cho nó chạy trên $G_{n,p}$. Kế tiếp, thuật toán tô màu đơn giản nhất mà chúng ta biết đó là thuật toán tô màu tham lam đã được thảo luận trong mục **5.1. Phép tô màu đỉnh**. Năm 1976, Bollobás và Erdős đã chỉ ra rằng, với xác suất cao, thuật toán này thực sự tạo ra một phép tô màu chỉ sử dụng số màu vào khoảng gấp đôi so với chặn dưới nêu ở trên. Năm 1988, Bollobás đã sử dụng một cách tiếp cận khác, phi thuật toán, để chứng minh rằng, trên thực tế, $\left(\frac{1}{2} + o(1)\right) \frac{\log n}{\log(1/q)}$ gần như là tốt nhất có thể:

Kết quả 7.3.3. (Bollobás)

Giả sử cho trước một số thực dương $0 < p < 1$.

Khi đó,

$$\chi(G_{n,p}) = \left(\frac{1}{2} + o(1)\right) \frac{\log n}{\log(1/q)},$$

với hầu hết mọi đồ thị $G_{n,p}$, trong đó $q = 1 - p$.

Điều mà Kết quả 7.3.3) chỉ ra đó là nếu $\varepsilon > 0$ thì

$$\lim_{n \rightarrow \infty} P_p \left(\left| \chi(G_{n,p}) \log(1/q) / \log n - \frac{1}{2} \right| < \varepsilon \right) = 1.$$

Các hàm ngưỡng tối ưu cung cấp cho chúng ta một lượng thông tin đáng kể về một tính chất, nhưng để có được một cái nhìn sâu sắc hơn về sự xuất hiện của một tính chất chúng ta sẽ cần phải xem xét thêm khái niệm thời điểm xảy ra (hoặc thời điểm xảy ra đầu tiên).

Giả sử cho trước một tính chất đơn điệu tăng Q , thời điểm τ tại đó Q xuất hiện trong một quá trình đồ thị $\tilde{G} = (G_t)_{t=0}^N$ được gọi là **thời điểm xảy ra** (hoặc **thời điểm xảy ra đầu tiên**) của Q :

$$\tau = \tau_Q = \tau(\tilde{G}; Q) = \min\{t: G_t \text{ có tính chất } Q\}.$$

Các hàm ngưỡng trong mô hình $\mathcal{G}(n, M)$ có thể dễ dàng được mô tả lại theo thuật ngữ thời điểm xảy ra.

Thật vậy, m_l là một hàm ngưỡng dưới của một tính chất Q nếu và chỉ nếu $\tau(\tilde{G}; Q) > m_l$ với hầu hết mọi quá trình đồ thị $\tilde{G}$; một hàm ngưỡng trên cũng được đặc trưng tương tự.

Có một số kết quả khá nổi bật chỉ ra rằng thời điểm xảy ra của hai tính chất dường như cách xa nhau gần như giống với các kết quả trong không gian các đồ thị ngẫu nhiên của chúng ta. Một ví dụ hay về nhận định này thuộc về tính chất liên thông, đã được giới thiệu trong Kết quả 7.3.3).

Đâu là trở ngại căn bản nhất của tính liên thông? Sự tồn tại của một đỉnh cô lập. Nói cách khác, nếu Q_1 là tính chất liên thông ($Q_1 = \text{“liên thông”}$) và Q_2 là tính chất có bậc nhỏ nhất ≥ 1 ($Q_2 = \text{“}\delta \geq 1\text{”}$) thì $\tau(\tilde{G}; Q_1) \geq \tau(\tilde{G}; Q_2)$ với mọi quá trình đồ thị $\tilde{G}$. Một điều đáng ngạc nhiên hơn ở đây đó là dấu đẳng thức xảy ra với hầu hết mọi quá trình đồ thị.

Kết quả 7.3.4.

Với hầu hết mọi quá trình đồ thị $\tilde{G}$, chúng ta luôn có $\tau(\tilde{G}; \text{liên thông}) = \tau(\tilde{G}; \delta \geq 1)$.

Mặc dù phép chứng minh của kết quả này chỉ phức tạp hơn một chút so với Kết quả 7.3.3), nhưng chúng ta sẽ không trình bày nó ở đây. Tuy nhiên, chúng ta sẽ vẫn mở rộng thêm khẳng định này. Điều mà Kết quả 7.3.3) chỉ ra rằng nếu chúng ta bắt đầu với một đồ thị rỗng xác định trên một tập lớn các đỉnh và, bằng cách bổ sung thêm liên tục vào nó các cạnh một cách ngẫu nhiên cho đến khi đồ thị không có chứa bất kì một đỉnh cô lập nào thì, với xác suất cao, đồ thị mà chúng ta nhận được là liên thông: chính các cạnh đó đã loại bỏ đi đỉnh cô lập cuối cùng và làm cho đồ thị nhận được liên thông. Thoạt nhìn, đây thực sự là kết quả bất ngờ nhất mà chúng ta nhận được.

Lưu ý rằng chúng ta cũng có thể dễ dàng suy ra Kết quả 7.3.3) từ Kết quả 7.3.4), bởi vì Kết quả 7.3.4) ngụ ý rằng tính chất liên thông có cùng hàm ngưỡng với tính chất có bậc nhỏ nhất ≥ 1 . Chúng ta cũng dễ dàng chứng minh được rằng $P_p(\delta(G_{n,p}) \geq 1) \rightarrow 0$ nếu và chỉ nếu $E_p(X_{n,p}) \rightarrow \infty$, trong đó $X_{n,p}$ là số các đỉnh cô lập của $G_{n,p}$. Hơn nữa, $P_p(\delta(G_{n,p}) \geq 1) \rightarrow 1$ nếu và chỉ nếu $E_p(X_{n,p}) \rightarrow 0$. Các khẳng định tương tự cũng đúng cho các tính chất k – liên thông và có bậc nhỏ nhất $\geq k$.

7.4. Các chu trình Hamilton

Trong các phép chứng minh mà chúng ta trình bày cho đến nay, chúng ta luôn áp dụng một ý tưởng tiếp cận trực tiếp vấn đề dù ít hay nhiều. Chúng ta hầu như không cần gì nhiều hơn từ lý thuyết đồ thị ngoài các định nghĩa của các khái niệm liên quan và trọng tâm của các phép chứng minh nằm ở việc vận dụng các kiến thức cơ bản của lý thuyết xác suất. Mục này của bài viết sẽ được dành cho việc trình bày một định lý đẹp của Pósa liên quan đến các chu trình Hamilton, phép chứng minh của chúng ta cơ bản là dựa trên một kết quả tao nhã trong lý thuyết đồ thị. Tất nhiên, việc sử dụng ý tưởng của các phương pháp xác suất trong lý thuyết đồ thị của chúng ta ở đây sẽ có sự kết hợp của tất cả các ý tưởng đã được trình bày trong bốn mục trước. Như vậy, chúng ta sẽ phải chuẩn bị một nền tảng đủ tốt để có thể sử dụng các kết quả không tầm thường trong lý thuyết đồ thị và sẽ áp dụng lý thuyết xác suất để thu lấy một số thông tin quan trọng về các đồ thị trong một không gian xác suất được thiết kế riêng cho vấn đề này. Chúng ta cũng sẽ chọn một đồ thị thích hợp mà sau đó chúng ta sẽ thay đổi nó đi đôi chút với sự trợ giúp của các công cụ lý thuyết đồ thị mạnh mẽ.

Như chúng ta đã thấy trong **Chủ đề 04. Các bài toán cực trị đồ thị**, các nghiên cứu về các chu trình Hamilton là một trong những mảng quan trọng nhất của lý thuyết đồ thị trong nhiều năm và, ngày nay, chúng ta cũng đã biết được rất nhiều điều kiện đủ khác nhau để một đồ thị có dạng Hamiltonian. Ở đây, chúng ta quan tâm đến một khía cạnh khác của bài toán về các chu trình Hamiltonian: “Điều gì xảy ra trong trường hợp trung bình? Giả sử cho trước một số tự nhiên n , với những giá trị nào của m thì một đồ thị điển hình với bậc n và kích thước m sẽ có chứa một chu trình Hamilton?”

Mọi đồ thị Hamiltonian với bậc n đều chứa $\geq n$ cạnh và mọi đồ thị phi Hamiltonian với bậc n đều chứa $\leq \binom{n-1}{2} + 1$ cạnh. Điều này mở ra cho chúng ta một “sự thật không chắc chắn” khá lớn: “Với $1 \leq n \leq m \leq \binom{n-1}{2} + 1$, một số đồ thị $G(n, m)$ sẽ là Hamiltonian và một số đồ thị khác sẽ là đồ thị phi Hamiltonian.”

Bằng cách thay thế hạn chế từ kích thước thành bậc nhỏ nhất, chúng ta thu được một số kết quả: “Mọi đồ thị Hamiltonian đều có bậc nhỏ nhất ≥ 2 ” và, từ định lý Dirac, “Mọi đồ thị phi Hamiltonian đều có bậc nhỏ nhất $\leq \lfloor (n-1)/2 \rfloor$ ”.

Điều hấp dẫn ở đây là nếu chúng ta không đòi hỏi **sự chắc chắn** mà chỉ cần **xác suất cao** thì vấn đề trên sẽ trở nên rất nhỏ.

Thay vì cố định kích thước, chúng ta sẽ cố định xác suất: như chúng ta biết, có rất ít sự khác biệt giữa hai cách tiếp cận và chúng ta cũng sẽ dễ dàng hơn khi làm việc với $\mathcal{G}(n, p)$ thay vì $\mathcal{G}(n, M)$. Như vậy, vấn đề chúng ta muốn giải quyết có thể được phát biểu lại như sau: “Với những giá trị nào của $p = p(n)$ thì $G_{n,p}$ có thể là một đồ thị Hamiltonian?” Từ các kết quả trong các mục trước của bài viết, chúng ta biết được một chặn dưới, mặc dù là một chặn dưới khá yếu: “Nếu $p = (\log n - \omega(n))/n$, trong đó $\omega(n) \rightarrow \infty$, thì hầu hết mọi đồ thị $G_{n,p}$ đều không liên thông, nên hầu như không có bất kì một đồ thị $G_{n,p}$ nào là Hamiltonian.” Năm 1976, Pósa đã đạt được một bước đột phá quan trọng khi ông đã chứng minh được rằng, với cùng một bậc của xác suất, hầu hết mọi đồ thị $G_{n,p}$ đều là Hamiltonian.

Cơ sở của phép chứng minh của kết quả này là Kết quả 4.3.5). Kí hiệu S là một x_0 – đường đi dài nhất trong một đồ thị H và kí hiệu L là tập các đỉnh cuối của các phép biến đổi của S . Kí hiệu N là tập các đỉnh láng giềng của các đỉnh của L trên S và kí hiệu $R = V(H) - L \cup N$. Khi đó, Kết quả 4.3.5) chỉ ra rằng H không có chứa bất kì một $L - R$ cạnh nào. Tất cả những gì chúng ta cần từ khẳng định này là nếu $|L| = l \leq |H|/3$ thì chúng ta luôn có thể tìm được các tập phân biệt với kích thước l và $|H| - 3l + 1$ được nối với nhau không bởi bất kì một cạnh nào của H .

Chúng ta sẽ bắt đầu công việc với một bổ đề đơn giản tương tự như trong Kết quả 7.1.4). Kí hiệu D_t là số các cặp (X, Y) của các tập con phân biệt của V sao cho $|X| = t$, $|Y| = n - 3t$ và G không có chứa bất kì một $X - Y$ cạnh nào.

Kết quả 7.4.1.

Giả sử cho trước các hằng số $c > 3$ và $0 < \gamma < \frac{1}{3}$. Kí hiệu $p = (c \log n)/n$.

Khi đó, trong $\mathcal{G}(n, p)$, chúng ta sẽ có

$$P_p(D_t > 0 \text{ với mọi } t, 1 \leq t \leq \gamma n) = O(n^{3-c}).$$

Chứng minh kết quả 7.4.1.

Kí hiệu $\beta = \frac{c-3}{4c}$.

Khi đó,

$$\begin{aligned} \sum_{t=1}^{\lfloor \gamma n \rfloor} E_p(D_t) &= \sum_{t=1}^{\lfloor \gamma n \rfloor} \binom{n}{t} \binom{n-t}{n-3t} (1-p)^{t(n-3t)} \\ &\leq n \binom{n-1}{2} (1-p)^{n-3} + \sum_{t=2}^{\lfloor \beta n \rfloor} \frac{1}{t!} n^{3t} (1-p)^{t(n-3t)} + \sum_{t=\lfloor \beta n \rfloor+1}^{\lfloor \gamma n \rfloor} 2^{2n} (1-p)^{t(n-3t)}. \end{aligned}$$

Kế tiếp, do $(1 - p)^n < n^{-c}$ nên chúng ta sẽ có

$$n^3(1 - p)^{n-3} < (1 - p)^{-3}n^{3-c};$$

nếu $2 \leq t \leq \beta n$ thì

$$n^{3t}(1 - p)^{t(n-3t)} < n^{t(3-(c(n-3t)/n))} \leq n^{3-c};$$

nếu $\beta n \leq t \leq \gamma n$ thì

$$(1 - p)^{t(n-3t)} < n^{2n/\log n - (n-3t)t/n} = O(n^{-\beta(1-3\gamma)n}).$$

Như vậy,

$$\sum_{t=1}^{\lfloor \gamma n \rfloor} E_p(D_t) = O(n^{3-c}).$$

Điều này ngầm khẳng định tính đúng đắn của khẳng định của chúng ta. □

Kết quả 7.4.2.

Kí hiệu $p = (c \log n)/n$ và xét không gian $\mathcal{G}(n, p)$.

Khi đó, nếu $c > 3$ và x và y là các đỉnh bất kì thì hầu hết mọi đồ thị đều có chứa một đường đi Hamilton từ x đến y ; nếu $c > 9$ thì hầu hết mọi đồ thị đều là Hamiltonian liên thông: mọi cặp đỉnh phân biệt đều được nối với nhau bởi một đường đi Hamilton.

Chứng minh kết quả 7.4.2.

Chọn $\gamma < \frac{1}{3}$ sao cho $c\gamma > 3$ nếu $c > 9$ và $c\gamma > 1$ nếu $c > 3$.

Chúng ta sẽ giới thiệu thêm một số kí hiệu sau đây cho các sự kiện nhất định trong $\mathcal{G}(n, p)$:

$D = \{D_t = 0 \text{ với mọi } t, 1 \leq t \leq \lfloor \gamma n \rfloor\}$, trong đó D_t là số các cặp (X, Y) của các tập con phân biệt của V sao cho $|X| = t$, $|Y| = n - 3t$ và G không có chứa bất kì một $X - Y$ cạnh nào;

$E(W, x) = \{G_{n,p}[W] \text{ có chứa một đường đi với chiều dài lớn nhất mà đỉnh cuối của nó được nối với } x\}$;

$E(W, x|w) = \{G_{n,p}[W] \text{ có chứa một } w - \text{đường đi với chiều dài lớn nhất trong số các } w - \text{đường đi mà đỉnh cuối của nó được nối } x\}$;

$F(x) = \{\text{mọi đường đi với chiều dài lớn nhất có chứa } x\}$;

$H(W) = \{G_{n,p}[W] \text{ có chứa một đường đi Hamilton}\}$;

$H(x, y) = \{G_{n,p} \text{ có chứa một } x - y \text{ đường đi Hamilton}\}$;

$HC = \{G_{n,p} \text{ là Hamiltonian liên thông}\}$.

Chúng ta sẽ đồng nhất một biến cố với tập con tương ứng của $\mathcal{G}(n, p)$, nên phần bù của một biến cố A được kí hiệu bởi $\bar{A} = \mathcal{G}(n, p) \setminus A$.

Lưu ý rằng, từ Kết quả 7.4.1), chúng ta có

$$P_p(\bar{D}) = 1 - P_p(D) = O(n^{3-c}).$$

Chúng ta sẽ chọn ra một đỉnh x và một tập $W \subset V \setminus \{x\}$ cố định với $|W| = n - 2$ hoặc $n - 1$. Mục đích đầu tiên của chúng ta là chứng minh rằng $P_p(D \cap \bar{E}(W, x))$ là khá nhỏ. Kí hiệu $G \in D \cap \bar{E}(W, x)$ và xét một đường đi $S = x_0 x_1 \dots x_k$ với chiều dài lớn nhất trong $G[W]$. (Bằng cách xác định một cấu trúc thứ tự trên W , chúng ta có thể dễ dàng nhận thấy rằng S được xác định bởi $G[W]$.) Kí hiệu $L = L(G[W])$ là tập các đỉnh cuối của các phép biến đổi của x_0 – đường đi S và kí hiệu R tương tự như trong Kết quả 4.3.5 (áp dụng cho $G[W]$). Nhắc lại rằng $|R| \geq |W| + 1 - 3|L|$ và không có bất kì một $L - R$ cạnh nào, nên cũng sẽ không có bất kì một $L - (R \cup \{x\})$ nào. Do $G \in D$ và $|R \cup \{x\}| \geq n - 3|L|$ nên chúng ta nhận thấy rằng $|L| \geq \gamma n$.

Kế tiếp, do L chỉ phụ thuộc vào các cạnh nằm trong $G[W]$, nên nó sẽ độc lập với các cạnh liên thuộc với x . Do đó,

$$P_p(D \cap \bar{E}(W, x)) \leq P_p(|L(G[W])| \geq \gamma n) P_p(N(x) \cap L' = \emptyset),$$

trong đó L' là một tập cố định gồm $\lceil \gamma n \rceil$ đỉnh của W .

Như vậy,

$$P_p(D \cap \bar{E}(W, x)) \leq (1 - p)^{\gamma n} < n^{-c\gamma},$$

nên xác suất trong câu hỏi của chúng ta thực sự là nhỏ.

Bằng cách sử dụng cùng một phép chứng minh tương tự, chúng ta cũng nhận thấy rằng

$$P_p(D \cap \bar{E}(W, x|w)) < n^{-c\gamma},$$

với điều kiện $|W| = n - 2$ hoặc $n - 1$, $w \in W$ và $x \notin W$.

Lưu ý rằng $\bar{F}(x) \subset \bar{E}(V - \{x\}, x)$, nên

$$\begin{aligned} P_p(\bar{H}(V)) &= P_p\left(\bigcup_{x \in V} \bar{F}(x)\right) \leq P_p\left(D \cap \bigcup_{x \in V} \bar{F}(x)\right) + P_p(\bar{D}) \\ &\leq \sum_{x \in V} P_p(D \cap \bar{F}(x)) + P_p(\bar{D}) \\ &\leq n P_p(D \cap \bar{E}(V - \{x\}, x)) + P_p(\bar{D}) \\ &\leq n^{1-c\gamma} + O(n^{3-c}). \end{aligned}$$

Điều này chứng tỏ rằng nếu $c > 3$ thì hầu hết mọi đồ thị đều có chứa một đường đi Hamilton.

Kế tiếp, kí hiệu x và y là các đỉnh phân biệt và kí hiệu $W = V - \{x, y\}$.

Từ phần thứ nhất của phép chứng minh,

$$P_p(\bar{H}(W)) \leq 2n^{1-c\gamma} + O(n^{3-c}).$$

Do

$$H(x, y) \supset H(W) \cap E(W, y) \cap E(W, x|y),$$

nên chúng ta có

$$\begin{aligned} P_p(\bar{H}(x, y)) &\leq P_p(\bar{H}(W)) + P_p(D \cap \bar{E}(W, y)) + P_p(D \cap \bar{E}(W, x|y)) + P_p(\bar{D}) \\ &\leq 2n^{1-c\gamma} + 2n^{-c\gamma} + O(n^{3-c}). \end{aligned}$$

Do đó, nếu $c > 3$ thì hầu hết mọi đồ thị đều có chứa một đường đi Hamilton từ x đến y .

Cuối cùng, do chúng ta có $\binom{n}{2}$ cách chọn cho một cặp không sắp thứ tự (x, y) , $x \neq y$, nên

$$P_p(\overline{HC}) \leq \sum_{x \neq y} P_p(\bar{H}(x, y)) \leq n^{3-c\gamma} + n^{2-c\gamma} + O(n^{5-c}).$$

Như vậy, nếu $c > 9$ thì hầu hết mọi đồ thị đều là Hamiltonian liên thông. □

Do mọi đồ thị Hamiltonian liên thông đều là Hamiltonian nên, từ Kết quả 7.4.3), chúng ta cũng biết rằng nếu $c > 9$ và $p = c \log n / n$ thì hầu hết mọi đồ thị đều là Hamiltonian. Độc lập với Pósa, Korshunov đã chứng minh được kết quả tốt nhất có thể có cho khẳng định này với mọi $c > 1$. Quan trọng hơn, vào năm 1983, Komlós và Szemerédi đã xác định được các hàm ngưỡng tối ưu của tính chất Hamiltonian.

Kết quả 7.4.3. (Komlós – Szemerédi)

Giả sử cho trước một hàm $\omega(n) \rightarrow \infty$.

Kí hiệu $p_l = (\log n + \log \log n - \omega(n))/n$ và $p_u = (\log n + \log \log n + \omega(n))/n$.

Khi đó, p_l là một hàm ngưỡng dưới và p_u là một hàm ngưỡng trên của tính chất Hamiltonian.

Trên thực tế, tương tự như Kết quả 7.3.4), chúng ta cũng có một kết quả về thời điểm xảy ra liên hệ giữa tính chất Hamiltonian với tính chất trở ngại của nó. Chúng ta nhận thấy rằng tính không liên thông là một tính chất trở ngại hiển nhiên của tính chất Hamiltonian. Trên thực tế, chúng ta còn có một tính chất trở ngại hiển nhiên hơn và rất dễ để phát hiện ra nó: tính chất có bậc nhỏ nhất ≤ 1 . Vào năm 1983, Bollobás đã chỉ ra rằng đây thật sự là tính chất trở ngại chính của tính Hamiltonian, với mọi quá trình đồ thị:

Kết quả 7.4.4. (Bollobás)

Với hầu hết mọi quá trình đồ thị $\tilde{G}$, chúng ta luôn có $\tau(\tilde{G}; \text{Hamiltonian}) = \tau(\tilde{G}; \delta \geq 2)$, trong đó “Hamiltonian” là tính chất Hamiltonian và “ $\delta \geq 2$ ” là tính chất có bậc nhỏ nhất ≥ 2 .

Như vậy, nếu chúng ta dùng một quá trình đồ thị ngẫu nhiên ngay khi chúng ta loại bỏ đi đỉnh cuối cùng với bậc ≤ 1 thì, với xác suất cao, chúng ta sẽ có một đồ thị Hamiltonian. Kết quả 7.4.4) có thể dễ dàng suy ra Kết quả 7.4.3); trên thực tế, chỉ với việc bổ sung thêm một chút công việc, nó còn ngụ ý cả dạng sâu sắc hơn sau đây của Kết quả 7.4.3):

Kết quả 7.4.3’.

Giả sử cho trước một số thực $c \in \mathbb{R}$. Kí hiệu $p = (\log n + \log \log n + c)/n$.

Khi đó, $\lim_{n \rightarrow \infty} P_p(G_{n,p} \text{ là Hamiltonian}) = e^{-e^{-c}}$.

Không nhất thiết phải nói ra, chúng ta cũng có thể biết rằng câu chuyện sẽ không dừng lại ở đây: chúng ta còn rất nhiều câu hỏi liên quan đến các đồ thị ngẫu nhiên và các chu trình Hamiltonian. Chẳng hạn, khi chúng ta đã khám phá ra chương ngại “chính” của vấn đề, cụ thể là sự tồn tại của các đỉnh với bậc ≤ 1 , chúng ta sẽ nhận ra rằng việc loại trừ chúng và xác định xác suất để một đồ thị ngẫu nhiên với bậc n và kích thước m , với điều kiện là nó có bậc nhỏ nhất ≥ 2 , là Hamiltonian là một điều tự nhiên. Khi điều đó xảy ra, chương ngại thứ hai mà chúng ta gặp phải đó là một “**mạng nhện**”: ba đỉnh với bậc 2 có chung với nhau một đỉnh láng giềng nào đó. Tuy nhiên, các kết quả của loại vấn đề này sẽ phù hợp với các bài viết chuyên luận mang tính chất chuyên môn cao hơn là cho một tài liệu đơn sơ về lý thuyết đồ thị ngẫu nhiên như chúng ta đang thực hiện.

7.5. Hiện tượng chuyển pha

Một đồ thị ngẫu nhiên “điển hình” G_M sẽ trông như thế nào? Cụ thể hơn, một quá trình đồ thị “điển hình” $(G_t)_{t=0}^N$ sẽ trông như thế nào? Đặc biệt, cấu trúc của một thành phần liên thông của G_t sẽ thay đổi như thế nào khi t tăng?

Khá dễ dàng để chúng ta có thể thấy rằng nếu t khá nhỏ thì G_t có xu hướng chỉ có các thành phần liên thông dạng cây, với bậc của các thành phần phụ thuộc vào độ lớn của t .

Kết quả 7.5.1.

Với hầu hết mọi quá trình đồ thị ngẫu nhiên, nếu $k \geq 2$ là một số tự nhiên cố định cho trước và $t = o(n^{(k-1)/k})$ thì mọi thành phần liên thông của G_t sẽ là một cây với bậc $\leq t$. Hơn nữa, nếu s là một hằng số và $t/n^{(k-2)/(k-1)} \rightarrow \infty$ thì G_t có chứa $\geq s$ thành phần liên thông với bậc k .

Phép chứng minh của khẳng định này đi theo đường lối chứng minh của Kết quả 7.3.1) và hơi nhạt nhẽo: chúng ta thậm chí còn không cần phải sử dụng đến việc có k^{k-2} cây với bậc k để chứng minh nó. Tất cả những gì chúng ta phải làm chỉ là ước tính các giá trị $E(X_k)$ và $E(X_k^2)$, trong đó X_k là số các cây với bậc k chứa trong G_t , chỉ bằng cách sử dụng việc có một số $t(k)$, $1 \leq t(k) \leq \binom{k}{k-1}$, cây xác định trên k đỉnh phân biệt.

Sự tăng trưởng của bậc lớn nhất của một thành phần liên thông được mô tả trong Kết quả 7.5.1) là khá ổn định và đều đặn: không có bất kỳ một sự thay đổi bất ngờ nào xảy ra. Điều mà Erdős và Rényi khám phá ra đó là khi giá trị t vào khoảng $t = \frac{n}{2}$ sự tăng trưởng này trở nên khá điên cuồng: nhìn một cách toàn cảnh vào các đồ thị, chúng ta sẽ thấy được sự thay đổi đột ngột về chất trong cấu trúc của các thành phần liên thông. Sự thay đổi về chất này được gọi là **hiện tượng chuyển pha** của một quá trình đồ thị ngẫu nhiên. Nói một cách đại khái, trước thời điểm $\frac{n}{2}$, mọi thành phần liên thông đều có $O(\log n)$ đỉnh; nhưng, sau thời điểm $\frac{n}{2}$, chúng ta chỉ có duy nhất một thành phần liên thông lớn nhất với bậc n (nghĩa là chứa một tỉ lệ không

đổi của các đỉnh). Hơn nữa, tất cả các thành phần liên thông khác vẫn có bậc $\log n$; trên thực tế, khi t tăng lên, giá trị của chúng sẽ ngày càng nhỏ hơn.

Để xây dựng dạng công thức cho kết quả này một cách chính xác, với mọi đồ thị G , chúng ta sẽ dùng $L^{(1)}(G) \geq L^{(2)}(G) \geq \dots$ để kí hiệu cho bậc của các thành phần liên thông của G , nghĩa là $\sum_i L^{(i)}(G) = |G|$. Từ Kết quả 7.5.1), chúng ta nhận thấy rằng, với hầu hết mọi đồ thị $G_t = G_{n,t}$, $L^{(1)}(G_t) = k$, nếu $t/n^{(k-1)/k} \rightarrow 0$ và $t/n^{(k-2)/(k-1)} \rightarrow \infty$. Với $t = \left\lfloor \frac{cn}{2} \right\rfloor$, và với c là một hằng số cho trước bất kì, kết quả sau đây là một phiên bản sâu sắc hơn của kết quả nổi tiếng của Erdős và Rényi về hiện tượng chuyển pha:

Kết quả 7.5.2. (Erdős – Rényi)

Giả sử cho trước một số thực dương $c > 0$, một số tự nhiên $h \geq 1$ cố định và một hàm $\omega(n) \rightarrow \infty$. Kí hiệu $\alpha = c - 1 - \log c$ và $t = t(n) = \left\lfloor \frac{cn}{2} \right\rfloor$. Khi đó, chúng ta có các kết quả sau:

i) Nếu $c < 1$ thì, với hầu hết mọi đồ thị ngẫu nhiên G_t ,

$$\left| L^{(i)}(G_t) - \frac{1}{\alpha} \left\{ \log n - \frac{5}{2} \log \log n \right\} \right| \leq \omega(n),$$

với mọi i , $1 \leq i \leq h$.

ii) Chúng ta luôn có thể tìm được các hằng số dương $c_2 > c_1 > 0$ sao cho, với mọi i , $1 \leq i \leq h$,

$$c_1 n^{2/3} < L^{(i)} \left(G_{\left\lfloor \frac{n}{2} \right\rfloor} \right) < c_2 n^{2/3},$$

với hầu hết mọi đồ thị $G_{\left\lfloor \frac{n}{2} \right\rfloor}$.

iii) Nếu $c > 1$ thì, với hầu hết mọi đồ thị ngẫu nhiên G_t ,

$$|L^{(1)}(G_t) - \gamma n| \leq \omega(n) n^{1/2},$$

trong đó $0 < \gamma = \gamma(c) < 1$ là nghiệm duy nhất của phương trình $e^{-c\gamma} = 1 - \gamma$.

Hơn nữa,

$$\left| L^{(i)}(G_t) - \frac{1}{\alpha} \left\{ \log n - \frac{5}{2} \log \log n \right\} \right| \leq \omega(n),$$

với mọi i , $2 \leq i \leq h$.

Để làm rõ bản chất của Kết quả 7.5.2), chúng ta hãy xem xét xem điều gì sẽ xảy ra với $L^{(1)}(G_t)$, bậc lớn nhất có thể có của một thành phần liên thông, trong một quá trình đồ thị ngẫu nhiên điển hình, khi t tăng từ $\left\lfloor \frac{cn}{2} \right\rfloor$ lên $\left\lfloor \frac{c'n}{2} \right\rfloor$, trong đó $c' = \frac{100}{99}c$. Nếu $0 < c < c' < 1$ hoặc $1 < c < c'$ thì $L^{(1)}(G_t)$ tăng theo một hệ số hằng. Tuy nhiên, nếu $c < 1 < c'$ thì $L^{(1)}(G_t)$ tăng lên một cách khá đáng kể, từ bậc $\log n$ lên thành bậc n . Khi đi qua điểm tới hạn $c = 1$, cấu trúc của thành phần liên thông thay đổi hoàn toàn về chất và **hiện tượng chuyển pha** xảy ra: trước điểm tới hạn, chúng ta có khá nhiều thành phần liên thông với bậc tối đại và bậc

tối đại này vào khoảng $O(\log n)$; nhưng sau điểm tới hạn, chúng ta chỉ có duy nhất một thành phần liên thông với bậc tối đại và nó cũng lớn hơn khá nhiều so với thành phần liên thông lớn thứ hai.

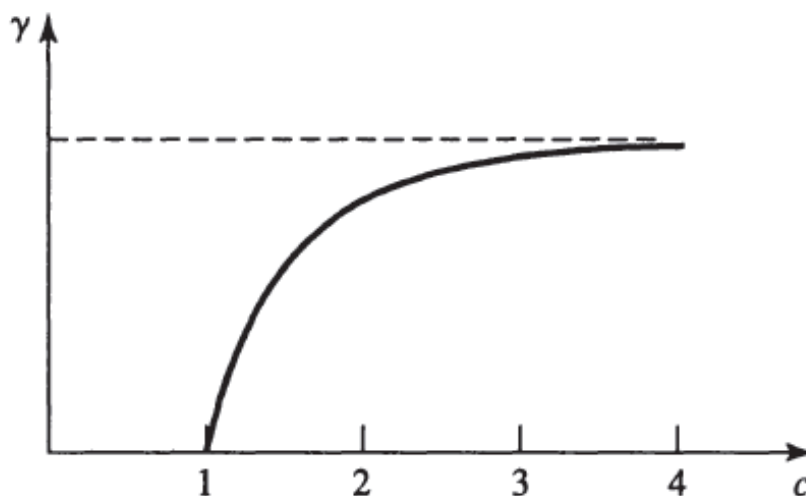
Nói cách khác, khi đi qua điểm tới hạn, một thành phần liên thông khổng lồ sẽ xuất hiện với khoảng γn đỉnh, với $\gamma = \gamma(c)$ là một hằng số dương nào đó, trong khi thành phần liên thông lớn thứ hai vẫn có bậc $O(\log n)$.

Đề thấy rằng $\gamma = \gamma(c)$ thật sự được xác định rõ ràng với mọi $c > 1$, kí hiệu $f_c(\gamma) = c\gamma$ và $g(\gamma) = -\log(1 - \gamma) = \gamma + \frac{\gamma^2}{2} + \frac{\gamma^3}{3} + \dots$.

Do, với mọi $\gamma \geq 0$, hàm $g(\gamma)$ là lồi chặt, $g'(0) = 1$ và $g(\gamma) \rightarrow \infty$ khi $\gamma \rightarrow 1^-$ nên thực sự chúng ta luôn có thể tìm được duy nhất một số thực $\gamma \in (0, 1)$ sao cho $f_c(\gamma) = g(\gamma)$, nghĩa là $e^{-c\gamma} = 1 - \gamma$.

Trên thực tế, chúng ta cũng có thể dễ dàng kiểm tra được rằng, với $c = 1 + \varepsilon > 1$,

$$\gamma(c) = \gamma(1 + \varepsilon) = 2\varepsilon - \frac{8}{3}\varepsilon^2 + \frac{28}{9}\varepsilon^3 + O(\varepsilon^4).$$



Hình vẽ minh họa cho đường cong $\gamma(c)$, với mọi $c > 1$.

Điều gì sẽ xảy ra ở gần điểm tới hạn vẫn còn là một bí ẩn trong hơn 20 năm với nhiều câu hỏi chưa được trả lời như: “Trong một quá trình đồ thị điển hình, với những giá trị nào của t thì thành phần liên thông lớn nhất sẽ lớn tối thiểu gấp đôi thành phần liên thông lớn thứ hai? Thành phần liên thông lớn thứ hai có thể lớn đến mức nào? Liệu nó có thể phát triển đến $n/\log n$ không? Một khi chúng ta nhận thấy sự xuất hiện của thành phần liên thông khổng lồ của một quá trình đồ thị thì tốc độ phát triển của nó là bao nhiêu?”

Những câu hỏi như thế này đã được giải quyết bởi Bollobás vào năm 1984, giúp chúng ta làm sáng tỏ rất nhiều vấn đề về hiện tượng chuyển pha, và nhiều kết quả chi tiết hơn đã được chứng minh bởi Łuczak vào năm 1990, và bởi Janson, Knuth, Łuczak, Pittel vào năm 1993.

Chẳng hạn, nếu $t = \frac{n}{2} + s$, $s/n^{2/3} \rightarrow \infty$, $s/n \rightarrow 0$ thì $L^{(1)}(G_t) = 4s + o(s)$, với hầu hết mọi đồ thị G_t ; nghĩa là, xét một cách trung bình, việc chúng ta bổ sung thêm một cạnh sẽ làm tăng thêm 4 đỉnh mới vào thành phần liên thông khổng lồ. Ngoài ra, $L^{(2)}(G_t)$ không tăng trưởng đáng kể trên $n^{2/3}$: chẳng hạn, trong $\tilde{G}^n$, chúng ta sẽ có $\lim_{n \rightarrow \infty} P(L^{(2)}(G_t) \geq n^{2/3+\varepsilon} \text{ với mọi } t) = 0$, với mọi $\varepsilon > 0$.

Chúng ta sẽ không đưa ra bất kì một phép chứng minh cụ thể nào cho Kết quả 7.5.2), nhưng chúng ta vẫn sẽ phác thảo một cách tiếp cận rất tao nhã do Karpn đề xuất cho việc chứng minh của Kết quả 7.5.2). Ý tưởng cốt lõi của Karpn là bỏ qua cấu trúc của một thành phần liên thông, tập trung vào phép tương ứng $x \mapsto |C(x)|$, trong đó $C(x)$ là tập đỉnh của thành phần liên thông có chứa đỉnh x trong đồ thị ngẫu nhiên của chúng ta và khai thác sự tương đồng của nó với một phân phối xác suất. Như thường lệ, sẽ thuận tiện hơn khi chúng ta làm việc với mô hình $\mathcal{G}(n, p)$ thay vì $\mathcal{G}(n, M)$: như trong mục 7.2. **Các tính chất cơ bản của hầu hết tất cả các đồ thị - Kỹ thuật sử dụng kỳ vọng**, khi đó chúng ta có thể dễ dàng chuyển đổi từ mô hình này sang mô hình khác.

Giả sử cho trước một đồ thị G và một đỉnh $x \in V$. Kí hiệu $C_G(x)$ là tập đỉnh của thành phần liên thông có chứa đỉnh x trong G . Để xây dựng $C = C_G(x)$, chúng ta sẽ tiến hành như sau: Kí hiệu $x_1 = x$, $A_0 = \emptyset$, $B_0 = \{x_1\}$. Khi đó, chúng ta sẽ kí hiệu $A_1 = \{x_1\}$ và bổ sung thêm vào B_0 tất cả các đỉnh láng giềng của x_1 trong G để thu được B_1 . Nếu $B_1 = A_1$ thì $C = B_1 = A_1$; trường hợp ngược lại, chúng ta sẽ chọn ra một đỉnh $x_2 \in B_1 - A_1$, kí hiệu $A_2 = A_1 \cup \{x_2\}$ và bổ sung thêm vào B_1 tất cả các đỉnh láng giềng của x_2 trong G để thu được B_2 . Nếu $B_2 = A_2$ thì $C = B_2 = A_2$; trường hợp ngược lại, chúng ta sẽ chọn ra một đỉnh $x_3 \in B_2 - A_2$, kí hiệu $A_3 = A_2 \cup \{x_3\}$ và bổ sung thêm vào B_2 tất cả các đỉnh láng giềng của x_3 trong G để thu được B_3 . Bằng cách thực hiện lặp lại quy trình trên, chúng ta sẽ nhận được một tập $C = B_l = A_l$; đây chính xác là tập $C_G(x)$ cần tìm. Lưu ý rằng tập này chỉ phụ thuộc vào G chứ không phụ thuộc vào các lựa chọn mà chúng ta đã thực hiện trong quá trình xây dựng. Tại thời điểm i , A_i là tập các đỉnh trong C mà chúng ta đã kiểm tra các đỉnh láng giềng. Do đó, tại thời điểm l , chúng ta hết các đỉnh mới để kiểm tra. Như vậy, $A_l = B_l$ chính xác là thành phần liên thông $C_G(x)$ cần tìm.

Lưu ý rằng trong cách xây dựng nêu ở trên, $|A_i| = i$, $A_i \subset B_i$, với $i = 0, 1, \dots, l$, và l là chỉ số đầu tiên sao cho $A_l = B_l$. Kế tiếp, nếu G là một đồ thị ngẫu nhiên $G_{n,p}$ thì, với $A_i \subset B_i$, $A_i \neq B_i$ đã được xây dựng, xác suất để một đỉnh trong $V - B_i$ được đưa vào B_{i+1} chính là p , xác suất này độc lập với tất cả các đỉnh khác và tất cả các đỉnh đã được lựa chọn trước đó. Do đó, chúng ta có thể cho chạy quy trình trên mà không cần tham chiếu đến các tập A_i (tập các đỉnh đã được kiểm tra trong i vòng đầu tiên): sau khi đã xây dựng B_i , chọn ra các đỉnh nằm trong $V - B_i$ độc lập với nhau và bổ sung thêm các đỉnh đã chọn của B_i để thu được B_{i+1} . Dãy $B_0 \subset B_1 \subset \dots$ có thể được kéo dài đến vô hạn. Tuy nhiên, các kết nối với $C(x)$ lại rất dễ được phục hồi: bằng cách dùng l để kí hiệu cho chỉ số nhỏ nhất sao cho $|B_l| = l$, tập B_l được phân phối chính xác như $C(x)$. Đặc biệt, xác suất để $C(x)$ có đúng k đỉnh tối đa là xác suất để $|B_k|$ đúng bằng k .

Chìa khóa mấu chốt cho vấn đề đã xuất hiện: mỗi $|B_i|$ đều có một phân phối rất đơn giản. Thật vậy, câu hỏi được đặt ra cho chúng ta ngay bây giờ đó là: “Xác suất để một đỉnh $y \in [n] - \{x\}$ không được đưa vào B_i là bao nhiêu?”. Chúng ta sẽ cố gắng thực hiện i lần bổ sung thêm y vào B_i , với mỗi lần thử thành công có xác suất p , nên $P(y \notin B_i) = (1 - p)^i$. Do tất cả các lựa chọn là độc lập nên $|B_i|$ có phân phối nhị thức với các tham số $n - 1$ và $1 - (1 - p)^i$:

$$P(|B_i| = k) = \binom{n-1}{k} (1 - (1 - p)^i)^k (1 - p)^{i(n-k-1)}.$$

Như đã trình bày trong lưu ý ở trên, xác suất này là một chặn trên của $P_p(|C(x)| = k)$.

Mối quan hệ này cho phép chúng ta chỉ ra rằng có một số các giá trị sẽ cực kỳ khó có thể xuất hiện như các bậc của các thành phần liên thông của $G_{n,p}$: với xác suất cao, chúng ta sẽ có một khoảng trống trong dãy các bậc của các thành phần liên thông.

Kết quả 7.5.3. (Karpn)

Giả sử cho trước một số thực $a \geq 2$ cố định.

Khi đó, nếu n đủ lớn, $\varepsilon = \varepsilon(n) < 1/3$, $p = p(n) = \frac{1+\varepsilon}{n}$ thì, với xác suất $\geq 1 - n^{-a}$, $G_{n,p}$ không có chứa bất kì một thành phần liên thông nào với bậc k thỏa mãn

$$\frac{8a}{\varepsilon^2} \log n \leq k \leq \frac{\varepsilon^2}{12} n.$$

Chứng minh kết quả 7.5.3.

Kí hiệu $k_0 = \lceil 8a\varepsilon^{-2} \log n \rceil$ và $k_1 = \lfloor \varepsilon^2 n / 12 \rfloor$. Khi đó, bằng cách dùng p_k để kí hiệu xác suất để một thành phần liên thông của $G_{n,p}$ có chứa một đỉnh cố định có k đỉnh, xác suất để $G_{n,p}$ có một thành phần liên thông với bậc k tối đa là np_k . Như vậy, để chứng minh tính đúng đắn của khẳng định trên, chúng ta chỉ cần chỉ ra rằng

$$\sum_{k=k_0}^{k_1} p_k \leq n^{-a-1}.$$

Không giảm tính tổng quát của vấn đề, chúng ta giả định rằng $k_0 \leq k_1$, nên $\varepsilon^4 \geq 96a (\log n)/n \geq 1/n$, bởi vì trong trường hợp ngược lại tính đúng đắn của khẳng định trên là hiển nhiên.

Khi đó, từ $P(|B_i| = k) = \binom{n-1}{k} (1 - (1-p)^i)^k (1-p)^{i(n-k-1)}$, chúng ta suy ra rằng

$$p_k \leq P(|B_i| = k) \leq \frac{n^k}{k!} e^{-k^2/2n} (kp)^k (1-p)^{k(n-k-1)},$$

bởi vì

$$\frac{n-1}{k} = \frac{n^k}{k!} \prod_{j=1}^k \left(1 - \frac{j}{n}\right) \leq \frac{n^k}{k!} e^{-k^2/2n},$$

và

$$(1-p)^k \geq 1 - kp.$$

Lưu ý rằng, với mọi $|\varepsilon| \leq 1/3$, $1 + \varepsilon \leq e^{\varepsilon - \varepsilon^2/3}$ và, từ công thức Stirling, chúng ta sẽ có

$$\begin{aligned} p_k &\leq \exp\{-k^2/2n - \varepsilon^2 k/3 + k^2(1 + \varepsilon)/n\} \\ &\leq \exp\{-\varepsilon^2 k/3 + k^2/n\} \\ &\leq e^{-\varepsilon^2 k/4}. \end{aligned}$$

Như vậy,

$$\begin{aligned}
\sum_{k=k_0}^{k_1} p_k &\leq \sum_{k=k_0}^{k_1} e^{-\varepsilon^2 k/4} \\
&= e^{-\varepsilon^2 k_0/4} (1 - e^{-\varepsilon^2/4}) \\
&\leq \frac{5}{\varepsilon^2} e^{-\varepsilon^2 k_0/4} \\
&\leq nn^{-2a} \\
&\leq n^{-a-1}.
\end{aligned}$$

□

Với việc thực hiện thêm một số công việc khác nữa, Kết quả 7.5.3) có thể được chứng minh lại trong một phiên bản mạnh hơn nhiều, phiên bản này cho chúng ta thấy rằng, trong một quá trình phát triển đều đặn, khoảng trống lớn hơn nhiều sẽ xuất hiện ngay sau thời điểm $n/2$ và kéo dài cho đến khi kết thúc quá trình này. Tuy nhiên, ngay cả trong hình thức này, nó cũng cho chúng ta biết rất nhiều thông tin quan trọng về các thành phần liên thông trong một quá trình đồ thị ngẫu nhiên. Chẳng hạn, giả sử cho trước các số thực dương $0 < \varepsilon_1 < \varepsilon_2 < 1/3$, chúng ta luôn có thể tìm được các hằng số dương α, β sao cho, với hầu hết mọi quá trình đồ thị $\tilde{G} = (G_t)_{t=0}^N$ sao cho $(1 + \varepsilon_1)n/2 \leq t \leq (1 + \varepsilon_2)n/2$, đồ thị G_t không chứa bất kì một thành phần liên thông nào có bậc nằm giữa $\alpha \log n$ và βn . Chúng ta sẽ gọi một thành phần liên thông là **nhỏ** nếu nó có $\leq \alpha \log n$ đỉnh và gọi nó là **lớn** nếu nó có $\geq \beta n$ đỉnh. Ngoài ra, kí hiệu $t_1 = \lfloor (1 + \varepsilon_1)n/2 \rfloor$ và $t_2 = \lfloor (1 + \varepsilon_2)n/2 \rfloor$. Khi đó, với hầu hết mọi quá trình đồ thị điển hình $(G_t)_{t=0}^N$, với mọi chỉ số t sao cho $t_1 \leq t \leq t_2$, mọi thành phần liên thông của G_t đều là nhỏ hoặc lớn.

Chúng ta sẽ khảo sát những thay đổi trong cấu trúc của các thành phần liên thông trong một quá trình điển hình $(G_t)_{t=0}^N$ khi t tăng từ t_1 lên t_2 . Tác động của việc bổ sung thêm một cạnh vào G_t để tạo ra G_{t+1} là gì? Nếu cạnh mới được bổ sung thêm vào một thành phần liên thông bất kì thì sẽ không có thay đổi gì xảy ra. Nếu cạnh mới nối một thành phần liên thông lớn với một thành phần liên thông khác thì hai thành phần liên thông đó sẽ được thay thế bởi một thành phần liên thông mới duy nhất. Quan trọng nhất, điều gì sẽ xảy ra nếu cạnh mới nối hai thành phần liên thông nhỏ? Hợp của các thành phần liên thông nhỏ này chắc chắn không thể lớn, bởi vì $2\alpha \log n < \beta n$, nên nó bắt buộc phải nhỏ. Đặc biệt, G_{t+1} có số thành phần liên thông lớn nhất tối đa nhiều như G_t ; thậm chí, nếu cạnh mới nối hai thành phần liên thông lớn thì G_{t+1} có số thành phần liên thông lớn nhất ít hơn một đơn vị so với G_t .

Với điều kiện $\tilde{G} = (G_t)_{t=0}^N$ là một quá trình đồ thị điển hình và G_t có chứa ít nhất hai thành phần liên thông lớn, xác suất để cạnh thứ $(t + 1)$ nối hai thành phần liên thông lớn nhất $\geq \beta^2 n^2 / \binom{n}{2} > 2\beta^2$. Kế tiếp, G_{t_1} có tối đa $1/\beta$ thành phần liên thông lớn, nên nếu $\omega(n) \rightarrow \infty$ thì, với xác suất $1 - o(1)$, sau khi bổ sung thêm $\omega(n)$ cạnh tiếp theo, tất cả các thành phần liên thông lớn nhất đều được hợp nhất thành một. Đặc biệt, trong một quá trình đồ thị điển hình G_{t_2} có duy nhất một thành phần liên thông lớn, thành phần liên thông khổng lồ của G_{t_2} và tất cả các thành phần liên thông khác đều có tối đa là $c_1 \log n$ đỉnh. Đây chính xác là phiên bản định tính của khẳng định (iii) của Kết quả 7.5.2).

Giá trị γ nêu trong khẳng định (iii) của Kết quả 7.5.2) có thể được xác định bằng cách khai thác sự tương tự giữa một phân phối xác suất và sự tăng trưởng của một thành phần liên thông có chứa một đỉnh cố định cho trước. Để tạo sự thuận tiện, chúng ta sẽ đưa ra một mô tả heuristic ngắn gọn về sự tương tự này.

Kí hiệu $p = c/n$, $c > 1$. Giả sử chúng ta đã biết rằng hầu hết mọi đồ thị $G_{n,p}$ đều có một thành phần liên thông khổng lồ duy nhất với $(\gamma + o(1))n$ đỉnh, trong đó $\gamma > 0$, và tất cả các thành phần liên thông khác đều nhỏ, với không quá $\alpha \log n$ đỉnh. Làm thế nào để chúng ta có thể tìm được γ ? Hiển nhiên, γ là giới hạn của xác suất để thành phần liên thông $C(x)$ có chứa một đỉnh x cố định cho trước là nhỏ: $|C(x)| \leq \alpha \log n$. Để ước tính xác suất này, chúng ta sẽ cho “tăng” $C(x)$, bắt đầu từ x , như trong chứng minh của Kết quả 7.5.3), nhưng vẫn theo dõi các láng giềng mà chúng ta đưa vào $C(x)$. Nói một cách chính xác, kí hiệu U_l là tập các đỉnh có khoảng cách đến x bằng l và kí hiệu $V_l = \bigcup_{i=1}^l U_i$ là tập các đỉnh có khoảng cách đến x không vượt quá l . Chúng ta sẽ dừng quá trình này lại nếu $|V_l| > \alpha \log n$ với một chỉ số l nào đó, bởi vì khi đó $C(x)$ lớn, và nếu $U_{l+1} = \emptyset$ và $|V_l| \leq \alpha \log n$ thì $C(x) = V_l$ là nhỏ.

Chúng ta sẽ xem xét cách mà U_{l+1} phát sinh từ (U_l, V_l) . Kí hiệu $U_l = \{u_1, u_2, \dots\}$, đầu tiên chúng ta lấy ra tất cả các đỉnh láng giềng mới của u_1 , sau đó là tất cả các đỉnh láng giềng mới của u_2 , v.v... và chúng ta sẽ dừng quá trình trên lại nếu chúng ta đạt đến $\alpha \log n$ đỉnh. Giả sử rằng chúng ta đã đạt tới $h \leq \alpha \log n$ đỉnh từ x khi chúng ta kiểm tra các đỉnh láng giềng mới của u_i . Phân phối xác suất của số các đỉnh láng giềng mới của u_i là gì?

Rõ ràng, với mọi $k \leq \alpha \log n$, chúng ta đều có

$$\begin{aligned} P(u_i \text{ có } k \text{ đỉnh láng giềng mới}) &= \binom{n-h}{k} p^k (1-p)^{n-h-k} \\ &= \frac{c^k}{k!} e^{-c} (1 + O((\log n)^2/n)). \end{aligned}$$

Điều này có nghĩa là phân phối của số các hậu duệ của u_i , nghĩa là số các đỉnh mà chúng ta bổ sung thêm vào $C(x)$ vì u_i , gần với phân phối Poisson ứng với giá trị trung bình c . Như vậy, phân phối của $|C(x)|$ gần với phân phối của tổng dân số trong quá trình ngẫu nhiên Poisson, với điều kiện $C(x)$ nhỏ.

Để định nghĩa quá trình này, kí hiệu $Z_{i,j}$, $i = 0, 1, \dots, j = 1, 2, \dots$, là các biến ngẫu nhiên Poisson độc lập, mỗi biến đều ứng với giá trị trung bình c :

$$P(Z_{ij} = k) = \frac{c^k}{k!} e^{-c}.$$

Kí hiệu $Z_0 = 1$. Giả sử chúng ta đã xác định Z_n , kí hiệu $Z_{n+1} = Z_{n,1} + Z_{n,2} + \dots + Z_{n,Z_n}$, trong đó, như thường lệ, chúng ta sẽ quy ước tổng rỗng có giá trị là 0. (Z_n có thể được hiểu là kích thước của quần thể ở thế hệ thứ n và $Z_{n,i}$ là số các hậu duệ của thành viên thứ i của thế hệ thế hệ thứ n)

Kết quả 7.5.4.

Giả sử cho trước quá trình ngẫu nhiên $(Z_n)_{n=0}^\infty$, với $c > 1$. Kí hiệu p_∞ là xác suất để $Z_n > 0$ với mọi n .

Khi đó, p_∞ là nghiệm duy nhất của phương trình $e^{-cp_\infty} = 1 - p_\infty$ trong khoảng $(0, 1)$.

Chứng minh kết quả 7.5.4.

Kí hiệu p_n là xác suất để $Z_n > 0$. Khi đó, $p_0 = 1$ và $p_\infty = \lim_{n \rightarrow \infty} p_n$. Đầu tiên, chúng ta sẽ chứng minh bằng phương pháp quy nạp toán học theo chỉ số n rằng $p_n \geq \gamma$, với mọi n , trong đó γ là nghiệm duy nhất

của phương trình $e^{-c\gamma} = 1 - \gamma$ trong $(0,1)$. Điều này hiển nhiên đúng với $n = 0$, bởi vì $p_0 = 1 \geq \gamma$. Giả sử $n \geq 0$ và $p_n \geq \gamma$. Với điều kiện $Z_1 = k \geq 1$, quá trình này là tổng của k quá trình độc lập có cùng một phân phối. Do $1 - p_t$ là xác suất để quá trình này kết thúc tại thời điểm t , nên

$$\begin{aligned} 1 - p_{n+1} &= P(Z_1 = 0) + \sum_{k=1}^{\infty} P(Z_1 = k)(1 - p_n)^k \\ &= \sum_{k=0}^{\infty} \frac{c^k}{k!} e^{-c} (1 - p_n)^k \\ &= e^{-cp_n} \sum_{k=0}^{\infty} \frac{(c(1 - p_n))^k}{k!} e^{-c(1-p_n)} \\ &= e^{-cp_n} \\ &\leq e^{-c\gamma}. \end{aligned}$$

Do đó, $p_{n+1} \geq \gamma$, như đã khẳng định, nên $p_{\infty} = \lim_{n \rightarrow \infty} p_n \geq \gamma$.

Bằng cách áp dụng lập luận trên cho $1 - p_{\infty}$ thay vì $1 - p_{n+1}$ và $1 - p_n$, chúng ta nhận thấy rằng

$$1 - p_{\infty} = P(Z_1 = 0) + \sum_{k=1}^{\infty} P(Z_1 = k)(1 - p_{\infty})^k = e^{-cp_{\infty}}.$$

Như vậy, p_{∞} là một nghiệm của phương trình $e^{-cp_{\infty}} = 1 - p_{\infty}$ thỏa mãn điều kiện $0 < p_{\infty} \leq 1$. □

Quay trở lại với kích thước $(\gamma + o(1))n$ của thành phần liên thông khổng lồ trong khẳng định (iii) của Kết quả 7.5.2), chúng ta biết rằng γ là giới hạn của xác suất để một đỉnh cố định cho trước $x \in [n]$ nằm trong thành phần liên thông khổng lồ. Do đó, γ là xác suất để Z_n nêu trong Kết quả 7.5.4) không bị chết, nên $\gamma = p_{\infty}$. Như vậy, $e^{-c\gamma} = 1 - \gamma$ như trong khẳng định (iii) của Kết quả 7.5.2).

Để làm cho tất cả những điều này trở nên chính xác về mặt học thuật, chúng ta sẽ phải làm nhiều việc hơn nữa, nhưng rõ ràng là phương pháp này có thể được sử dụng để thiết lập các đặc điểm chính của hiện tượng chuyển pha. Trên thực tế, phương pháp trên chỉ là một trong số một số cách khác nhau để chúng ta có thể khảo sát hiện tượng chuyển pha. Đặc biệt, Erdős, Rényi, Bollobás, Łuczak sử dụng một cấu trúc mịn hơn của các thành phần liên thông còn Janson, Knuth, Łuczak, Pittel dựa vào các hàm sinh và các công cụ giải tích mạnh để thu được các kết quả rất chi tiết về sự xuất hiện của thành phần liên thông khổng lồ.