

Về một điều kiện đủ cho đồ thị ngẫu nhiên đường kính nhỏ, giúp phân tích mạng thế giới nhỏ

Analyzing Small-Worlds: A Sufficient Condition for Obtaining Small Diameter in A New Random Graph Model

Nguyễn Khanh Văn

Abstract: Network structures and graphs that feature the Small-World property have drawn a strong interest from the research community in two perspectives: 1) The Small-world effect is a popular phenomenon amongst several real-world complex networks; 2) Small-world graphs are considered very useful tools to model real-world complex networks as well as to design new topologies for certain applications in computer networks. We propose to study a new general random graph model that can be used to analyze the small-world effect (such an approach is already widely used). Our main result is to construct a general sufficient condition for this graph model to have logarithmic diameter, i.e. $O(\log n)$ for n as the number of the vertices. This result can help to assess and analyze several new graphs model for small-worlds.

Keyword: Small-world networks, diameter, routing, random graphs, network design.

I. GIỚI THIỆU

Tính chất thế-giới-nhỏ (TGN) là một đặc trưng tương đối phổ quát trong rất nhiều cấu trúc mạng phức tạp được quan sát trong rất nhiều mặt của khoa học và đời sống, như các mạng xã hội, các mạng sinh học, mạng lưới cung cấp điện, hay các mạng liên kết vật lý của Internet. Sự biểu lộ của TGN được mô tả bởi hai yếu tố: 1) có đường kính đồ thị rất nhỏ (thường là đa thức của logarit của kích thước mạng) và 2) xu hướng tạo cộng đồng con (clustering). Hiện tượng TGN lần đầu tiên được đề cập trong giới khoa học qua

công trình của Milgram trong mạng xã hội dựa trên thí nghiệm chuyển thư giới thiệu để xác lập chuỗi quen biết [18]. Watt và Strogatz chính thức đặt nền móng cho địa hạt nghiên cứu này [22], thu hút sự quan tâm từ nhiều giới khác nhau, như toán học và vật lý (nghiên cứu về cấu trúc chung), xã hội học (các mạng xã hội, ví dụ mạng quan hệ diễn viên, quan hệ đồng tác giả, quan hệ qua Facebook), sinh học (các mạng sinh học) và các nhà tin học (Internet và các dạng mạng máy tính và cấu trúc tô-pô liên kết). Kleinberg đưa ra một mô hình cơ bản khác [10], phát triển từ mô hình Watt-Strogatz, đem lại một cách nhìn mới cho hiện tượng này, đậm nét ý nghĩa thuật toán (định tuyến và tìm kiếm thông tin).

Các mô hình TGN nói trên đều dựa vào một tiếp cận cơ bản. Đó là việc tạo ra một tô-pô mạng ngẫu nhiên thông qua việc cải biến một đồ thị cơ sở ban đầu bằng cách thêm vào (hay thay thế bằng) các mối liên kết ngẫu nhiên (random link). Thông thường các đồ thị cơ sở là khá đơn giản, có thể chỉ là một lưới một chiều hoặc nhiều chiều hơn (ring, grid,...), còn các liên kết ngẫu nhiên có thể tuân theo một phân phối xác suất tương đối đơn giản nào đó. Trong mô hình Watt-Strogatz, đó là phân phối ngẫu nhiên đồng nhất (uniform random) còn trong mô hình Kleinberg, đó là phân bố có tính tới yếu tố khoảng cách địa lý, được mô tả chi tiết trong phần II của bài báo này. Rất nhiều mô hình hoặc thiết kế mạng ứng dụng tính chất TGN cũng sử dụng tiếp cận này. Tuy nhiên còn rất ít nghiên cứu đề cập một mô hình mang tính tổng quát theo tiếp cận nói trên và đưa ra những điều kiện tổng quát để

đảm bảo cho đường kính nhỏ, điều kiện tiên quyết của tính TGN.

Trong bài báo này, chúng tôi đưa ra một cách tiếp cận tổng quát, đề xuất một mô hình đồ thị ngẫu nhiên khái quát, sử dụng tiếp cận “thêm liên kết ngẫu nhiên vào một đồ thị cơ sở” nói trên. Chúng tôi khảo sát mô hình này và cho thấy nhiều mô hình TGN và thiết kế tô-pô cụ thể đã có có thể coi là trường hợp riêng của mô hình phổ quát này. Mặt khác, đóng góp chính của chúng tôi là *đề xuất một điều kiện đủ tổng quát đảm bảo tính chất đường kính nhỏ ở bậc lo-ga-rit của kích thước mạng*. Định lý tổng quát này cho phép khảo sát rất hiệu quả đường kính của một loạt mạng TGN đã biết, qua đó thể hiện là một công cụ hữu ích cho việc phân tích các đồ thị TGN mới, đồng thời có thể giúp xây dựng các thiết kế tô-pô TGN cho nhiều mô hình mạng thực tế. Trong nhiều trường hợp ta có thể chứng minh đường kính đồ thị $O(\log n)$ nhanh gọn hơn rất nhiều so với cách làm đã biết trước kia.

Sau đây để thuận tiện, các chữ viết tắt ĐTNN thay cho “đồ thị ngẫu nhiên” và LKNN thay cho “liên kết ngẫu nhiên” sẽ được dùng trong bài báo.

II. TỔNG QUAN VỀ MÔ HÌNH THẾ GIỚI NHỎ VÀ NGHIÊN CỨU LIÊN QUAN

Trong mục này chúng tôi xin giới thiệu một số mô hình TGN cơ bản và những kết quả nghiên cứu quan trọng liên quan.

Watts và Strogatz đã đặt nền tảng cho khái niệm mạng TGN và các cơ sở [22]. Trong mô hình Watts-Strogatz, đồ thị cơ sở là một tập các nút trên một vòng tròn (*ring lattice*, tức lưới 1 chiều), mà mỗi nút đều có cạnh nối đến k nút láng giềng gần nhất (trên *ring* này) với một tham số k cho trước. Sau đó với mỗi cạnh (u,v) trên đồ thị cơ sở, với một xác suất β (tham số cho trước), thay thế nó bằng một LKNN (u,w) mà w được chọn theo ngẫu nhiên đồng nhất, nhưng tránh để xảy ra liên kết vòng hoặc đúp. Với β là hằng số dương <1 , đồ thị thu được thể hiện hai đặc trưng TGN là đường kính nhỏ và hệ số tương hỗ (clustering coefficient) lớn đáng kể.

Nghiên cứu hiện tượng tìm đường thành công (*searchability*) mà chỉ dùng thông tin cục bộ trong thí nghiệm Milgram, mô hình Kleinberg đã đưa ra cải tiến như sau: đồ thị cơ sở là một lưới 2 chiều, mỗi nút có 4 liên kết cơ sở đến các nút láng giềng. Sau đó với mỗi nút u ta thêm vào k LKNN đến các đỉnh v lựa chọn theo luật tỷ lệ nghịch với hàm lũy thừa khoảng cách: xác suất của sự kiện “ u nối với v ” là $\sim d^{-\alpha}(u,v)$; trong đó d là khoảng cách *mahantan* (đường đi trên lưới ngắn nhất giữa u và v) còn α là một tham số hệ thống.

Kleinberg chứng minh rằng với $\alpha=2$ và chỉ duy nhất giá trị này thì mô hình với đầy đủ tính chất “*searchability*” đã phản ánh thực tế: với 2 nút $s-t$ đã cho, chỉ bằng giải thuật định tuyến tham lam và chỉ cần sử dụng thông tin cục bộ của các nút đã đi qua, thì vẫn tìm được đường đi $s-t$ có độ dài rất ngắn (cụ thể là $O(\log^2 n)$). Tuy nhiên sau đó, Leebhar & Schabanel [13] đã chỉ ra đó không phải là đường đi tối ưu, đồng thời Martel và Nguyễn cho thấy đường kính của đồ thị chỉ là $\theta(\log n)$ [17] (hơn thế nữa, khi $\alpha > 4$, đồ thị sẽ trở thành “thế-giới-lớn” có đường kính là $\theta(n^c)$, với c là hằng số dương). Trong [7,8], các tác giả đã tổng quát hóa mô hình Kleinberg với việc sử dụng đồ thị cơ sở tùy ý để khảo sát tính “*searchability*” dạng khái quát. Bài toán xây dựng đường đi tối ưu mà chỉ viếng thăm số nút hạn chế đã được giải quyết trong [9].

Bên cạnh hai thuộc tính TGN (về đường kính và hệ số tương hỗ), nhiều mạng phức hợp trong thế giới thực còn thể hiện một thuộc tính phổ quát khác là phân bố bậc đỉnh không đều, mà trái lại theo hàm lũy thừa. Tính chất phổ quát này thường được gọi là Luật Lũy thừa (power law), một địa hạt rất được quan tâm và nghiên cứu đa ngành. Trong số nhiều mô hình nghiên cứu, chúng tôi chú ý đến mô hình quan trọng của Chung-Lu vì nó cũng chia sẻ nhiều điểm chung với các dạng mô hình thế giới nhỏ quan tâm trong bài báo này [3,4].

Mô hình Chung-Lu, ký hiệu $G(w)$, có đồ thị cơ sở là một tập n đỉnh mà được gán với một vec-tơ trọng số $w=(w_1, w_2, \dots, w_n)$ thể hiện chuỗi giá trị bậc đỉnh chờ đợi ở đồ thị thu được (“sequence of expected

degrees”). Các LKNN được sinh độc lập cho mỗi cặp đỉnh u_i và u_j với một xác suất $\sim w_i w_j$. Các tác giả đã sử dụng mô hình này để nghiên cứu thành công một vài dạng đồ thị thế giới thực như đồ thị liên kết vật lý Internet, mà ở đó luật Lũy thừa được quan sát thấy. Sử dụng mô hình đó các tác giả cũng đã cho thấy tính chất TGN, tức là đường kính nhỏ, thể hiện trên các đồ thị liên kết Internet.

Khảo sát trên cho thấy tính phổ biến của mô thức xây dựng mô hình qua sự kết hợp của một đồ thị cơ sở giàu liên kết cục bộ với một phân phối liên kết ngẫu nhiên nào đó, mà có thể phản ánh yếu tố địa lý (mô hình Kleinberg). Điều này đặc biệt phù hợp với đồ thị liên kết vật lý Internet. Yook và cộng sự đã quan sát thấy qui luật liên kết vật lý Internet có xu hướng phổ biến hơn với khoảng cách ngắn hơn [23]. Tính “distance bias” này cũng được quan sát trong một nghiên cứu về chiến thuật tìm đường địa lý (“geographic routing”) trong mạng xã hội [14]. Trong bài báo nổi tiếng về Luật Lũy thừa của Faloutsos et al [6] cũng đề cập nhiều yếu tố địa lý trong việc hình thành mạng Internet.

Các mô hình TGN đem lại nhiều ứng dụng rất lý thú trong nhiều bài toán thiết kế tô pô mạng đa dạng. Ứng dụng mô hình TGN của Kleinberg đã tạo nên một số kiến trúc mạng đồng đẳng có hiệu năng cận tối ưu [15,16]. Việc sử dụng các LKNN theo mẫu hình TGN đem lại một tiếp cận mới trong thiết kế mạng liên kết cho các siêu máy tính hay các trung tâm dữ liệu cỡ lớn hiện đại. So với các tiếp cận truyền thống hơn, tiếp cận mới cho phép giảm thiểu độ trễ truyền tin [12], nâng cao khả năng mở rộng mạng mềm dẻo (scalability) [21], hay tối ưu chi phí cáp truyền tin [20].

ĐTNN truyền thống được Erdős và Rényi và đặt nền móng [5]; các kết quả quan trọng đã được hệ thống hóa và trình bày lại trong sách chuyên khảo của Bollobas [2]. Bài báo này thừa kế một tiếp cận nghiên cứu chung về TGN đã phát triển từ lâu của chúng tôi [17,19,20]. Bài báo này có thể nói là kết quả của sự đi sâu hơn nữa theo hướng tổng quát hóa, nhằm tìm đến những tri thức chung nhất về TGN (tức là đề xuất điều

kiện đủ để đạt đường kính đồ thị logarit), có nghĩa là hướng tới đóng góp vào lý thuyết cơ bản đang thành hình về TGN. Ý tưởng khái quát, đồng thời là sự khác biệt của bài báo này với các công trình trước, nằm ở việc đưa ra một mô thức khái quát hóa sẽ được chúng tôi mô tả chi tiết ở mục III.2.

III. MÔ HÌNH ĐỒ THỊ NGẪU NHIÊN ĐỀ XUẤT

III.1. Định nghĩa và các khái niệm cơ sở

Định nghĩa 3.1 Gọi $H(V, w, E)$ là một đồ thị cơ sở vô hướng có tập đỉnh V với các trọng số xác định bằng vec-tơ $w = \{w_u, u \in V\}$ và với tập cạnh E (mà cách cạnh được gọi là liên kết cục bộ).

Gọi τ là một họ các phân phối xác suất $\{\tau = \tau_u, u \in V\}$ mà mỗi phân phối xác định trên tập đỉnh V , tức là ứng với mỗi đỉnh $u \in V$ thì có một hàm xác suất $\tau_u(v)$ xác định cho các $v \in V$.

Với một cặp (H, τ) cho trước, các mạng ngẫu nhiên $\mathcal{NAN}(H, \tau)$ là những đồ thị phát triển từ cơ sở H theo cách sau: với mỗi đỉnh $u \in V$ ta thêm vào w_u liên kết ngẫu nhiên (cũng gọi là liên kết tầm xa), mỗi cái đi từ đỉnh u tới một đỉnh $v \in V$ nào đó được chọn với xác suất $\tau_u(v)$.

($\mathcal{NAN}$ được viết tắt theo tên tiếng Anh: Node-based Augmented Networks)

Chúng tôi cũng sử dụng một số khái niệm và ký pháp liên quan như sau. Ký pháp $G = \mathcal{NAN}(H, \tau)$ cho biết rằng đồ thị ngẫu nhiên G đã được sinh từ mô hình phát triển $\mathcal{NAN}$ theo định nghĩa trên. Để tiện ta dùng ký hiệu $v \xleftarrow{\tau} V$ để chỉ rằng đỉnh v được chọn ngẫu nhiên từ tập đỉnh V theo phân phối τ , ký hiệu $v \xleftarrow{uni} V$ khi chọn theo ngẫu nhiên đồng nhất.

Với mỗi tập con $S \subseteq V$, trọng số của S , ký hiệu w_S , là tổng các trọng số của các đỉnh trong S , $\sum_{u \in S} w_u$. Với một hằng số $k > 0$, đồ thị cơ sở H được gọi là k -heavy nếu mọi đỉnh của V đều có trọng số ít nhất là bằng k .

Ví dụ 3.1 Mạng thế giới nhỏ Kleiberg [13], dạng vô hướng, có thể được coi là được tạo bởi $\mathcal{NAN}(H, \tau)$ trong đó H là các đồ thị lưới (grid) $n \times n$ với các đỉnh có

trọng số 1; còn τ là một phân phối nghịch-bình phương: $\tau_u(v) \sim d^{-2}(u,v)$. Khoảng cách $d(u,v)$ là độ dài của đường ngắn nhất để đi từ u đến v trên đồ thị lưới H .

Bên cạnh đó, chúng tôi cũng xem xét một mô hình khác có sự mở rộng bổ sung, thuận lợi cho mô tả luật Lũy thừa; xin xem chi tiết tại phụ lục mục 2.

Với G là một đồ thị liên thông, ta ký hiệu $diam(G)$ là đường kính của đồ thị G , tức là $\max\{d(u,v) \mid u,v \text{ là đỉnh của } G\}$, trong đó $d(u,v)$ là độ dài của đường đi ngắn nhất giữa 2 đỉnh u và v .

Để thuận tiện, chúng tôi sử dụng một số ký hiệu tiệm cận định nghĩa riêng, khi xét số đỉnh n của một ĐTNN tiến ra vô cùng. Trước hết, chúng tôi viết $eNeg(n)$ thay cho $e^{-O(n)}$, tức là một lớp các hàm số coi là vô cùng nhỏ theo cấp độ hàm mũ (exponentially negligible). Cụ thể là: hàm dương $f(n)$ là thuộc $eNeg(n)$, ký hiệu $f(n)=eNeg(n)$, nếu

$$\exists c>0, c'>0, \forall n>c': 0 \leq f(n) \leq e^{-cn} \quad (1)$$

Với hàm $g(n)$ đã cho nào đó ta cũng viết $f(n)=eNeg(g(n))$, nếu:

$$\exists c>0, c'>0, \forall x>c': 0 \leq f(x) \leq e^{-cg(n)} \quad (2)$$

Ngoài ra, ta ký hiệu $eNP(n)$ cho lớp $eNeg(g(n))$ trong đó $g(n)$ là mọi đa thức của n .

Ta ký hiệu xác suất của một sự kiện E là $\Pr[E]$. Sự kiện $E(n)$ được gọi là xảy ra với xác suất rất lớn, ký hiệu VHP, nếu $\Pr[E(n)] = 1 - eNeg(n)$ với n tiến ra vô cùng. Đồng thời nói $E(n)$ xảy ra với xác suất VHP(f) nếu $\Pr[E(n)] = 1 - eNeg(f(n))$.

Một biến ngẫu nhiên $X=X(n)$ được gọi là có xu thế không thua $Y=Y(n)$, ký hiệu $X \geq_{st} Y$ nếu với mọi $a>0$, $\Pr[X < a] \leq \Pr[Y < a] + eNP(n)$, tức là khả năng X ở phía dưới mọi hằng $a>0$ nào đó là bé hơn hoặc hơn không đáng kể khả năng Y ở phía dưới chính a đó; cũng có nghĩa là khả năng X vượt a là cao hơn hoặc không thua kém đáng kể khả năng Y vượt a .

Bổ đề sau đây là một kết quả dễ dàng thu được (không chứng minh) nhờ dùng các bất đẳng thức xác suất phần đuôi (tail inequalities) như là bất đẳng thức Chernoff; xin xem thêm chi tiết tại phụ lục.

Bổ đề 3.1 Với mọi hằng số $\alpha>1$ và hằng dương $\beta<1$, tổng của n biến ngẫu nhiên đồng nhất Bec-nu-li có kỳ vọng $p=p(n)$ là $\leq \alpha np$ với VHP($\theta(n)$) và $\geq \beta np$ với VHP($\theta(np)$).

Các hằng số trong $\theta(n)$ và $\theta(np)$ của bổ đề 3.1 được xác định theo phiên bản bất đẳng thức Chernoff mà được trình bày trong phụ lục.

III.2. Phát biểu kết quả chính và các khái niệm liên quan

Một cái nhìn trực quan phổ biến trong các mô hình ĐTNN là các đồ thị cơ sở sẽ cung cấp các mối liên kết cục bộ địa phương, tức là các kết nối “ngắn”, còn các LKNN, được thêm vào sẽ cung cấp các kết nối “xa”; có như vậy sự phát triển quan hệ mới lan tỏa (xét từ 1 nút nào đó) khiến cho đường kính đồ thị trở nên nhỏ (cỡ logarit hay đa thức logarit của n). Nếu các LKNN không đi đủ “xa” hoặc là chỉ co cụm ở một khu vực nào đó thì không đảm bảo tính “phát tán” này và đồ thị sẽ không có đường kính nhỏ. Vì vậy ở đây chúng tôi tập trung đưa ra những khái niệm để mô hình hóa tính phát tán của phân phối τ trong mô hình $\mathcal{NAN}(H, \tau)$ đang xét. Sau đây là các định nghĩa hình thức cần thiết xung quanh tính “phát tán” nói trên; trước hết ta nói về tính *giãn*, một khía cạnh của phát tán.

Định nghĩa 3.3 Xét mô hình ĐTNN $\mathcal{NAN}(H, \tau)$. Với các hằng số μ và $\xi \in (0,1)$, phân phối τ được gọi là (μ, ξ) -*giãn* nếu với mọi đỉnh u , với mọi tập $C \subset V$ với không quá n^μ đỉnh, một LKNN đi từ u theo τ sẽ thoát khỏi C với xác suất $\geq \xi$ (tức là rơi vào C với xác suất $\leq 1-\xi$). Một cách hình thức, τ là (μ, ξ) -*giãn* nếu:

$$\forall u \in V, \forall C \subset V: |C| \leq n^\mu \Rightarrow \tau_u(C) \leq 1-\xi \quad (3)$$

Định nghĩa trên nêu lên một khía cạnh của tính phát tán của τ là sự không tồn tại của một tập con với chỉ n^μ đỉnh mà thống trị hoàn toàn, tức là thu hút hầu hết các LKNN từ một đỉnh cho trước nào đó (tức là có sự *giãn* ra, không tập trung hết vào một nơi). Tuy nhiên, ở chiều ngược lại, cũng cần đảm bảo tính “công bằng” ở một mức độ nào đó, sao cho không một “góc”

nào của V là gần như không có độ hấp dẫn để thu hút được LKNN.

Định nghĩa 3.4 Xét ĐTNN $\mathcal{NAN}(H, \tau)$. Với hằng số $v \in (0,1)$, phân phối τ được gọi là v -cb (cb đọc là “công bằng”) nếu với mọi đỉnh u , mọi tập con $C \subset V$ với ít nhất n^v đỉnh sẽ có xác suất thu hút LKNN từ u là $\geq \frac{1}{n}$, nghĩa là:

$$\forall u \in V, \forall C \subset V: |C| \geq n^v \Rightarrow \tau_u(C) \geq \frac{1}{n} \quad (4)$$

Với các hằng số μ và $v \in (0,1)$, phân phối τ được gọi là (μ, v) -pt (pt đọc là “phát tán”) nếu τ là v -cb và đồng thời (μ, ξ) -giãn với một hằng số $\xi > 0$ nào đó.

Bên cạnh tính phát tán thì để thu được đường kính nhỏ, đương nhiên ta cần có lượng LKNN đủ lớn, hay chính xác hơn là mật độ LKNN là đủ lớn.

Định nghĩa 3.5 Đồ thị $H(V, w, E)$ là ξ -nặng nếu mọi đỉnh u có trọng số $w_u \geq \xi$

Với các điều kiện khái quát đã đề xuất ở trên, chúng tôi xin đưa ra kết quả chính của bài báo.

Định lý 1. Xét ĐTNN $G = \mathcal{NAN}(H, \tau)$ với H là đồ thị liên thông. Giả sử tồn tại các hằng số dương λ, μ, v mà $1 > \mu > v$ sao cho H là λ -nặng còn τ là (μ, v) -pt. Thế thì G có đường kính $O(\log n)$ với xác suất $1 - o(n^{-2})$.

III.3. Ý nghĩa ứng dụng của kết quả

Định lý 1 và các khái niệm đề xuất liên quan đã đưa ra một tiếp cận để phân tích đánh giá đường kính đồ thị của một mạng TGN nào đó. Xin đưa ra một số ví dụ ứng dụng cụ thể.

Để thấy mô hình Watts-Strogatz có thể coi là một thể hiện của $\mathcal{NAN}(H, \tau)$, trong đó τ là một phân phối ngẫu nhiên đồng nhất, tức là $\tau_u(v_1) = \tau_u(v_2) \quad \forall v_1 \neq u, v_2 \neq u: v_1 \neq v_2$; từ đó dễ thấy đồ thị này thỏa mãn (μ, v) -pt $\forall \mu, v \in [0,1)$. Vì vậy theo định lý 1, mô hình Watts-Strogatz với tham số hằng $\beta > 0$ có đường kính đồ thị là $O(\log n)$.

Ở trên ta đã đề cập mô hình Kleinberg cơ bản ($\alpha=2$) là một thể hiện của $\mathcal{NAN}(H, \tau)$ với τ tuân theo luật tỷ lệ nghịch bình phương. Khảo sát mô hình này,

dễ thấy rằng đồ thị này là (μ, ξ) -giãn với các hằng số bất kỳ $\mu, \xi \in [0,1)$ nhưng thỏa mãn $\mu + \xi < 1$. Cũng dễ thấy $\tau_u(v) = O(\frac{1}{\log n} \frac{1}{d^2(u,v)})$; nên $\forall u \in V, \exists c > 0, \forall C \subset V: |C| \geq c \log n \Rightarrow \tau_u(C) \geq n^{-1}$. Do đó đồ thị này cũng thỏa mãn (μ, v) -pt $\forall \mu, v \in (0,1)$. Vì vậy theo định lý 1, mô hình Kleinberg với $\alpha=2$ có đường kính đồ thị là $O(\log n)$. Ta có cũng có thể chứng minh một kết quả tương tự với một mô hình TGN khác của Kleinberg, trong đó đồ thị cơ sở là một cây thay vì một lưới [10].

Ta hãy thử đánh giá khái quát đường kính đồ thị liên kết vật lý Internet (theo cấp router hoặc theo cấp AS, tức là hệ thống tự trị) theo tiếp cận sử dụng mô hình $\mathcal{NAN}(H, \tau)$. Faloutsos et al. quan sát thấy rằng với bán kính R đủ nhỏ, mỗi lân cận bán kính R (khoảng cách địa lý) sẽ chứa số nút mạng tỉ lệ với R^α ; mặc dù α biến đổi theo vùng nhưng $\alpha \approx 1$ [6]. Mặt khác ở khoảng cách tương đối xa, xu hướng kết nối được quan sát là có xác suất tỷ lệ nghịch với lũy thừa bậc β của khoảng cách địa lý, trong đó β biến đổi nhưng nằm trong khoảng $(1,2)$ [23,6]. Vậy ta có thể mô phỏng đồ thị Internet bằng $\mathcal{NAN}(H, \tau)$ với H là một lưới xấp xỉ một chiều, còn τ thỏa $\tau_u(v) \sim d^{-\beta}(u,v)$. Tính toán cho thấy đồ thị này cũng thỏa mãn (μ, v) -pt nếu có $1 > \mu > v > \beta - 1$ (thỏa mãn được do $\beta \in (1,2)$)¹. Vậy đường kính đồ thị của mạng này cũng là $O(\log n)$.

Trên đây chúng tôi đã nêu khảo sát ngắn gọn trên một số mô hình TGN khác nhau, với tiếp cận ứng dụng mô hình $\mathcal{NAN}(H, \tau)$. Còn có nhiều hướng tiếp cận khác, nhưng do điều kiện hạn chế về dung lượng bài báo chúng tôi không nêu ra ở đây, như khảo sát mô hình Kleinberg tổng quát (α bất kỳ) hay đồ thị mô phỏng luật lũy thừa (chúng tôi sẽ công bố các kết quả này trên một báo cáo nghiên cứu trong tương lai).

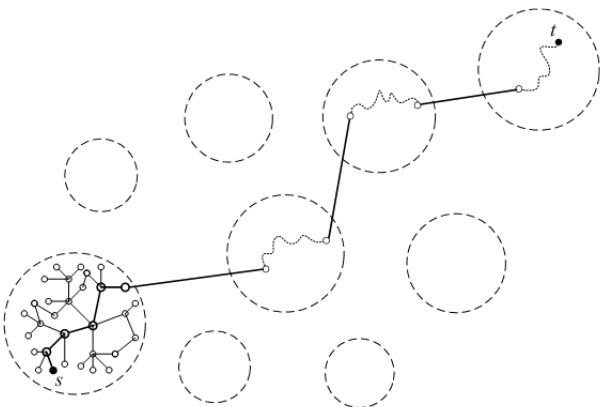
Có thể nói, đóng góp chính của bài báo này là cung cấp công cụ mô hình lý thuyết để phân tích các đồ thị TGN được giới nghiên cứu quan tâm trước nay. Tuy

¹ Bên cạnh đó, chú ý rằng $\forall \mu \in (0,1) \exists \xi > 0$ để đồ thị là (μ, ξ) -giãn nhưng $\mu > \xi$

nhiên kết quả mang tính lý thuyết cơ bản này của chúng tôi cũng mang lại ý nghĩa thực tiễn vì thông thường sự hiểu biết sâu sắc về các mô hình TGN có thể đem lại một tư tưởng thiết kế mới; ví dụ như một số thiết kế mới của mạng đồng đẳng (P2P) có ảnh hưởng của mô hình TGN [15,16]. Thật vậy, như đã điểm qua ở mục II, một công trình khác của chúng tôi về thiết kế mạng liên kết (ứng dụng cho siêu máy tính hoặc trung tâm dữ liệu hiện đại) [20] cũng đã ứng dụng ý tưởng của mô hình TGN; trong đó chúng tôi thiết kế các liên kết xa (được gọi là shortcut) theo cách mô phỏng các LKNN trong mô hình TGN tổng quát ở đây (có tính “giãn” và “phát tán”) nhằm đạt được đường kính đồ thị logarit. Để thấy rõ ý nghĩa thực tiễn này xin tham khảo chi tiết tại [20].

III.4. Ý tưởng và cấu trúc chứng minh định lý 1

Để chứng minh định lý 1, trước hết chúng tôi sẽ khảo sát trường hợp đơn giản cơ bản, khi phân phối τ là phân phối đồng nhất (mỗi đỉnh v đều có cơ hội như nhau để thu hút LKNN từ mỗi đỉnh u). Đây là bài toán con có tầm quan trọng và tính độc lập nhất định do có sự liên hệ chặt chẽ với các đồ thị ngẫu nhiên truyền thống (Erdos-Renyi). Sau đó chúng tôi sẽ sử dụng kết quả khảo sát bài toán con này (được phát biểu thành định lý 2 như ở mục IV) làm cơ sở để chứng minh định lý 1.



Hình 1. Minh họa sự thu giảm về siêu đồ thị mà tồn tại đường dẫn độ dài $O(1)$ giữa mọi cặp đỉnh.

Ý tưởng cơ bản là, sử dụng tính (μ, ν) -pt, ta có thể phân rã V thành các tập đỉnh con rời nhau, mỗi tập có n^μ đỉnh, sao cho đường kính của đồ thị con nội tại tạo thành từ mỗi tập con đỉnh là $O(\log n)$. Vì vậy, không mất tính tổng quát, ta có thể thu giảm bài toán thành xét một siêu đồ thị mà mỗi siêu đỉnh đại diện cho một tập con đỉnh nói trên (có $n^{1-\mu}$ siêu đỉnh) như minh họa trong Hình 1. Tiếp đó sử dụng Định lý 2 để chứng minh siêu đồ thị này có đường kính là $O(1)$, tức là suy ra đpcm (điều phải chứng minh)!

IV. KHẢO SÁT TRƯỜNG HỢP τ LÀ PHÂN PHỐI ĐỒNG NHẤT

Trong mục này, ta sẽ khảo sát đường kính của DTNN theo mô hình $\mathcal{NAN}(H, \tau)$ trong đó τ là phân phối đồng nhất; có nghĩa là khi tạo một LKNN cho đỉnh u thì theo τ , ta chọn đích v theo luật chọn ngẫu nhiên đồng nhất (uniform random) từ tập đỉnh $V \setminus \{u\}$. Mô hình trường hợp $\mathcal{NAN}(H, \tau = \text{uniform})$ này có thể được định nghĩa lại như sau.

Mô hình đồ thị ngẫu nhiên J - Định nghĩa 4.1.

Một đồ thị ngẫu nhiên $J(n, Z)$ được sinh theo mô hình J , ký hiệu $J = J(n, Z)$, là một đồ thị có n đỉnh và các cạnh được tạo ra như sau: từ mỗi đỉnh u của đồ thị, độc lập sinh ra Z cạnh nối tới các đỉnh $v \xleftarrow{\text{uni}} V$.

Rõ ràng $J(n, Z)$ chính là $\mathcal{NAN}(H, \tau = \text{uniform})$ với H là đồ thị gốc chỉ có n đỉnh (không cạnh) đều với trọng số Z ; dễ thấy, kỳ vọng của bậc mỗi đỉnh là $2Z$.

Mô hình này là khá gần gũi với mô hình đồ thị ngẫu nhiên truyền thống Erdős-Rényi, ký hiệu $G(n, p)$ trong đó một cạnh có thể được tạo ra ngẫu nhiên với xác suất p cho mỗi cặp đỉnh bất kỳ (trong số n đỉnh cho trước). Nếu chọn $p = Z/2n$, các đồ thị sinh từ 2 mô hình này sẽ là những phiên bản khá giống nhau, tuy nhiên vẫn có chút sự khác biệt: có sự phụ thuộc phần nào trong sự sinh ra giữa các cạnh của $J(n, Z)$. Vì vậy, bậc của đỉnh trong J sẽ ít nhất là Z , còn trong G bậc đỉnh có thể rất nhỏ, thậm chí là 0 dù với xác suất rất nhỏ.

Định lý sau đây là kết quả chính của mục này.

Định lý 2. Xét một đồ thị ngẫu nhiên $J = J(n, Z)$ trong đó $Z = Z(n)$ thỏa mãn rằng có một số nguyên dương d để $Z^d/n \rightarrow 0$ và $Z^{d+1}/(n \log n) \rightarrow \infty$, khi $n \rightarrow \infty$. Thế thì $\text{diam}(J) \leq d+4$ với VHP(Z).

Để chứng minh định lý 2, ta sẽ thông qua một kết quả gần tương tự đối với đường kính đồ thị $G(n, p)$. Đây là một kết quả khá cổ điển đối với mô hình truyền thống này (có thể tham khảo trong [2]).

Bổ đề 4.2 Xét một đồ thị ngẫu nhiên theo mô hình Erdős–Rényi $G = G(n, Z)$ thỏa mãn rằng có một hằng số nguyên dương d để $(np)^d/n \rightarrow 0$ và $(np)^{d+1}/(n \log n) \rightarrow \infty$. Thế thì $\text{diam}(G) = d+4$ với VHP(np).

Ví dụ để điều kiện trên thỏa mãn có thể chọn sao cho $p = n^{\frac{1}{d}-1} \log n$ (tức là $np = n^{\frac{1}{d}} \log n$). Thậm chí Bollobas còn cung cấp một kết quả mạnh hơn: $\text{diam}(G)$ sẽ tập trung quanh hai giá trị d và $d+1$ với d là hằng số hay một hàm tăng chậm theo n .

Việc sử dụng kỹ thuật kỹ thuật cổ điển (như dùng trong [1] để chứng minh bổ đề 4.2) để chứng minh định lý 2 có thể thực hiện được nhưng khá khó và phức tạp. Vì vậy, ở đây chúng tôi phát triển một phương pháp riêng, đó là dựa vào bổ đề 4.2, thông qua ý tưởng so sánh đường kính của các đồ thị theo các mô hình gần tương tự. Như phân tích ở trên, ta đã thấy mô hình $J(n, Z)$ và $G(n, Z)$ gần tương tự. Mỗi liên hệ đường kính đồ thị của 2 dạng mô hình này được khảo sát và xây dựng thông qua việc so sánh đường kính đồ thị của 2 mô hình trung gian, đem lại sự chuyển biến khác biệt nhỏ hơn nữa giữa mô hình J và G . Sau đây chúng tôi giới thiệu 2 mô hình trung gian $\mathcal{H}$ và $\mathcal{I}$. Lưu ý rằng cả 4 mô hình đồ thị mà ta xét ở đây (G , $\mathcal{H}$, $\mathcal{I}$ và J) đều là vô hướng.

Mô hình ĐTNN $\mathcal{I}$ - Định nghĩa 4.2. Một đồ thị ngẫu nhiên $\mathcal{I}(n, m)$ được sinh theo mô hình $\mathcal{I}$, ký hiệu $\mathcal{I} = \mathcal{I}(n, m)$, là một đồ thị có n đỉnh và với mỗi đỉnh $u \in V(\mathcal{I})$ ta chọn ngẫu nhiên một tập hợp $m < n$ đỉnh

phân biệt $v_1, v_2, \dots, v_m$ từ $V \setminus \{u\}$ rồi tạo m cạnh nối (u, v_i).

So sánh $\mathcal{I}$ với J . Sự khác biệt là trong $\mathcal{I}$ các cạnh sinh ngẫu nhiên đều là cạnh đơn (nghĩa là không thể có nhiều hơn 2 cạnh cùng nối 2 đỉnh nào đó).

Mô hình đồ thị ngẫu nhiên $\mathcal{H}$ - Định nghĩa 4.3. Một đồ thị ngẫu nhiên $\mathcal{H}(n, p)$ được sinh theo mô hình $\mathcal{H}$, ký hiệu $\mathcal{H} = \mathcal{H}(n, p)$, là một đồ thị có n đỉnh và từ mỗi đỉnh u , ta tạo ngẫu nhiên một cạnh nối với mỗi đỉnh v khác với xác suất p .

So sánh giữa $\mathcal{H}$ với G . Sự khác biệt tương đối nhỏ là trong G với mỗi cặp đỉnh (u, v) ta tạo ngẫu nhiên 1 cạnh với xác suất p , còn trong $\mathcal{H}$, xác suất tồn tại một cạnh giữa một cặp đỉnh (u, v) là $2p - p^2 > p$.

Chứng minh định lý 2. Ý tưởng cơ bản của chứng minh này là như sau. Với n và Z cho trước, ta chọn $p = \frac{Z}{2n}$ và xét việc sinh các đồ thị ngẫu nhiên $G = G(n, p = \frac{Z}{2n})$, $\mathcal{H} = \mathcal{H}(n, p)$ và $\mathcal{I} = \mathcal{I}(n, m = \frac{3Z}{4})$. Ở đây ta chọn p và m , để sao cho mật độ các cạnh sinh ra là dày dần lên trong dãy $G, \mathcal{H}, \mathcal{I}$ và J . Do đó ta có thể chứng minh $\text{diam}(G) \geq_{xt} \text{diam}(\mathcal{H}) \geq_{xt} \text{diam}(\mathcal{I}) \geq_{xt} \text{diam}(J)$. Do đó suy ra điều phải chứng minh nhờ sử dụng bổ đề 4.2.

Để có thể chứng minh được chuỗi bất đẳng thức xác suất về đường kính trên, chúng tôi đề xuất một kỹ thuật so sánh dựa trên các quan sát riêng như sau. Để chứng minh $\text{diam}(X) \geq_{xt} \text{diam}(Y)$ với X và Y là 2 trong các đồ thị dạng trên, chúng tôi đưa ra một khái niệm quan hệ mới giữa 2 đồ thị ngẫu nhiên X và Y , ký hiệu $X \succ Y$, như sau. Ta nói $X \succ Y$ nếu tồn tại một quá trình ngẫu nhiên có 2 giai đoạn: giai đoạn 1 với xuất phát là một đồ thị rỗng có n đỉnh ta thêm dần vào một số cạnh ngẫu nhiên, và kết thúc giai đoạn này ta thu được một thể hiện X ; giai đoạn 2 ta cũng tiếp tục thêm vào một số cạnh ngẫu nhiên khác song song với việc tĩa bỏ một số cạnh kiểu suy biến vòng (nguồn và đích là một), và cuối cùng ta thu được một thể hiện Y .

Nói cách khác, quá trình ngẫu nhiên nói trên là sinh dần các LKNN và có thể được thể hiện qua một mã

chương trình M gồm 2 phần kế tiếp M_1 và M_2 , ta viết $M=M_1||M_2$, sao cho M sinh được Y còn M_1 sinh được X . Những cạnh suy biến vòng và song song (thêm vào nhưng trùng lặp với LKNN đã có) ta coi là không đáng quan tâm.

Mở rộng khái niệm, ta nói $X \succ_e Y$ nếu tồn tại một quá trình ngẫu nhiên bổ sung dần các LKNN không suy biến vòng, gồm 2 giai đoạn, thể hiện dạng mã chương trình $M=M_1||M_2$, mà với xác suất $1-eNeg(n)$ thì M_1 sinh được X còn M cũng sinh được Y . Tương tự, ta cũng nói $X \succ_{e(f(n))} Y$ với xác suất nói trên là $1-eNeg(f(n))$.

Từ đây ta có 2 quan sát khá hiển nhiên.

Quan sát RG-A. Nếu $X \succ Y$ hay $X \succ_e Y$ thì $diam(X) \geq_{xt} diam(Y)$.

Quan sát này là hiển nhiên do thuộc tính bất biến này của đường kính đồ thị: Nếu X là đồ thị con của đồ thị Y thì $diam(X) \geq diam(Y)$.

Quan sát RG-B. Giả sử $H=(V,W,E)$ là một đồ thị cơ sở có vector trọng số W . Xét 2 phân phối LKNN τ và τ' trên H . Giả sử τ có xu thế kém τ' nghĩa là với mọi cặp đỉnh phân biệt $u, v \in V$ thì $\tau_u(v) \leq \tau'_u(v)$. Thế thì với hai ĐTNN sinh từ cơ sở H là $X = \mathcal{NAN}(H, \tau)$ và $Y = \mathcal{NAN}(H, \tau')$ ta có $X \succ Y$.

Để minh chứng cho quan sát này, ta có thể xây dựng một quá trình ngẫu nhiên với mã $M=M_1||M_2$ như sau. Với mỗi đỉnh u , M_1 sẽ thực hiện sinh thêm W_u liên kết ngẫu nhiên theo phân phối τ , tức là tạo ra một thể hiện X . Đoạn mã M_2 thực hiện việc sau: với mỗi cạnh vòng (đỉnh nguồn và đích là một) sinh ra bởi M_1 ta xem xét thay thế nó bằng một LKNN mới đi từ u tới một đỉnh v khác với xác suất $\frac{\tau'_u(v) - \tau_u(v)}{\tau_u(u)}$. Dễ thấy M là một quá trình sinh dần các LKNN và cuối cùng sinh ra một thể hiện Y . Vậy suy ra $X \succ Y$.

Theo quan sát **RG-A**, rõ ràng chúng ta sẽ chứng minh được định lý 2 nếu chứng minh được rằng $G \succ H \succ_e I \succ_e J$ với $G = \mathcal{G}(n, p = \frac{Z}{2n})$, $H = \mathcal{H}(n, p)$ và $I =$

$\mathcal{I}(n, m = \frac{3Z}{4})$. như đã biết. Sau đây là các bước chứng minh cụ thể.

Chứng minh $G \succ H$ là khá hiển nhiên thông qua quan sát **RG-B** và theo sự so sánh giữa $\mathcal{G}$ và $\mathcal{H}$ ta đã nói ở trên (ngay sau phần định nghĩa $\mathcal{H}$), ta có thể coi H là sinh bởi $\mathcal{G}(n, 2p-p^2)$ và bổ sung thêm một số cạnh song song không đáng quan tâm.

Giờ ta sẽ chứng tỏ $H \succ_e I$. Ta xây dựng một quá trình ngẫu nhiên $M=M_1||M_2$ như sau. Trước hết sinh H , M_1 cho mỗi đỉnh u sinh một LKNN nối với $v \neq u$ với xác suất p . Ta gọi R_u là biến ngẫu nhiên đếm số liên kết được sinh bởi mỗi đỉnh u như vậy. Sau đó M_2 thực hiện kiểm tra với mỗi node u nếu $R_u \leq m = \frac{3Z}{4}$ thì sinh thêm $m_u = (m - \frac{3Z}{4})$ liên kết mới đi từ u : chọn ngẫu nhiên một tập m_u các đỉnh từ V' , tập các đỉnh v mà đã được chọn ứng với u trong M_1 ở trên, rồi nối u đến tập m_u đỉnh đó.

Với mỗi đỉnh u thì R_u chính là tổng của $n-1$ biến ngẫu nhiên Bec-nu-li đồng nhất với xác suất p , nên theo bổ đề 3.1, ta có $R_u \leq 1,5np = m = \frac{3Z}{4}$ với VHP. Hãy gọi E là sự kiện $R_u \leq m$ với mọi đỉnh u , tức là $M=M_1||M_2$ sinh được một thể hiện Y . Rõ ràng $\Pr[\bar{E}] \leq n * \Pr[R_u > m] = n * eNeg(n)$ và vẫn là $eNeg(n)$. Vì vậy với VHP ta có $M=M_1||M_2$ sinh ra một thể hiện Y . Tức là, $H \succ_e I$.

Ta chỉ còn cần chứng minh $I \succ_{eN} J$. Ta có thể sử dụng kỹ thuật tương tự như trên. Tuy nhiên để giản lược trình bày, về bản chất có thể thấy rằng ta chỉ cần đánh giá xác suất các sự kiện E_u như sau: trong J , khi mỗi đỉnh u sinh Z liên kết ngẫu nhiên (độc lập) đi từ u , ta gọi E_u là sự kiện mà u tạo ra được ít nhất $m = \frac{3Z}{4}$ liên kết phân biệt. Ta cần chứng tỏ rằng E , sự kiện mà E_u xảy ra với mọi u , là xảy ra với VHP. Xét việc u sinh LKNN thứ k , $1 \leq k \leq Z$. Xác suất để LKNN này là “mới”, tức là khác biệt với $k-1$ lần trước là $\geq 1 - (k-1)/(n-1) \geq 1 - (Z-1)/(n-1)$. Dễ thấy với n đủ lớn (vì $Z/n \rightarrow 0$) ta có xác suất trên $\geq 0,9$. Do đó R_u , số LKNN tạo bởi u , có thể được chặn dưới bằng tổng của Z biến

ngẫu nhiên Bec-nu-li độc lập đồng nhất với xác suất thành công 0,9. Theo bổ đề 3.1, ta có $\Pr[R_u \geq m = \frac{3Z}{4}]$ là VHP(Z). Tức là $1 - \Pr[E_u] = \text{eNeg}(Z)$. Vì E không xảy ra nếu như chỉ cần có một đỉnh u mà E_u không xảy ra, nên $1 - \Pr[E] \leq n \cdot (1 - \Pr[E_u]) = n \cdot \text{eNeg}(Z) = \text{eNeg}(Z)$, vì $n/Z^d \rightarrow 0$. Tức là E xảy ra với VHP(Z).

Vậy, $G \succ H \succ I \succ J$, và theo bổ đề 4.2 ta có $\text{diam}(J) \leq d+4$ với VHP(Z), tức là đpcm.

V. ĐỊNH LÝ ĐIỀU KIỆN ĐỦ XÁC LẬP ĐƯỜNG KÍNH ĐỒ THỊ NHỎ

Trong mục này chúng tôi xây dựng chứng minh cho định lý 1. Để tiện theo dõi xin nhắc lại phát biểu như sau.

Định lý 1. Xét ĐTNN $G = \mathcal{NAN}(H, \tau)$ với H là đồ thị liên thông. Giả sử tồn tại các hằng số dương λ, μ, ν mà $1 > \mu > \nu$ sao cho H là λ -nặng còn τ là (μ, ν) -pt. Thế thì G có đường kính $O(\log n)$ với VHP.

Để chứng minh, trước hết chúng tôi đưa ra một số khái niệm và bổ đề hỗ trợ như sau.

V.1. Các khái niệm và kết quả hỗ trợ

Xét một đồ thị $G(V, E)$ liên thông. Giả sử U là một tập con đỉnh của V , ta ký hiệu G_U là một đồ thị con của G mà thu được bằng cách lấy từ G chỉ các đỉnh thuộc U và các cạnh giữa chúng. Ta gọi một lân cận của đỉnh u với bán kính k trên G , ký hiệu là $G_u(k)$ hay G_u^k , là tập tất cả các đỉnh v mà tồn tại đường dẫn từ u tới v trên G với độ dài là $\leq k$.

Định nghĩa 5.1 Xét đồ thị $G(V, E)$ và một phân hoạch tập đỉnh V thành các tập con rời nhau: $\mathcal{V} = \{V_1, V_2, \dots, V_k\}$, $V = \bigcup_{i=1, k} V_i$. Ta gọi $G^\mathcal{V}$ là một siêu đồ thị thu được từ G bằng cách thu giảm G_{V_i} thành một siêu đỉnh; chính xác hơn, $G^\mathcal{V}$ có k đỉnh là $G_1^\mathcal{V}, G_2^\mathcal{V}, \dots, G_k^\mathcal{V}$, sao cho với mọi $1 \leq i \neq j \leq k$, giữa $G_i^\mathcal{V}$ và $G_j^\mathcal{V}$ tồn tại một cạnh nếu tồn tại một cạnh $(u, v) \in E$ mà $u \in G_i^\mathcal{V}$ và $v \in G_j^\mathcal{V}$.

Bổ đề 5.1 Giả sử với đồ thị liên thông $G(V, E)$ có tồn tại một phân hoạch tập đỉnh V thành các tập con rời nhau: $\mathcal{V} = \{V_1, V_2, \dots, V_k\}$, $V = \bigcup_{i=1, k} V_i$ sao cho V_i

đều nằm trong các đồ thị con G_i liên thông nào đó và $\forall i, \text{diam}(G_i) \leq p$, với $p > 0$ là một hằng số nguyên nào đó. Thế thì, $\text{diam}(G) \leq p * q$ với $q = \text{diam}(G^\mathcal{V})$.

Để đảm bảo tính liên tục, chúng tôi để chứng minh của bổ đề 5.1 trong phụ lục. Trong hai định nghĩa tiếp theo đây, ta đều lấy cơ sở là xét một đồ thị liên thông $G(V, E)$ và khái niệm khoảng cách $d(u, v)$ là độ dài của đường dẫn ngắn nhất giữa 2 đỉnh cho trước u và v nào đó.

Định nghĩa 5.2 Với số tự nhiên δ bất kỳ ta gọi δ -lưới là một tập đỉnh con $U = \{u_1, u_2, \dots, u_t\}$ sao cho mọi khoảng cách giữa các đỉnh trong U là $> \delta$ đồng thời $V = \bigcup_{i=1, t} G_{u_i}(\delta)$.

Một δ -lưới có thể xây dựng theo nguyên tắc tham lam đơn giản như sau: Lấy bất kỳ $u_1 \in V$ và sau đó với $k=1, 2, 3, \dots$ lấy bất kỳ u_{k+1} từ $V - \bigcup_{i=1, k} G_{u_i}(\delta)$ chừng nào mà tập này còn khác rỗng. Để kiểm tra thấy cách xây dựng này thỏa mãn định nghĩa 5.2.

Định nghĩa 5.3 Xét một 2δ -lưới gọi là $U = \{u_1, u_2, \dots, u_t\}$. Ta xét phân hoạch $\mathcal{V} = \{V_1, V_2, \dots, V_t\}$, $V = \bigcup_{i=1, t} V_i$ xây dựng theo cách sau: với mỗi đỉnh bất kỳ $v \in V$, tìm $i \in 1..t$ sao cho khoảng cách giữa $d(u_i, v)$ nhỏ nhất và đưa v vào tập V_i . Ta gọi phân hoạch này là U -dựa.

Với $\mathcal{V}$, một phân hoạch U -dựa như trên, vì khoảng cách giữa các đỉnh của U là $\geq 2\delta$ suy ra tất cả các lân cận $G_{u_i}(\delta)$ là rời nhau. Để thấy rằng $G_{u_i}(\delta) \subseteq V_i$. Vì vậy từ bổ đề 5.1, ta suy ra kết quả hiển nhiên sau.

Bổ đề 5.2 Xét ĐTNN $G = \mathcal{NAN}(H, \tau)$ với H là đồ thị liên thông, U là 2δ -lưới và $\mathcal{V}$ là một phân hoạch U -dựa (theo các định nghĩa 5.2, 5.3). Ta có:

- i) $\forall j=1, t: G_{u_j}(\delta) \subseteq V_j$;
- ii) $\text{diam}(G) \leq 2\delta * \text{diam}(G^\mathcal{V})$.

Bổ đề 5.2 đã cho thấy ý tưởng chính của việc chứng minh định lý 1: ta cần xây dựng các lưới U thích hợp và phân hoạch $\mathcal{V}$ (với các tập đỉnh con đủ lớn) để chuyển về bài toán tìm đường kính của $G^\mathcal{V}$ (mà ở đây ta có thể áp dụng định lý 2). Bổ đề tiếp theo sau

đây sẽ chỉ ra cách xây dựng tập đỉnh con đủ lớn như là các lân cận với bán kính chỉ là $O(\log n)$.

Bổ đề 5.3 Xét ĐTNN $G = \mathcal{NAN}(H, \tau)$ với H là đồ thị liên thông và λ -nặng, và τ là (μ, ξ) -giãn với μ, ξ bất kỳ thuộc $(0, 1)$ sao cho $\lambda * \xi > 1$. Vậy tồn tại một hằng số $C > 0$ để cho mọi lân cận $G_u(k)$ của tâm $u \in V$ và bán kính $k = C \log n$ đều có kích thước là $\geq n^\mu$ với xác suất $1 - o(n^{-2})$, tức là:

$$\exists C > 0: \Pr[\forall u \in V, k = C \log n: |G_u(k)| \geq n^\mu] = 1 - o(n^{-2}) \quad (5)$$

(Thực ra ta có thể bỏ bớt điều kiện $\lambda * \xi > 1$, nhưng chứng minh sẽ dài hơn và không thực sự cần thiết.)

Chứng minh bổ đề 5.3. Với mỗi đỉnh $u \in V$ ta gọi $R_\tau(u)$ là tập gồm các đỉnh v mà (u, v) là một LKNN trên G ‘được sinh’ tại u theo phân phối τ . Tất nhiên $|R_\tau(u)|$ là phụ thuộc và không vượt quá trọng số w_u . Với một đỉnh u cho trước, ta xét chuỗi các lân cận $\{S_i\}$ xây dựng theo cách sau:

- $S_0 = \emptyset$
- $S_1 = H_u(c \log n)$, tức là một lân cận của u bán kính $c \log n$ trong đồ thị cơ sở H ; $c > 0$ là một hằng số đủ lớn mà ta sẽ chọn sau.
- Với $i = 1, 2, 3, \dots$, $S_{i+1} = S_i \cup T_i$, trong đó $T_i = \bigcup_{v \in V_i} R_\tau(v)$

Để cho tiện, ta có thể ‘tưởng tượng’ là mỗi đỉnh v có w_v hạt giống và các LKNN được sinh dần như sau: từ S_i sinh ra S_{i+1} thì dùng các hạt giống mới chưa dùng, tức là thuộc vào $S_i - S_{i-1}$, để tạo ra các LKNN mới cho S_{i+1} .

Bây giờ ta sẽ chứng minh là: $|S_i| < n^\mu / (1 + \alpha) \rightarrow |S_{i+1}| / |S_i| > (1 + \alpha)$ với xác suất $1 - o(n^{-4})$, α là hằng số tùy ý $\in (1, \lambda \xi)$.

Ta hãy mô tả chi tiết quá trình sinh dần S_{i+1} từ S_i như sau đây:

- Đặt $S = S_i - S_{i-1}$, $T = \emptyset$
- Với mỗi đỉnh $v \in S$, ta lần lượt sinh $v_i = R_\tau(v)$ với $i = 1..w_v$, nếu $v_i \notin S_i \cup T$ thì ta thêm nó vào T : $T = T \cup \{v_i\}$
- Sau khi sử dụng hết ‘hạt giống’ trong S , ta thu được $S_{i+1} = S_i \cup T$.

Ta đánh giá $\Pr[|T|/|S| > \alpha]$. Nếu như $|S_i \cup T| < n^\mu$ thì do τ là (μ, ξ) -giãn ta thấy xác suất để mỗi hạt giống trong S có thể sinh ra một LKNN đến đỉnh nào đó (v_i) không nằm trong $S_i \cup T$ là ít nhất ξ . Tổng số hạt giống của các đỉnh trong S là $w_S = \sum_{v \in S} w_v$; vì vậy ta có thể chặn dưới $|T|$ bằng tổng của w_S các biến Becnuli độc lập đồng nhất có xác suất thành công (=1) là ξ . Mặt khác $w_S \geq \lambda |S|$; do đó nếu ta chọn $\alpha: 1 < \alpha < \lambda \xi$ thì theo bổ đề 3.1, ta có $|T| \geq \alpha |S|$ với VHP($c' |S| \xi$) với hằng số c' nào đó. Vậy chọn $c > 4/(c' \xi)$ thì ta có: $|S| \geq c \log n \rightarrow \Pr[|T|/|S| > \alpha] = 1 - o(n^{-4})$.

Do đó dễ thấy, chừng nào $|S_j| < n^\mu$, $j > 1$, ta luôn có với xác suất $1 - o(n^{-3})$ thì:

$$|S_j| \geq |S_1| (1 + \alpha + \alpha^2 + \dots + \alpha^{j-1}) = |S_1| (\alpha^j - 1) / (\alpha - 1).$$

Do đó với $j = (\mu / \log \alpha) \log n + 1$, ta sẽ có $\Pr[|S_j| \geq n^\mu] = 1 - o(n^{-3})$.

Dễ thấy các đỉnh trong S_j có thể đến được từ u bằng đường dẫn với tối đa là $c \log n + j$ cạnh; chú ý rằng các đỉnh trong S_1 có thể kết nối được từ u với tối đa $c \log n$ cạnh trên đồ thị cơ sở H . Tức là nếu ta chọn $C = c + (\mu / \log \alpha) + 1$ thì $\Pr[|G_u(C \log n)| \geq n^\mu] = 1 - o(n^{-3})$.

Suy ra với $C = c + (\mu / \log \alpha) + 1$ ta có:

$\Pr[\forall u \in V, k = C \log n: |G_u(k)| \geq n^\mu] = 1 - o(n^{-2})$, đó là đpcm!

Bây giờ ta có đủ các kết quả chuẩn bị cần thiết để chứng minh trực tiếp định lý 1.

V.2. Chứng Minh Định lý 1

Trước hết, ý tưởng cụ thể của chứng minh là như sau:

- A) Đưa ra một phân hoạch tập đỉnh V thành các tập con rời nhau: $\mathcal{V} = \{V_1, V_2, \dots, V_m\}$, $V = \bigcup_{i=1, m} V_i$ sao cho các tập con V_i có kích thước xấp xỉ và $\geq n^\mu / 2$, đồng thời các $V_i \subseteq V'_i$ nào đó mà $G_{V'_i}$ có đường kính là $O(\log n)$. Ta sẽ sử dụng các bổ đề 5.2 và 5.3 ở đây.
- B) Chứng minh siêu đồ thị G^* có đường kính $O(1)$. Từ đây kết hợp với kết quả của A) và theo bổ đề 5.1, đường kính của G cũng là $O(\log n)$.

Để thực hiện B), ta sẽ vận dụng lại kỹ thuật so sánh các mô hình ĐTNN trong mục V. Cụ thể là ta sẽ chứng minh $J \succ_{e(n^c)} K \succ G^z$, với hằng số $c > 0$ nào đó, trong đó $J = J(m, Z)$ sẽ được mô tả chi tiết ở phần sau, còn $K = K(m, z, p)$ theo một mô hình ĐTNN mở rộng của mô hình $J(n, Z)$ như định nghĩa dưới đây. Nhờ đó ta có $\text{diam}(G^z) = O(\text{diam}(J)) = O(1)$.

Mô hình đồ thị ngẫu nhiên K . Một ĐTNN $K(n, m)$ được sinh theo mô hình K , ký hiệu $K = K(m, z, p)$, là một đồ thị có m đỉnh và với mỗi đỉnh $u \in V(K)$ ta độc lập tạo ra z LKNN, mỗi trong chúng sẽ chọn đến một đỉnh $v \neq u$ với xác suất p và quay lại về u với xác suất $1 - (m-1)p$.

Sau đây là chứng minh cụ thể của định lý 1.

Trước hết, chú ý rằng các giả thiết của định nghĩa 5.1 chưa cho phép ta dùng ngay bổ đề 5.3, nơi mà ta cần thêm điều kiện $\lambda^* \xi > 1$. Tuy nhiên ta có thể bổ sung giả thiết này vào mà không làm yếu định lý 1 vì lý do sau. Bổ đề 5.2 chỉ ra rằng, ta có $\text{diam}(G) = O(\text{diam}(G^z))$ trong đó: U là một 2δ-lưới với $\delta > 0$ là một hằng số nào đó và $\mathcal{Z}$ là một phân hoạch U -dựa (theo các định nghĩa 5.2, 5.3); mặt khác ta có thể chọn hằng số δ đủ lớn để các đỉnh của G^z là đủ nặng và thỏa mãn điều kiện $\lambda^* \xi > 1$. Bởi vì mỗi siêu đỉnh của G^z sẽ gom các “hạt giống” (sinh LKNN) từ các đỉnh nội thuộc từ đồ thị G , tức là đồ thị G^z tối thiểu cũng là δλ-nặng. Tức là ta chỉ còn cần chứng minh định lý 1 với đồ thị G^z đã thỏa mãn $\lambda^* \xi > 1$. Do đó, để cho tiện và không mất tính tổng quát ta đưa luôn giả thiết $\lambda^* \xi > 1$ cho đồ thị G ban đầu. Tức là ta có thể dùng bổ đề 5.3 ngay với đồ thị G . Nói rộng hơn ta luôn có thể giả thiết hằng số λ là đủ lớn tùy ý.

Theo bổ đề 5.3, tồn tại một hằng số $C > 0$ để cho với xác suất $1 - o(n^{-2})$, mọi lân cận $G_u(k)$ của tâm $u \in V$ và bán kính $k = C \log n$ đều có kích thước là $\geq n^\mu$. Gọi U là một $2k$ -lưới và $\mathcal{Z}' = \{V'_1, V'_2, \dots, V'_m\}$ là một phân hoạch U -dựa (theo các định nghĩa 5.2, 5.3). Với mỗi tập đỉnh con V'_i ta lại chia nhỏ hơn thành các tập đỉnh nhỏ hơn sao cho kích thước của chúng đều nằm trong

khoảng $(n^\mu/2, n^\mu)$. Nhờ đó ta thu được một phân hoạch $\mathcal{Z} = \{V_1, V_2, \dots, V_m\}$ mà các tập đỉnh con đều có kích thước nằm trong khoảng $(n^\mu/2, n^\mu)$. Các tập đỉnh con này đều nằm trong các lân cận có bán kính $C \log n$ của G . Tức là $\mathcal{Z}$ thỏa mãn:

- Phủ kín V , nghĩa là: $V = \bigcup_{i=1, m} V_i$
- Các V_i có kích thước $\in (n^\mu/2, n^\mu)$
- Các V_i đều thuộc các lân cận $G_{V'_i}$ mà có đường kính là $O(\log n)$.

(6)

Bây giờ ta sẽ chứng minh rằng siêu đồ thị G^z có đường kính $O(1)$. Trước hết ta chứng tỏ rằng $K \succ G^z$ trong đó $K = K(m, z, p)$ với $z = \lambda n^\mu/2$ và $p = n^{\mu-\nu-1}/2$. Hai đồ thị này đều thuộc mô hình NAN ; ta hãy ký hiệu $G^z = NAN(S_1, \tau_1)$ và $K = NAN(S_2, \tau_2)$, trong đó 2 đồ thị cơ sở S_1 và S_2 cùng có kích thước m (phân hoạch $\mathcal{Z}$ có m thành phần). Tuy nhiên ta có quan sát:

- Trọng số mỗi đỉnh của S_1 là cao hơn của mỗi đỉnh S_2 :

$$w_{V_i} = \sum_{v \in V_i} w_v \geq \lambda^* |V_i| \geq \lambda n^\mu/2 = z \quad (7)$$

- Phân phối τ_2 trên K là có xu thế kém τ_1 trên G^z : do tính (μ, ν) -pt giả thiết của G thì sức hút về mỗi siêu đỉnh trên G^z (bao gồm $\geq n^\mu/2$ đỉnh của V) là

$$\geq (n^\mu/2)/n^\nu * n^{-1} = n^{\mu-\nu-1}/2 = p \quad (8)$$

Vì vậy theo quan sát RG-B (mục IV), ta có

$$K \succ G^z \quad (9)$$

Bây giờ ta sẽ chứng tỏ rằng $J \succ_{e(n^{\mu-\nu})} K$ với $J = J(m, Z)$ và $Z = zmp/2$. Điều này có thể chứng minh khá dễ dàng bằng kỹ thuật đã sử dụng trong mục V. Trước hết lưu ý rằng một thể hiện của K có thể được tạo ra như sau: lần lượt với mỗi đỉnh $u \in S_2$ ta lần lượt tiến hành z thí nghiệm Becnuli độc lập với xác suất thành công $(m-1)p$, và nếu thành công thì chọn $v \xleftarrow{\text{uni}} V \setminus \{u\}$ và tạo liên kết (u, v) . Như vậy số LKNN sinh bởi u là tổng của z biến ngẫu nhiên Becnuli đồng nhất với xác suất thành công $(m-1)p$. Do đó ta có thể thiết kế một chương trình $M_1 || M_2$ để sinh K như sau. Trong M_1 ta thực hiện việc tạo một thể hiện của J , tức là với mỗi đỉnh u , tạo ra Z

LKNN đến một đỉnh v_i nào đó: $v_i \xleftarrow{uni} V \setminus \{u\}$ (các đỉnh v_i này có thể chọn trùng lặp). Trong M_2 ta thực hiện lấy mẫu z biến ngẫu nhiên Becnuli đồng nhất và lấy tổng z của chúng, nếu $z > Z$ thì ta sinh thêm $(z-Z)$ LKNN đi từ u đến một đỉnh v_i nào đó: $v_i \xleftarrow{uni} V \setminus \{u\}$.

Sử dụng bổ đề 3.1, dễ thấy là

$$P_{\text{BAD}} = \Pr[z \leq Z] = e\text{Neg}(zmp) = e\text{Neg}(n^{\mu-\nu}) \quad (10)$$

Vì ta có $z = \lambda n^\mu/2$, $p = n^{\mu-\nu-1}/2$ và cũng dễ thấy có $n^{1-\mu} \leq m \leq 2n^{1-\mu}$.

Cũng dễ thấy:

$$\begin{aligned} \Pr[\text{với mọi đỉnh } u: z > Z] &\geq 1 - m * P_{\text{BAD}} \\ &= 1 - n^{1-\mu} * e\text{Neg}(n^{\mu-\nu}) = 1 - e\text{Neg}(n^{\mu-\nu}) \end{aligned} \quad (11)$$

Do đó ta có:

$$J \succ_{e(n^{\mu-\nu})} K \quad (12)$$

Bây giờ ta cần chứng minh nốt rằng $\text{diam}(J) = O(1)$. Nhắc lại là $J = J(m, Z)$ và $Z = zmp/2$, trong đó như đã nêu $z = \lambda n^\mu/2$ và $p = n^{\mu-\nu-1}/2$ và $n^{1-\mu} \leq m \leq 2n^{1-\mu}$. Suy ra $Z = O(n^{\mu-\nu})$.

Vậy nếu chọn $\Delta = \min\left\{\frac{1-\mu}{\mu-\nu} - \frac{1}{2}, \frac{1-\mu}{2(\mu-\nu)}\right\}$ thì rõ ràng $Z^\Delta/m \rightarrow 0$ và $Z^{\Delta+1}/(m \log m) \rightarrow \infty$, khi $n \rightarrow \infty$. Thế thì theo định lý 2, ta có:

$$\text{diam}(J) \leq \Delta + 4 \text{ với } VHP(Z) \quad (13)$$

Kết hợp (9), (12), (13) ta có $\text{diam}(G^{\mathcal{P}}) \leq \Delta + 4$ với $VHP(Z + n^{\mu-\nu})$ tức là với $VHP(n^{\mu-\nu})$. Điều này kết hợp với (6) và theo bổ đề 5.1 ta suy ra rằng đường kính của cả đồ thị $G = \mathcal{NAN}(H, \tau)$ là $(\Delta + 4)\text{Clog}n$, tức là $O(\log n)$, với xác suất $(1 - o(n^{-2})) * (1 - O(e^{-n^{\mu-\nu}}))$, tức là với xác suất $(1 - o(n^{-2}))$ do $e^{-n^{\mu-\nu}} = o(n^{-2})$. Vậy ta có đpcm!

VI. KẾT LUẬN

Trong bài báo này, chúng tôi đã xây dựng một mô hình phổ quát cho các mạng TGN, đồng thời đề xuất và chứng minh một điều kiện đủ khái quát để đạt đường kính đồ thị nhỏ. Từ đó, chúng tôi đã giới thiệu một tiếp cận mới cho việc phân tích đánh giá đường kính đồ thị cho các dạng mô hình TGN. Do hạn chế khuôn khổ bài báo, chúng tôi mới chỉ đưa ra một số ví

dụ minh họa ngắn gọn cho ứng dụng của tiếp cận này mà chưa đi sâu phân tích các ví dụ phức tạp. Tuy nhiên, chúng tôi nhận thấy đây là một tiếp cận đầy triển vọng, bởi vì, một số kết quả đã biết về đường kính đồ thị trước đây đã được nhiều tác giả chứng minh rất phức tạp trong các bài báo riêng rẽ thì có thể chứng minh lại khá giản dị bằng cùng một cách tiếp cận này. Vì vậy có thể xem đây là một kết quả có ý nghĩa đáng kể cho quá trình chung xây dựng một lý thuyết mới về các mô hình mạng dựa trên cơ sở DTNN.

Trong tương lai chúng tôi hy vọng sẽ tiếp tục phát triển thành công tiếp cận này để đề xuất và khảo sát thêm nhiều dạng mô hình mạng phức tạp, xuất phát từ thực tế, trong đó chúng tôi đã có các kết quả ban đầu (sẽ công bố sau) trong việc mô hình hóa đồ thị luật Lũy thừa cũng như một vài dạng mô hình đồ thị của mạng vật lý Internet mà các yếu tố địa lý đóng góp vai trò chủ đạo trong sự hình thành.

LỜI CẢM ƠN

Chúng tôi xin cảm ơn các đồng tác giả trong [5, 21, 18] đã đóng góp nhiều ý kiến gợi ý.

PHỤ LỤC

1. Bất đẳng thức Chernoff

Chúng tôi sử dụng phát biểu sau lấy từ nguồn [1].

Giả sử $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên Becnuli độc lập với $E[X_i] = p_i$. Đặt $X = \sum_{i=1}^n X_i$, $p = \frac{1}{n} \sum_{i=1}^n p_i$ và $q = 1 - p$. Thế thì với $0 < t < q$ ta có:

$$\Pr[X \geq n(p+t)] \leq e^{-2t^2n} \quad (14)$$

Hơn nữa với $0 < \epsilon < 1$ ta có

$$\Pr[X \leq pn(1-\epsilon)] \leq e^{-\epsilon^2 pn/2} \quad (15)$$

2. Mô hình mở rộng $\mathcal{NAN}(H, \tau, p)$

Chúng tôi cũng xem xét một mô hình khác có sự mở rộng bổ sung, trong đó số LKNN không phải một hằng số cố định mà là một biến ngẫu nhiên nhị thức (Binomial) gắn với thông số tại đỉnh u , tức là theo phân phối $\text{Bin}(w_u, p)$. Nói một cách nôm na, mỗi đỉnh u được phát w_u “hạt giống”; mỗi hạt giống này có thể sinh được một LKNN với xác suất p .

Định nghĩa 3.2 Với một cặp (H, τ) như ở định nghĩa 1 và một hằng số $p \in (0,1]$, ta định nghĩa các mạng ngẫu nhiên $\mathcal{NAN}(H, \tau, p)$ là những mạng sinh từ cơ sở H theo cách sau: với mỗi đỉnh $u \in V(H)$ bổ sung $X_u \sim \text{Bin}(w_u, p)$ LKNN, mà mỗi liên kết này sẽ đi từ đỉnh u đến một đỉnh v được chọn với xác suất $\tau_u(v)$.

Phù hợp với mô hình này, trọng số của một tập con $S \subseteq V(H)$ là $\sum_{u \in S} pw_u$.

Ví dụ 2. Xét $G = \mathcal{NAN}(H, \tau, p)$ trong đó cơ sở là H là một tập đỉnh đơn có trọng số (tức là $E(H) = \emptyset$), và đồng thời $\tau_u(v)$ tỷ lệ thuận ($\propto$) với w_v . Không khó để thấy rằng bậc của đỉnh u là một biến ngẫu nhiên có kỳ vọng $2pw_u$ (trong đó một nửa số liên kết tầm xa sinh bởi chính u , còn một nửa còn lại sinh bởi các đỉnh khác rồi đi về u). Tương tự như mô hình Chung-Lu, dễ thấy mô hình đang xét cũng mô phỏng một đồ thị power-law với vector trọng số bất kỳ gán cho H .

Phân khảo sát chi tiết và các kết quả liên quan đến mô hình mở rộng $\mathcal{NAN}(H, \tau, p)$ này được dành cho các công bố sau.

3. Chứng minh Bổ đề 5.1.

Với mọi cặp đỉnh $s, t \in V$ thì ta có thể xây dựng một danh sách $V_{i_0}, V_{i_1}, V_{i_2}, \dots, V_{i_k}$ gồm các tập đỉnh phân biệt của $\mathcal{T}$ sao cho:

- $s \in V_{i_0}$ còn $t \in V_{i_k}$ và $k \leq \text{diam}(G^{\mathcal{T}})$
- Giữa 2 tập đỉnh kề nhau trong chuỗi trên, gọi là V_i và V_j , luôn tồn tại các đỉnh $u \in V_i$ và $v \in V_j$ mà (u, v) là một cạnh của G .
- Ngoài ra, với 2 đỉnh u và v bất kỳ cùng thuộc một tập đỉnh thì $d(u, v) \leq p$.

Từ trên rõ ràng ta có thể xây dựng được một đường đi s - t trải qua chuỗi tập đỉnh nói trên mà có độ dài $\leq p \cdot k$, tức là $d(s, t) \leq p \cdot q$. *Dpcm!*

TÀI LIỆU THAM KHẢO

- [1] B. BOLLOBAS, "Random graphs," vol. 44, pp. 1–20, 1991.
- [2] B. BOLLOBAS, *Random Graphs*, 2nd ed. Cambridge University Press, 2001, ch. 10.
- [3] F. CHUNG and L. LU, "Connected components in random graphs with given degree sequences," *Annals of Combinatorics*, vol. 6, pp. 125–145, 2002.
- [4] F. CHUNG and L. LU, "The average distances in random graphs with given expected degrees," *Internet Mathematics*, vol. 1, pp. 91–114, 2003.
- [5] P. ERDOS and A. RENYI, "On random graphs", *Publicationes Mathematicas* 6, p. 290-297, 1959.
- [6] C. FALOUTSOS, P. FALOUTSOS, and M. FALOUTSOS, "On power-law relationships of the internet topology," in *Proc. of ACM SIGCOMM*, 1999.
- [7] P. FRAIGNIAUD, E. LEBHA, and Z. LOTKER, "A doubling dimension threshold $\theta(\log \log n)$ for augmented graph navigability," in *14th Annual European Symposium on Algorithms (ESA)*, 2006.
- [8] P. FRAIGNIAUD & G. GIAKKOUPIS, "On the searchability of small-world networks with arbitrary underlying structure", In *Proc. of the 42nd ACM Symposium on Theory of Computing (STOC 2010)*.
- [9] G. GIAKKOUPIS & N. SCHABANEL, "Optimal path search in small worlds: dimension matters", In *Proc. of the 43rd ACM Symposium on Theory of Computing (STOC 2011)*.
- [10] J. KLEINBERG, "The small-world phenomenon: An algorithmic perspective," in *Proc. of ACM Symp. on Theory of Computing (STOC)*, 2000.
- [11] J. KLEINBERG, "Small-world phenomena and the dynamics of information," in *Neural Info. Proce. Sys. (NIPS)*, 2001.
- [12] M. KOIBUCHI, H. MATSUTANI, H. AMANO, D. FRANK HSU, H. CASANOVA, "A Case for Random Shortcut Topologies for HPC Interconnects", In *Proc. of the 39th International Symposium on Computer Architecture (ISCA'2012)*.
- [13] E. LEBHAR & N. SCHABANEL, "Almost optimal decentralized routing in long-range contact networks," in *International Colloquium on Automata, Languages and Programming*, 2004, pp. 894–905.
- [14] D. LIBEN-NOWELL, J. NOVAK, R. KUMAR, P. RAGHAVAN, and A. TOMKINS, "Geographic routing in social networks," In *Proceedings of the National Academy of Sciences*, vol. 102, no. 33, pp. 11 623–11 628, 2005.
- [15] D. MALKHI, M. NAOR & D. RATAJCZAK, "Viceroy: A Scalable and Dynamic Emulation of the Butterfly", in *Proc. of the 21st annual symposium on Principles of distributed computing (PODC'2002)*.
- [16] G. MANKU, M. BAWA, P. RAGHAVAN, "Symphony: Distributed Hashing in a Small World",

USENIX Symposium on Internet Technologies and Systems, 2003.

[17] C. MARTEL and V. NGUYEN, “Analyzing Kleinberg’s (and other) smallworld models,” in Proc. of ACM Symp. On Princ. of Dist. Comp. (PODC), 2004.

[18] S. Milgram, “The small world problem,” Psychology Today, vol. 22, pp. 61–67, 1967.

[19] V. NGUYEN and C. MARTEL, “Augmented Graph Models for Small-World Analysis with Geographic Factors”, in Proc. of the Fifth ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO), 2008.

[20] V. NGUYEN, N. T. LE-XUAN, M. KOIBUCHI, and I. FUJIWARA, “Distributed shortcut networks: Layout-aware low-degree topologies exploiting small-world effect,” in Proc. of the 42nd International Conference on Parallel Processing (ICPP-2013).

[21] A. SINGLA, C.-Y. HONG, L. POPA, and P. B. GODFREY, “Jellyfish: Networking Data Centers Randomly,” in Proc. of USENIX Symposium on Network Design and Implementation (NSDI), 2012.

[22] D. WATTS and S. STROGATZ, “Collective dynamics of small-world networks,” Nature, vol. 393, pp. 440–32, 1998.

[23] S. YOOK, H. JEONG, and A. BARABASI, “Modeling the internet large-scale topology,” in Proceedings of the National Academy of Sciences, vol. 99, 2002, pp. 13 382– 13 386.

Nhận bài ngày: 29/03/2014

SƠ LƯỢC VỀ TÁC GIẢ

NGUYỄN KHANH VÂN

Sinh năm 1970 tại Hà nội.

Tốt nghiệp trường ĐH Bách khoa Hà nội năm 1992. Nhận bằng Thạc sỹ năm 2000 tại ĐH Wollongong, Úc. Nhận bằng Tiến sỹ năm 2006 tại ĐH California-



Davis, Hoa kỳ.

Hiện công tác tại Viện CNTT&TT, trường ĐH Bách khoa Hà Nội.

Hướng nghiên cứu chính: Tính toán phân tán, Mô hình mạng mới và thuật toán, Mạng máy tính, An toàn thông tin.

Email: vannk@soict.hust.edu.vn