

Phân tích Mạng Xã hội theo Chủ đề và Ứng dụng vào Công tác Tư vấn Học tập cho Sinh viên

Hồ Trung Thành

NCS Khóa 5, Trường ĐH Công nghệ Thông tin
thanht@uel.edu.vn

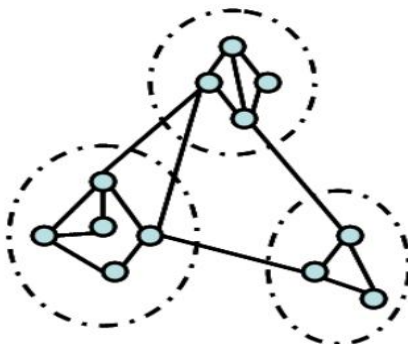
Tóm tắt. Trong luận án này, trước tiên chúng tôi tập trung nghiên cứu về mạng xã hội và những ứng dụng thực tế mang lại từ mạng xã hội cho cá nhân, tổ chức. Thứ hai, chúng tôi nghiên cứu các mô hình trong phân tích mạng xã hội như xác định hệ thống chủ đề và phân lớp theo đặt trung chủ đề được trao đổi thảo luận trên mạng xã hội theo hướng tiếp cận mạng Bayesian. Thứ ba, chúng tôi nghiên cứu mô hình xác định cấu trúc mạng và tìm ra trung tâm lan truyền thông tin trên mạng xã hội. Cuối cùng chúng tôi tập trung ứng dụng kết quả nghiên cứu vào việc tư vấn học tập, nghiên cứu, trao đổi của sinh viên trên mạng xã hội.

Từ khóa: Mạng Bayes, phân tích mạng xã hội, mô hình xác suất, mô hình chủ đề, rút trích cộng đồng.

1 Giới thiệu

Thế giới đã chứng kiến sự tiến triển vượt bậc của mạng xã hội cũng như những lợi ích mà mạng xã hội đem lại. Tầm quan trọng của mạng xã hội đã được khẳng định trong rất nhiều lĩnh vực kinh tế, giáo dục, nghiên cứu, xã hội [1] [2][4][16][19]. Thông qua các mô hình ứng dụng mạng xã hội, các cá nhân và tổ chức có thể làm chủ thông tin trên trên đó thông qua những trao đổi, kết nối cộng đồng của họ với các cá nhân và tổ chức khác[5]. Những kết nối đó đã tạo điều kiện cho việc hình thành một cộng đồng mạng trao đổi, chia sẻ thông tin (Hình 1) và cũng từ đó mạng xã hội đã mở ra cho nhiều nhà nghiên cứu, nhà khoa học phát triển và đem lại nhiều mô hình, kỹ thuật giúp khai thác và phân tích dữ liệu thu thập từ việc trao đổi thảo luận của cộng đồng trên mạng xã hội [5][6][7].

Ngày nay, mạng xã hội giáo dục đã được phát triển nhằm để hỗ trợ cho việc giảng dạy và học tập trong trường đại học. Mạng xã hội là những cộng đồng ảo mang lại sự cộng hưởng cho môi trường học thuật truyền thống. Như nhiều nhà nghiên cứu đã chỉ ra, mạng xã hội có thể mang lại nhiều lợi thế cho việc giảng dạy và học tập. Mạng xã hội cung cấp các phương thức mới để học tập, đặc biệt ở mức độ đại học. Các kiến thức và kỹ năng xã hội có thể được huấn luyện thông qua các giao tiếp giữa sinh viên trong mạng xã hội. Thêm vào đó, sinh viên có một khoảng thời gian học tập tại trường đại học, do đó sẽ tốt hơn nếu các mạng xã hội giáo dục có thể cung cấp các dịch vụ xã hội để họ có thể liên lạc với những thành viên khác ngay trong môi trường học tập, nghiên cứu.



Hình. 1. Cộng đồng trong mạng xã hội (Nguồn [16])

Mạng xã hội đã được sử dụng hiệu quả trong giáo dục, được triển khai như môi trường thảo luận để hỗ trợ cho các phương pháp học tập truyền thống và đã mang lại hiệu quả tốt cho kết quả học tập của sinh viên [20]. Mạng xã hội cũng được dùng để khuyến khích sinh viên tham gia vào giao tiếp xã hội [21]. Nghiên cứu trên đã cho thấy thành tích học tập của sinh viên tỷ lệ với mối quan hệ và hoạt động của sinh viên đó trên mạng xã hội. Một mạng xã hội cũng được sử dụng để tạo môi trường cho học tập nhóm và làm việc tương tác [22]. Phương pháp sử dụng mạng xã hội trong giảng dạy nghiên cứu đã thành công trong việc nâng cao tính sáng tạo và học tập nhóm trong sinh viên.

Sự tương tác hai chiều trên mạng xã hội sẽ giúp cho giảng viên, sinh viên hiểu nhau nhiều hơn trong vấn đề đào tạo và giáo dục tại trường đại học. Giảng viên biết sinh viên thiếu kiến thức gì và hiểu sinh viên cần gì để lấp vào lỗ hổng kiến thức đó. Bên cạnh đó, việc thông tin hai chiều và mạng xã hội có sự trao đổi, kết bạn của cả cộng đồng và mức độ ảnh hưởng lan truyền của một thông tin tốt hay xấu có thể nói dẫn đến một kết quả tốt hay xấu điều này rất quan trọng. Như vậy, vấn đề đặt ra là làm sao người giảng viên hiểu được sinh viên của mình đang trao đổi những chủ đề trong học tập, nghiên cứu, việc làm, giảng viên, môn học, đào tạo, ngành nghề,...? Từ cơ sở hiểu sinh viên thông qua những chủ đề, thông tin, dữ liệu thu thập được, người giảng viên có thể kịp thời cung cấp những thông tin, kiến thức để kịp thời tư vấn cho sinh viên trong việc định hướng nghiên cứu, học tập,... theo chủ đề mà sinh viên quan tâm. Như vậy, vấn đề đặt ra là làm sao phân tích thông tin, dữ liệu theo chủ đề được sinh viên trao đổi trên mạng xã hội mà hệ thống thu thập được thật sự có ích, giúp kịp thời cung cấp thông tin hiệu quả và đáp ứng được nhu cầu trong công tác tư vấn học tập cho sinh viên. Trên đây là những vấn đề cốt lõi mà luận án của chúng tôi tập trung nghiên cứu làm rõ.

2 Công trình Liên quan

Trên thế giới, đã có nhiều công trình nghiên cứu liên quan đến nghiên cứu được chúng tôi đặt ra. Dưới đây là một trong số những công trình tiêu biểu:

Trong công trình nghiên cứu [12][13], các tác giả đã trình bày những nghiên cứu về mô hình ART (Author-Recipient-Topic) trong phân tích mạng xã hội. Trong mô hình nghiên cứu này tác giả tập trung vào việc phân lớp các chủ đề dựa trên sự lắng nghe trực tiếp giữa những trao đổi của các thực thể (người sử dụng mạng xã hội). Mô hình nghiên cứu này được xây dựng dựa trên việc nghiên cứu mô hình LDA (Latent Dirichlet Allocation)[12] và mô hình AT (Author-Topic) [4], thêm vào đó, các tác giả đã nghiên cứu xây dựng những thuộc tính khóa trên từng chủ đề để giúp phân biệt những trao đổi theo từng chủ đề của cả người gửi và người nhận – đồng thời chỉ định cho việc khám phá những chủ đề được trao đổi tùy thuộc vào mối quan hệ giữa những người sử dụng với nhau.

Trong một số kết quả nghiên cứu khác về phân tích mạng xã hội, các tác giả trình bày về những nghiên cứu liên quan đến mô hình xác suất để xác định chủ đề, phân tích mạng thông tin liên kết giữa những trao đổi của người sử dụng [12][14] bằng hai cách tiếp cận dựa trên chủ đề và liên kết thông tin trong mạng xã hội. Trong hai công trình nghiên cứu [16], tác giả đề cập đến những hạn chế của phương pháp phân tích mạng xã hội hoàn toàn dựa trên chủ đề [11] là vấn đề ngôn ngữ bản địa. Tác giả đề cập cùng một chủ đề (một từ, hay cụm từ) nhưng ở địa phương này hiểu khác và ở địa phương khác lại hiểu khác và kết quả hai người cùng trao đổi một nội dung, chủ đề nhưng không được phát hiện là cùng một cộng đồng trao đổi [8].

Trong công trình nghiên cứu [16], tác giả đề xuất mô hình CART (Community-Author-Recipient-Topic) được xem là mô hình mở rộng của ART theo hướng tiếp cận mạng Bayesian. CART là mô hình Cộng đồng-Người gửi-Người nhận-Chủ đề, mục tiêu của việc xây dựng mô hình này là giúp rút trích cộng đồng trên mạng xã hội trên cùng chủ đề.

Bên cạnh các công trình nghiên cứu trên, công trình nghiên cứu [18] đã trình bày nghiên cứu về xác định và rút trích cộng đồng thông qua việc trao đổi của nhóm người trên mạng xã hội. Cũng trong công trình trên, tác giả trình bày giả định rằng một thành viên trong cộng đồng có sự tương tác và thông tin liên lạc với thành viên khác trong cộng đồng đó. Một cộng đồng có thể tham gia thảo luận trên nhiều chủ đề và một chủ đề có thể có nhiều cộng đồng tham gia. Điều này giúp tác giả khám phá được sự quan tâm và cũng như lợi ích của cộng đồng đến từng chủ đề được trao đổi.

Hầu hết những kết quả nghiên cứu được trình bày ở trên tập trung vào nghiên cứu và đề xuất các mô hình như AT, ART, LDA, CART và các mô hình xác suất trong phân tích mạng xã hội để phát hiện chủ đề, phân tích nội dung thảo luận, rút trích cộng đồng trao đổi theo chủ đề mà người sử dụng mạng xã hội quan tâm mà chưa thấy rõ được những ứng dụng những kết quả nghiên cứu vào thực tiễn. Chính vì vậy, trong nghiên cứu này chúng tôi tập trung vào việc nghiên cứu các mô hình đã đề cập trên để ứng dụng vào việc xác định và phân lớp theo đặt trưng của chủ đề. Trong việc phân lớp theo đặt trưng chủ đề chúng tôi sẽ sử dụng kết quả nghiên cứu nghiên cứu có sẵn để thực hiện. Sau đó, chúng tôi tiếp tục nghiên cứu mô hình đồ thị xác suất và các mô hình khác để ứng dụng vào việc xác định trung tâm của sự lan truyền thông tin trên mạng xã hội theo chủ đề. Từ những kết quả nghiên cứu trên, chúng tôi tập trung ứng dụng vào việc tư vấn học tập cho sinh viên theo từng chủ đề được sinh viên trao đổi trên mạng xã hội.

3 Mục tiêu và Ý nghĩa Nghiên cứu

3.1 Mục tiêu Nghiên cứu

Một là, nghiên cứu vấn đề phân tích thông tin, xác định hệ thống chủ đề theo đặc trưng chủ đề được thảo luận trên mạng xã hội

Hai là, nghiên cứu các giải thuật phân lớp theo đặt trưng của chủ đề

Ba là, xác định cấu trúc mạng và tìm ra trung tâm của sự lan truyền thông tin theo đặt trưng chủ đề trên mạng xã hội và ứng dụng vào tư vấn học tập, nghiên cứu của sinh viên.

3.2 Ý nghĩa Nghiên cứu

Một vấn đề mà có thể tất cả chúng ta thấy được là hiệu quả của mạng xã hội ngày nay mang lại cho cả cá nhân và tổ chức giúp họ gần nhau hơn. Với nghiên cứu trong lĩnh vực này, chúng tôi hy vọng sẽ đem lại 3 ý nghĩa khoa học và thực tiễn:

- Nghiên cứu, phân tích và đánh giá các mô hình đã được nghiên cứu trên thế giới như mô hình AT, LDA, ART, CART và các mô hình đồ thị xác suất theo hướng tiếp cận mạng Bayes trong phân tích mạng xã hội. Với kết quả nghiên cứu này, chúng tôi sẽ chọn lựa mô hình phù hợp trên cơ sở phân tích, đánh giá để giải quyết bài toán xác định và phân lớp theo đặt trưng chủ đề trên mạng xã hội.
- Sau khi lựa chọn được mô hình, chúng tôi tiếp tục nghiên cứu để tìm ra được mô hình xác định trung tâm của sự lan truyền thông tin trên mạng xã hội theo chủ đề.
- Ứng dụng kết quả vào tư vấn học tập cho sinh viên. Đây là một ứng dụng mới, thực tiễn mà theo tìm hiểu của chúng tôi chưa thấy có nghiên cứu trước đây thực hiện.

3.3 Bài toán Đặt ra

Với mục tiêu và ý nghĩa nghiên cứu trình bày ở trên, chúng tôi đặt ra ba bài toán cần giải quyết:

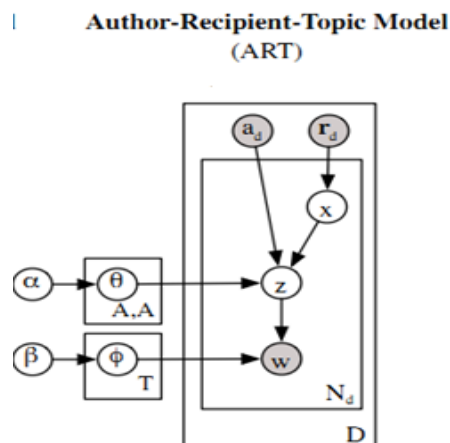
- Bài toán thứ nhất là xây dựng hệ thống chủ đề theo đặt trưng của chủ đề trong học tập và nghiên cứu của sinh viên.
- Bài toán thứ hai là phân lớp chủ đề theo đặt trưng của chủ đề được sinh viên trao đổi và thảo luận trên mạng xã hội.
- Bài toán thứ ba là tìm ra trung tâm lan truyền thông tin trên mạng xã hội theo từng chủ đề và ứng dụng vào tư vấn học tập cho sinh viên.

4 Kết quả Đạt được

Qua thời gian gần 10 tháng nghiên cứu thực hiện luận án, chúng tôi đã nghiên cứu được khoảng gần 30 bài báo, công trình tiêu biểu liên quan đến đề tài. Trong đó, chúng tôi tập trung vào nghiên cứu các mô hình AT, ART, LDA, CART, và các mô hình xác suất khác trong vấn đề xác định và phân lớp chủ đề theo hướng tiếp cận mạng Bayesian. Hiện tại, chúng tôi đã có đầy đủ cơ sở lý luận cũng như những kiến thức nền tảng về vấn đề nghiên cứu, xác định rõ được hướng nghiên cứu và đã chuẩn bị một hệ thống tài liệu tham khảo.

Trong phần này kết quả đạt được này, chúng tôi trình bày chi tiết hơn phần nghiên cứu về mô hình ART.

Mô hình ART mô tả các nút tương tác với nhau thông qua phân tích thông tin truyền đi qua các cạnh giữa các nút trong mạng. Một chủ đề được trao đổi trên mạng sẽ gắn với một cặp người gửi (Author) và người nhận (Recipient) và khám phá vai trò của người gửi và người nhận trong quá trình truyền đạt thông tin, thông điệp qua lại với nhau trên cùng một chủ đề (Topic).



Hình. 2. Mô hình ART (Nguồn [12])

Quy ước:

d: thông tin

a_d : một người gửi thông tin

r_d : tập hợp người nhận thông tin

x: một người nhận thông tin

z: một chủ đề

$\theta_{a_d x}$: phân phối chủ đề cụ thể cho từng cặp người gửi - người nhận ($a_d ; x$)

ϕ_z : các từ quang trọng trong chủ đề z được phân phối trong mạng

T: tập hợp các chủ đề

D: tập hợp các thông tin (document, email, messages)

A: tập hợp các tài khoản của những người gửi và người nhận

V: tập hợp các từ khoá trong thông tin được gửi

N_d: tập hợp các ký hiệu hay từ mà tạo nên thông tin d

Để tiếp cận với mô hình ART trong phân tích mạng xã hội, ta cần quan tâm đến những đặc điểm sau:

- Các loại nút trong mạng xã hội
- Giá trị các thuộc tính trên một nút trong mạng xã hội
- So sánh các thuộc tính trên một nút với giá trị các thuộc tính đó trên các nút khác mà nút đó kết nối đến
- Sự tồn tại của một loại liên kết (cạnh) cụ thể giữa hai nút
- Giá trị của một thuộc tính trên một liên kết giữa hai nút
- Tính toán các giá trị của thuộc tính liên kết giữa hai nút (ví dụ: khả năng dự đoán hai nút có mối quan hệ với nhau)
- Số lượng các liên kết giữa hai nút
- Chú ý các nút quan trọng, chủ chốt (đây chính là trung tâm của sự lan truyền thông tin), khi sử dụng lý thuyết đồ thị và kết hợp các thuật toán tìm kiếm khác để phân tích mạng xã hội theo hướng tiếp cận mạng Bayesian

Bên cạnh một số kết quả nghiên cứu trên, chúng tôi còn đang tập trung vào nghiên cứu và giải quyết bài toán thứ nhất về xây dựng hệ thống chủ đề theo đặt trung của chủ đề trong học tập và nghiên cứu của sinh viên.

5 Kết luận

Cho đến thời điểm này, chúng tôi có thể nói rằng hướng nghiên cứu của chúng tôi là phù hợp, mang tính khoa học và thực tiễn cao. Chúng tôi luôn kỳ vọng vào những kết quả nghiên cứu tiếp theo mà chúng tôi đã có kế hoạch thực hiện. Việc giải quyết bài toán xác định trung tâm của sự lan truyền thông tin theo chủ đề, chúng tôi xác định đây là bài toán khó và chưa có nhiều nghiên cứu. Tuy nhiên, chúng tôi sẽ tập trung nghiên cứu nhiều hơn nữa để có những kết quả như mong muốn. Bên cạnh đó, việc nghiên cứu và thừa kế những kết quả, công trình nghiên cứu đã được công bố trên thế giới chúng tôi sẽ tiếp tục quan tâm và nghiên cứu.

6 Kế hoạch Nghiên cứu Dự kiến

Công việc tiếp theo trong quá trình nghiên cứu, chúng tôi tập trung hoàn thành bài toán thứ nhất và tiếp tục nghiên cứu và giải quyết hai bài toán thứ hai và thứ ba. Bên cạnh đó, chúng tôi sẽ xây dựng mạng xã hội từ hệ thống mã nguồn mở để đưa vào phục vụ cho việc học tập và nghiên cứu của sinh viên. Từ những trao đổi và thảo luận theo từng chủ đề của sinh viên, chúng tôi có được hệ thống dữ liệu. Tiếp sau đó, chúng tôi xây dựng hệ thống chủ đề theo đặt trung của chủ đề từ hệ thống chủ đề trong lĩnh vực giáo dục đại học mà tập trung vào học tập và nghiên cứu của sinh viên. Từ đó, chúng tôi tiếp tục nghiên cứu các mô hình và giải thuật để phân lớp đặt trung theo chủ đề và nghiên cứu giải quyết bài toán xác định trung tâm của sự lan truyền

thông tin theo chủ đề trên mạng xã hội. Cuối cùng, chúng tôi tập trung vào ứng dụng kết quả nghiên cứu vào công tác tư vấn học tập cho sinh viên trên mạng xã hội.

Tài liệu tham khảo

1. Rohit Parimi and Doina Caragea: Predicting Friendship Links in Social Networks Using a Topic Modeling, Computing and Information Sciences, Kansas State University, Manhattan, KS, USA (2011).
2. Approach Wasserman, S., & Faust, K.: Social network analysis: Methods and applications, Cambridge University Press (1994).
3. Mohammad Javad Mosadegh, Mehdi Behboudi: Using social network paradigm for developing a conceptual framework in CRM, Australian Journal of Business and Management Research (2011).
4. Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T.: Probabilistic author-topic models for information discovery (2004).
5. Tom Chapman (2008): Social network marketing, engagement marketing and brand, Website: www.socialnetworkmarketinguk.com.
6. David Jensen and Jennifer Neville: Data Mining in Social Networks, University of Massachusetts (2010).
7. Jaideep Srivastava, Muhammad A. Ahmad, Nishith Pathak, David Kuo-Wei Hsu: Data Mining Based Social Network Analysis from Online Behaviour, University of Minnesota (2008).
8. Paola Velardi, Roberto Navigli, Alessandro Cucchiarelli, Fulvio D'Antonio: A New Content-Based Model for Social Network Analysis, University of Roma "La Sapienza" and University of Marche (2009).
9. Alessandro Cucchiarelli And Fulvio D'antonio, Roberto Navigli And Paola Velardi: Semantically Interconnected Social Networks, Polytechnic University of Marche and University of Roma Sapienza (2010).
10. T. Andrew Yang, Dan J. Kim, and Vishal Dhalwani: Social Networking as a New Trend in e-Marketing, University of Houston – Clear Lake Houston, Texas, USA (2008).
11. Zsolt Katona, Peter Pal Zubcsek and Miklos Sarvary: Network Effects and Personal Influences: The Diffusion of an Online Social Network (2010).
12. Andrew McCallum, Andreis Corrada-Emmanuel, Xuerui Wang: Topic and Role Discovery in Social Networks, University of Massachusetts Amherst, USA (2005).
13. Kevin R. Canini, Lei Shi, Thomas L. Griffiths: Online Inference of Topic with Latent Dirichlet Allocation, AISTATS 12th, USA (2009).
14. Amit Goyal, Francesco Bonchi, Laks V.S. Laskhmanana: Learning Influences Probabilities In Social Networks, WSDM'10, USA (2010).
15. Lulu, Jia Tang, Jiawei Han, Meng Jiang, Shiqiang Yang: Mining Topic-level Influences In Heterogeneous Networks, CIKM'10, Canada (2010).
16. Nishith Pathak, Colin DeLong, Arindam Banerjee: Social Topic Models for Community Extraction, SNA-KDD'08, USA (2008).
17. Anna Goldenberg, Andrew W. Moore: Bayes Net Graphs to Understand Co-authorship Networks?, LinkKDD '05, USA (2005).
18. Rongjing Xiang, Jennifer Neville, Monica Rogati: Modeling Relationship Strength in Online Social Networks, IW3C2, North Carolina, USA (2010).
19. Mrinmaya, Sachan Danish, Contractor Tanveer, A. Faruque, L. Venkata Subramaniam: Using Content and Interactions for Discovering Communities in Social Networks, IBM Research India (2012).

20. H.-L. Yang and J.H. Tang: Effects of social network on students' performance: A web-based forum study in Taiwan, *Journal of Asynchronous Learning Networks*, vol. 7, no. 3, pp. 93-197 (2003).
21. A. Calvó-Armengol, E. Patacchini, and Y. Zenou: Peer Effects and Social Networks in Education, *Review of Economic Studies*, vol. 76, no. 4, pp. 1239-2167 (2009).
22. A. Mora-Soto, M. Sanchez-Segura, F. Medina-Dominguez, and A. Amescua: Collaborative Learning Experiences Using Social Networks, in *International Conference on Education and New Learning Technologies (EDULEARN09)*, Barcelona, Spain, pp. 4260-4270 (2009).