

MỤC LỤC

Lời nói đầu	1
Chương 1. GIỚI THIỆU VỀ MẠNG XÃ HỘI VÀ TÁC HẠI CỦA THÔNG TIN SAI LỆCH	11
1. Giới thiệu chung về mạng xã hội	11
1.1. Đặc điểm của xã hội	12
1.2. Lợi ích của mạng xã hội	13
2. Sự lây lan và tác hại của thông tin sai lệch trên mạng xã hội . . .	16
2.1. Định nghĩa thông tin sai lệch và tính chấn	16
2.2. Thực trạng sự phát tán của thông tin sai lệch trên mạng xã hội	18
Kết luận chương	21
Chương 2. CƠ CHẾ LAN TRUYỀN THÔNG TIN SAI LỆCH	23
1. Mô hình lan truyền thông tin	24
1.1. Mô hình tầng độc lập	25
1.2. Mô hình ngưỡng tuyến tính	27
2. Một số hướng nghiên cứu liên quan bài toán hạn chế lan truyền thông tin sai lệch trên mạng xã hội	29
Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI	31
1. Đặt vấn đề, định nghĩa bài toán	31
2. Tổng thể về giải pháp	33
2.1. Công cụ kiểm thử Selenium	34
2.2. Kỹ thuật lấy dữ liệu trên MXH	36
2.3. Giải pháp hạn chế: Limiting the spread of epidemics within time constraint on online social network (Ngăn chặn khả năng lan truyền của thông tin sai lệch trong thời gian giới hạn trên một mạng xã hội)	41

2.4. Giải pháp hạn chế Targeted Misinformation Blocking (Xác định và ngăn chặn thông tin sai lệch trên mạng xã hội) . . .	50
2.5. Áp dụng các giải pháp đề xuất trên dữ liệu thực	54
Tài liệu tham khảo	56

LỜI NÓI ĐẦU

1. Lí do chọn đề tài

Cùng với sự phát triển của Internet, các mạng xã hội đã phát triển mạnh mẽ và trở thành một xu hướng mới thu hút nhiều người sử dụng trên toàn thế giới. Theo những khảo sát gần đây, có gần một nửa dân số thế giới, tức là hơn 3 tỷ người sử dụng mạng xã hội. Nhờ có mạng xã hội, người dùng có thể trao đổi thông tin với nhau một cách nhanh chóng bất kể khoảng cách về địa lý và thời gian. Bên cạnh đó, mạng xã hội còn cung cấp cho người dùng rất nhiều ứng dụng hữu ích, làm cho cuộc sống của con người ngày càng trở nên thuận tiện hơn. Ngoài những đặc tính kế thừa của mạng lưới xã hội thực như: Tương tác giữa người dùng, lan truyền thông tin, tạo ảnh hưởng trong cộng đồng thì mạng xã hội còn mang nhiều đặc tính mới như: Cập nhật thông tin thực lên mạng xã hội một cách nhanh chóng, sự lan truyền thông tin giữa người dùng xảy ra trong thời gian ngắn, sự bùng nổ thông tin với các nguồn tin tức khác nhau, v.v... Có thể nói, hiện nay mạng xã hội đang từng bước trở thành một kho tri thức mới mà con người có thể dễ dàng tiếp cận

Tuy vậy, đi cùng với những lợi ích mạng xã hội mang lại thì còn nhiều nguy cơ và hiểm họa tới người dùng, trong đó có các yếu tố xấu như tin đồn sai lệch được lây lan nhanh chóng. Các yếu tố này gây ra những tác hại lớn đối với cộng đồng người sử dụng mạng xã hội. Không chỉ ở Việt Nam, những tác hại này diễn ra trên phạm vi toàn thế giới. Việc lan truyền thông tin sai lệch có thể gây ra những tác động xã hội tiêu cực, thậm chí là tổn thất lớn về kinh tế. Chẳng hạn, những tin đồn không hay về sức khỏe của tổng thống Mỹ ông Obama vào tháng 4 năm 2013 đã dẫn đến sự bất ổn của thị trường tài chính ở phố Wall. Trước thềm Đại hội đại biểu toàn quốc lần thứ XII của Đảng Cộng sản Việt Nam, đã có những thông tin về gia đình cũng như tài sản của Thủ tướng Nguyễn Xuân Phúc lúc đó vẫn còn là Phó Thủ tướng Chính phủ, nhằm hạ uy tín gây mất lòng tin của nhân dân đối với Thủ tướng.

Twitter, Youtube và Facebook đang phải đối mặt với việc gia tăng áp lực từ

chính phủ của nhiều nước đe dọa áp đặt các đạo luật mới và khoản phạt nếu các mạng xã hội này không có biện pháp loại bỏ nhanh chóng các nội dung tuyên truyền cực đoan cũng như các nội dung vi phạm luật pháp các nước sở tại.

Trong thời gian gần đây, Đảng và Nhà nước ta đã đề ra rất nhiều biện pháp nhằm ngăn chặn thông tin sai lệch trên mạng xã hội và đặc biệt là bộ Thông tin và truyền thông đã tiến hành ngăn chặn, xử lý những thông tin sai sự thật, nghiêm túc xem xét đánh giá, xử lý sai phạm, bộ Thông tin và truyền thông đã bắt các nhà MXH gỡ bỏ các thông tin sai lệch được đăng tải [1]. Bộ quan tâm trước hết đến báo chí và chất lượng báo chí, có những bài viết với số lượng nhiều nhưng chất lượng kém. Bên cạnh đó, ta chủ động phát hiện những thông tin sai lệch, kịp thời chỉnh sửa, đính chính, đưa ra thông tin chính xác nhất.

Những thông tin sai lệch, không kiểm chứng, giật tít tạo sự tò mò cho người dùng trên mạng xã hội còn tiềm ẩn các nguy cơ phát tán mã độc, lừa đảo trên mạng gây mất an ninh, an toàn thông tin. Với thực trạng đó, các nhà khoa học đã nghiên cứu những giải pháp hiệu quả để ngăn chặn thông tin sai lệch. Trong đó việc mô hình hóa quá trình lan truyền thông tin trên mạng là nền tảng trong tiếp cận của họ. Các mô hình lan truyền thông tin, hay còn gọi là mô hình khuếch tán thông tin được các nhà khoa học đưa ra đã mô hình hóa toán học và mô tả một cách tương đối chính xác một mạng lưới và quá trình lan truyền thông tin, lan truyền dịch bệnh ở trên mạng lưới đó.

Thúc đẩy bởi những hiện tượng trên và yêu cầu bức thiết của việc giải quyết và ngăn chặn những tác hại do virus và tin đồn trên mạng xã hội mang lại. Nhóm tác giả đã mạnh dạn nghiên cứu đề tài nhằm mục đích tìm ra giải pháp hợp lý để giải quyết những vấn đề nêu trên. Đặc biệt là trong bối cảnh ở Việt Nam ngày càng mở rộng hợp tác quan hệ với các nước trên thế giới, đi kèm theo nhiều lợi ích đó là đối mặt với nhiều nguy cơ thách thức về An ninh quốc gia. Đặc biệt là các thế lực thù địch lợi dụng MXH để thực hiện những âm mưu, hoạt động nhằm xuyên tạc, chống phá đường lối, chính sách của Đảng, pháp luật của Nhà nước. Nghiêm trọng hơn, trong thời gian gần đây, chúng thực hiện những thủ đoạn cực kỳ tinh vi: Thuê người làm dữ liệu giả (ảnh, video clip, audio hiện

trường) để đưa lên mạng xã hội, các diễn đàn, xuyên tạc sự thật, cung cấp thông tin sai lệch về các cá nhân, các cán bộ, Đảng viên; dùng kỹ thuật và công nghệ để chỉnh sửa dữ liệu cũ, chỉnh sửa hình ảnh, cắt ghép, tạo bằng chứng và thông tin giả; tự bịa ra các bài phỏng vấn nhân vật, sự kiện, bịa đặt các trang hồ sơ liên quan đến các nhân vật nổi tiếng, các nhà lãnh đạo, thân nhân của họ và kích thích trí tò mò của công chúng bằng “thông tin lè trái, thông tin bí mật”, những việc làm này gây hậu quả vô cùng nghiêm trọng đối với tình hình An ninh quốc gia của Việt Nam.

2. Tình hình nghiên cứu liên quan đến đề tài

Ngăn chặn sự phát tán của thông tin sai lệch, tin đồn hiện nay là một chủ đề nóng được nhiều nhà khoa học quan tâm. Rất nhiều những nghiên cứu và bằng sáng chế liên quan vấn đề này có tính ứng dụng cao được áp dụng trên các mạng xã hội, ứng dụng trong giáo dục, y tế sức khỏe, ứng dụng cho chính phủ cũng như các vấn đề bảo vệ An ninh Quốc gia.

Không chỉ ở Việt Nam, đối với các nước trên thế giới, sự phát tán của thông tin sai lệch hiện nay đã trở thành vấn nạn nghiêm trọng. Một số nước đã thành lập các trung tâm chống tin giả sử dụng các biện pháp, giải pháp xã hội cũng như kỹ thuật để ngăn chặn, hạn chế sự phát tán của thông tin sai lệch tới công chúng.

Nguồn phát tán thông tin sai lệch có thể được phát hiện thông qua khảo sát người dùng hoặc các phương pháp khai phá dữ liệu. Ngoài ra, tin đồn còn có thể được phát hiện thông qua việc sử dụng các đặc trưng như: thời gian, cấu trúc và ngữ nghĩa [27]. Về phương diện mô hình hóa toán học, để phát hiện được nguồn phát tán thông tin sai lệch, Zhang và cộng sự đề xuất vấn đề Time Constrained Misinformation Detection (TCMD) tìm kiếm tập đỉnh nhỏ nhất sao cho phát hiện được thông tin sai lệch lớn hơn ngưỡng t . Christakis [2] và cộng sự đề xuất một phương thức phát hiện sự lan truyền của “ổ dịch” bằng cách theo dõi người dùng sử dụng phương pháp lựa chọn ngẫu nhiên.

Nguyen và cộng sự [27] đề xuất chiến lược cho việc “khử nhiễm” thông

tin sai lệch trên mạng xã hội bằng cách dùng thông tin tốt, chính thống để “khử nhiễm” thông tin xấu. Tuy nhiên, khi người dùng đã tin theo thông tin sai lệch sẽ là khó để thay đổi quan điểm của họ và chúng ta cần biết nội dung của thông tin xấu để đưa ra những thông tin chính thống phù hợp.

Gần đây, các nhà khoa học đã đề xuất một chiến lược chung trong việc ngăn chặn sự phát tán của tin đồn, thông tin sai lệch, đó là ngăn chặn những tài khoản và liên kết có vai trò quan trọng trong quá trình lan truyền, có một số phương pháp heuristic hiệu quả đã được đề xuất. Khalil [3] và cộng sự đề xuất bài toán xóa bỏ tập cạnh để giảm thiểu ảnh hưởng của tập nguồn, họ cũng đề xuất một thuật toán giải quyết có tỷ lệ xấp xỉ $1 - 1/e - \varepsilon$. Ceren Budak [4] nghiên cứu bài toán hạn chế sự lây lan của một chiến dịch lan truyền thông tin “xấu” trong mạng xã hội, dựa trên các mô hình lan truyền và tính cạnh tranh của thông tin trên mạng xã hội trực tuyến.

Canh.V Pham [5] và các cộng sự nghiên cứu việc chặn thông tin sai lệch trong một thời gian và ngân sách cho trước. Cui [6] và cộng sự tập trung vào việc nghiên cứu lựa chọn những điểm đặt các cảm biến sao cho phát hiện sự bùng nổ thông tin với xác suất cao nhất. Zhang đề xuất vấn đề t - Monitors Placement ngăn ngừa việc lan truyền thông tin sai lệch với ngưỡng đảm bảo $t \in [0,1]$ đề xuất chiến lược tiêm vaccin cho k đỉnh còn lại sao cho số đỉnh bị lây lan là nhỏ nhất trong mô hình IC. Các tác giả cho thấy rằng vấn đề tối ưu hóa này là NP-khó ngay cả việc tính toán hàm mục tiêu. Ngoài ra, có rất nhiều nghiên cứu liên quan đến vấn đề này trong thời gian gần đây.

Tuy nhiên, hầu hết các nghiên cứu trên đều được thực hiện trên dữ liệu có sẵn đã được công bố và chúng bị hạn chế theo những khía cạnh sau:

- Việc tính toán hàm mục tiêu là NP-Khó trên các mô hình phát tán thông tin cổ điển dẫn đến việc khó thực hiện giả pháp trên các mạng xã hội thực.
- Yếu tố thời gian chưa được tập trung xem xét trong ngăn chặn thông tin sai lệch. Đây là điều rất quan trọng vì thông tin chỉ phát tán nhanh trong một khoảng thời gian nhất định, nếu thông tin sai lệch phát tán trong thời gian càng dài thì càng gây nhiều thiệt hại và ảnh hưởng tới công chúng.

Hơn nữa, việc ngăn chặn ngăn thông tin sai lệch với nguồn lực cho trước có một sự hạn chế là ta không biết ngân sách bao nhiêu để ngăn chặn sự bùng phát của thông tin sai lệch.

- Các nghiên cứu mới chỉ dừng lại ở mô hình và thuật toán mà chưa kết hợp với việc thu thập dữ liệu hiện thực, đặc biệt chưa áp dụng ở các mạng xã hội ở Việt Nam.

Trong đề tài này, nhóm tác giả tập trung khắc phục các hạn chế trên của những nghiên cứu đã có. Đặc biệt, nhóm tác giả đã thu thập dữ liệu của các đối tượng phát tán thông tin sai lệch ở Việt Nam và đưa ra giải pháp ngăn chặn thông tin sai lệch trên các dữ liệu được thu thập đó. Qua đó để xây dựng một hệ thống ngăn chặn thông tin sai lệch trên dữ liệu thu thập được bằng việc đưa ra danh sách người dùng mà việc ngăn chặn họ tin theo, nghe theo, phát tán thông tin sai lệch sẽ hạn chế được tối đa thông tin sai lệch.

3. Mục tiêu, nhiệm vụ nghiên cứu

- **Mục tiêu nghiên cứu:** Đề tài thực hiện để đạt được các mục tiêu sau:
 - + Tìm hiểu, làm rõ tác hại của thông tin sai lệch trên các mạng xã hội.
 - + Tìm hiểu cơ chế lây lan tin đồn và virus trên mạng xã hội. Qua đó đề xuất mô hình lan truyền thông tin phù hợp có yếu tố thời gian.
 - + Thiết lập bài toán hạn chế tin đồn và virus trên mạng xã hội trên mô hình lan truyền thông tin mới.
 - + Đưa ra giải pháp, thuật toán hiệu quả để ngăn chặn sự lây lan tin đồn và virus trên mạng xã hội.
 - + Thu thập dữ liệu về người dùng trên mạng xã hội thực, đặc biệt là dữ liệu các đối tượng phát tán thông tin. Đưa ra kết quả thực nghiệm và đánh giá.
- **Nhiệm vụ nghiên cứu:** Để đạt được mục tiêu nghiên cứu, nhóm tác giả giải quyết lần lượt các nhiệm vụ sau:

- + Nghiên cứu mạng xã hội, cách thức lan truyền thông tin trên mạng xã hội, mô hình hóa toán học của mạng xã hội trong các nghiên cứu từ trước tới nay trên các hội nghị khoa học, tạp chí khoa học uy tín qua các tài liệu, bài báo cáo.
- + Nghiên cứu các thuật toán liên quan đã có giải quyết bài toán ngăn chặn sự lan truyền thông tin sai lệch trên mạng xã hội trực tuyến, cập nhật các nghiên cứu trong thời gian gần đây.
- + Đưa ra thuật toán hiệu quả cho bài toán ngăn chặn sự lây lan của tin đồn và virus trên mạng xã hội, so sánh với các thuật toán mới nhất. Nghiên cứu mô hình hóa mạng xã hội và quá trình lan truyền thông tin trên mạng xã hội.
- + Nghiên cứu giải pháp lấy dữ liệu trên các mạng xã hội, trong phạm vi của đề tài là mạng xã hội Facebook.
- + Đề xuất giải pháp hạn chế thông tin sai lệch trên mạng xã hội và thực nghiệm kiểm tra giải pháp.
- + Thực nghiệm giải pháp trên dữ liệu thực thu thập từ mạng xã hội Facebook. Đưa ra danh sách người dùng có vai trò nhất trong việc phát tán thông tin sai lệch từ một người cho trước. Từ đó đưa ra những giải pháp thích hợp với những người dùng này.

4. Đối tượng và phạm vi nghiên cứu của đề tài

- Đối tượng nghiên cứu:

- + Các mạng xã hội trên thế giới đặc biệt là các mạng xã hội phổ biến ở Việt Nam được đa số người dùng sử dụng.
- + Hành vi phát tán thông tin trên mạng xã hội.
- + Các đối tượng phát tán thông tin sai lệch.

- Phạm vi nghiên cứu

- + Thời gian: Các mạng xã hội từ năm 2001 tới nay.
- + Không gian: Mạng Internet.

5. Phương pháp nghiên cứu

Trên cơ sở những điều kiện và đặc điểm nêu trên, để đạt được mục tiêu nghiên cứu của đề tài, nhóm tác giả sử dụng tổng hợp các cách tiếp cận và phương pháp nghiên cứu sau:

- Phương pháp nghiên cứu tài liệu:

- + Phương pháp phân tích và tổng hợp lý thuyết: Nhóm tác giả đã tham khảo các bài báo, tạp chí, hội nghị có uy tín trong nước và thế giới về những nghiên cứu có liên quan đến đề tài, từ đó phân tích đánh giá, có cái nhìn gần hơn, ví dụ như :

- Science direct: <http://www.sciencedirect.com/>.
- IEEE Xplore Digital Library: <http://ieeexplore.ieee.org>.
- The ACM Digital Library: <http://dl.acm.org/>.
- Google Scholar: <https://scholar.google.com>.

- + Phương pháp phân loại và hệ thống hóa lý thuyết: Từ những lý thuyết, tài liệu đã có, nhóm tác giả phân loại, hệ thống hóa, sắp xếp một cách khoa học, dễ hiểu. Đồng thời, có những nhận định, đánh giá về những lý thuyết, công trình nghiên cứu trước đây, qua đó, nhóm tác giả có những đề xuất tìm ra phương pháp giải quyết tối ưu hơn cho đề tài.

- Phương pháp nghiên cứu thực tiễn

- + Phương pháp hỏi chuyên gia: Thu thập thông tin tài liệu, ý kiến của chuyên gia, đánh giá vấn đề, hỏi ý kiến từ các chuyên gia hàng đầu về lĩnh vực công nghệ thông tin và an toàn thông tin để có cái nhìn tổng quát hơn về vấn đề đang nghiên cứu.
- + Từ thực tiễn đề xuất mô hình phù hợp với thực tế hơn.
- + Phương pháp mô hình hóa: Từ những mô hình mạng xã hội, nhóm tác giả mô hình hóa thành các bài toán phù hợp, từ đó sẽ giải quyết những yêu cầu của bài toán bằng các công cụ của toán học, khoa học

máy tính,...

- + Phương pháp thống kê: Tổng hợp, thống kê những số liệu thu được, mô hình hóa bằng các biểu đồ để có cái nhìn tổng quát hơn, từ đó dễ dàng đánh giá, nhận định vấn đề, đánh giá thuật toán và mô hình đề xuất.

6. Ý nghĩa khoa học, ý nghĩa thực tiễn của đề tài

- Ý nghĩa khoa học:

- + Cung cấp cái nhìn tổng quan về thông tin sai lệch và thực trạng lan truyền thông tin sai lệch hiện nay.
- + Đưa ra những lý thuyết mới có ý nghĩa, những giải pháp có tính khả thi trong nghiên cứu cơ bản cũng như thực tiễn hạn chế thông tin sai lệch trên mạng xã hội.
- + Đưa ra được mô hình mới và thuật toán mới khắc phục được những hạn chế của các mô hình và thuật toán đã công bố.

Qua đó cho thấy, một phần công trình nghiên cứu của nhóm được các đồng nghiệp trên thế giới đánh giá cao và có giá trị về mặt học thuật cũng như giá trị trong thực tiễn. Nhóm tác giả đã vận dụng những nghiên cứu quan trọng này trong đề tài, đề đưa ra giải pháp ngăn chặn thông tin sai lệch có hiệu quả.

- Ý nghĩa thực tiễn:

- + Giúp người dùng có cái nhìn rõ hơn về những tác hại của mạng Internet và đặc biệt là thực trạng về thông tin sai lệch.
- + Đánh giá thực trạng lan truyền, quy luật lan truyền của thông tin sai lệch trên mạng xã hội. Đề xuất một giải pháp hạn chế thông tin sai lệch trên mạng xã hội.
- + Là nền tảng để xây dựng một hệ thống để chặn thông tin sai lệch và có ý nghĩa trong thực tiễn. Đặc biệt trong bối cảnh hiện nay, sự phát

- tán rộng rãi của thông tin sai lệch là vấn nạn đối với các quốc gia. Một số nước phát triển đã xây dựng trung tâm phòng chống tin giả.
- + Qua thu thập dữ liệu người phát tán thông tin ta góp phần vào công tác Công an trong điều tra vụ án.
 - + Từ việc ngăn chặn thông tin sai lệch, ta hạn chế được rất nhiều những hậu quả nghiêm trọng do thông tin sai lệch gây ra.

7. Cấu trúc của đề tài

Ngoài phần mở đầu, kết luận, danh mục tài liệu tham khảo và phụ lục, đề tài được cấu trúc gồm 4 chương:

Chương 1: Giới thiệu về mạng xã hội và tác hại của thông tin sai lệch

Chương này giới thiệu tổng quan về mạng xã hội bao gồm: Định nghĩa mạng xã hội, những đặc trưng cơ bản, lợi ích và tác hại của mạng xã hội. Đặc biệt, trong chương này trình bày thực trạng lan truyền thông tin sai lệch trên mạng xã hội đối với thế giới và Việt Nam, từ đó đặt ra vấn đề cấp thiết trong việc tìm hiểu thông tin sai lệch, cơ chế lan truyền và giải pháp hạn chế tác hại của nó.

Chương 2: Thông tin sai lệch và cơ chế lan truyền thông tin sai lệch

Chương này nhóm tác giả trình bày định nghĩa thông tin sai lệch, những nguy cơ và hậu quả do thông tin sai lệch gây ra đối với các cá nhân, tổ chức. Đồng thời, phân tích cơ chế lan truyền thông tin và những đặc tính của hai mô hình lan truyền thông tin đang được sử dụng rộng rãi bao gồm: Mô hình tầng độc lập và mô hình ngưỡng tuyến tính. Ngoài ra, ở Chương 2 tổng quan một số hướng nghiên cứu liên quan đến bài toán hạn chế lan truyền thông tin sai lệch trên mạng xã hội trực tuyến.

Chương 3: Giải pháp hạn chế thiệt hại do thông tin sai lệch gây ra trên mạng xã hội trực tuyến

Từ thực trạng đã nêu trong Chương 2 và xuất phát từ những công trình nghiên cứu liên quan trước đó, nhóm tác giả đã phát biểu tổng quát bài toán Cực tiểu hóa thiệt hại do thông tin sai lệch gây ra trên mạng xã hội trực tuyến, với dữ

liệu thu thập trực tiếp từ mạng xã hội Facebook xung quanh các đối tượng liên quan đến các vấn đề xâm phạm An ninh Quốc gia, gây rối trật tự an toàn xã hội. Nghiên cứu, chứng minh bài toán thuộc lớp #P – khó, đồng thời đề xuất các giải pháp hiệu quả để giải quyết bài toán này.

CHƯƠNG 1

GIỚI THIỆU VỀ MẠNG XÃ HỘI VÀ TÁC HẠI CỦA THÔNG TIN SAI LỆCH

1. Giới thiệu chung về mạng xã hội

Mạng xã hội (MXH), hay còn gọi là mạng xã hội ảo (Social Network) là dịch vụ nối kết các thành viên cùng sở thích trên Internet với nhiều mục đích khác nhau không phân biệt không gian và thời gian. Những người tham gia vào MXH còn được gọi là “cư dân mạng”.

MXH bao gồm rất nhiều các dịch vụ mang các tính năng riêng biệt: chat, mail, blog, video, chia sẻ thông tin giúp kết nối người dùng dựa vào các mối quan hệ đặc trưng như: Quan hệ họ hàng, sở thích, ý tưởng. Chính vì thế, mạng xã hội dễ dàng giúp con người có thể tìm kiếm kết nối với các mối quan hệ dựa trên các nhóm, trường, cơ quan, dựa trên các thông tin cá nhân, địa chỉ.

MXH bao gồm 2 đặc điểm cơ bản. Đặc điểm thứ nhất là có sự tham gia trực tuyến của các cá nhân hay các chủ thể. Đặc điểm thứ hai là mạng xã hội sẽ có các trang web mở, người dùng tự xây dựng nội dung trong đó và các thành viên trong nhóm đây sẽ biết được các thông tin mà người dùng viết.

Hiện nay thế giới có rất nhiều mạng xã hội khác nhau, thị trường Bắc Mỹ và Tây Âu nổi tiếng với MySpace và Facebook; Nam Mỹ với Orkut và Hi5; Friendster tại Châu Á và các đảo quốc Thái Bình Dương. Mạng xã hội khác gặt hái được thành công đáng kể theo vùng miền như Bebo tại Anh Quốc, CyWorld tại Hàn Quốc, Mixi tại Nhật Bản. Tại Việt Nam, theo một nghiên cứu mới đây của DoubleClick Ad Planner, các trang MXH có lượng truy cập gần 16 triệu lượt/tháng trong đó có ba mạng xã hội lớn nhất tại Việt Nam: Facebook, Zing Me và Yume.

Mục tiêu là tạo ra một hệ thống trên nền Internet cho phép người dùng giao lưu và chia sẻ thông tin một cách có hiệu quả, vượt ra ngoài những giới hạn về địa lý và thời gian. Xây dựng lên một mẫu định danh trực tuyến nhằm phục vụ những yêu cầu công cộng chung và những giá trị của cộng đồng. Qua đó, nhằm

nâng cao vai trò của mỗi công dân trong việc tạo lập quan hệ và tự tổ chức xoay quanh những mối quan tâm chung trong những cộng đồng thúc đẩy sự liên kết các tổ chức xã hội.

1.1. Đặc điểm của xã hội

MXH trên Internet bao gồm các đặc điểm nổi bật: Tính liên kết cộng đồng, tính tương tác, khả năng truyền tải và lưu trữ lượng thông tin khổng lồ.

a) Tính liên kết cộng đồng

Đây là đặc điểm nổi bật của mạng xã hội ảo cho phép mở rộng phạm vi kết nối giữa người với người trong không gian đa dạng. Người sử dụng có thể liên kết với nhau, việc liên kết này tạo ra một cộng đồng mạng với số lượng thành viên lớn.

b) Tính đa phương tiện

MXH có rất nhiều tiện ích nhờ sự kết hợp giữa các yếu tố chữ viết, âm thanh, hình ảnh, Sau khi đăng kí tài khoản người sử dụng có thể tạo ra một không gian riêng cho bản thân. Nhờ các tiện ích đa phương tiện mà người sử dụng có thể chia sẻ thông tin, hình ảnh, video. Đặc điểm này được phản ánh trong cấu trúc phân lớp ứng dụng của MXH.

c) Tính tương tác

Tính tương tác được thể hiện không chỉ ở thông tin được di truyền và sau đó nhận được phản hồi từ phía người nhận mà còn phụ thuộc vào cách người dùng sử dụng các ứng dụng

d) Khả năng truyền tải và lưu trữ lượng thông tin khổng lồ

Tất cả các MXH đều có những ứng dụng tương tự nhau như đăng trạng thái, nhạc, video clip, viết bài nhưng được phân bố với dung lượng khác nhau. Các trang MXH lưu trữ thông tin và nhóm sắp xếp chúng theo một thứ tự thời gian, nhờ đó người sử dụng có thể truy cập và tìm kiếm thông tin.

1.2. Lợi ích của mạng xã hội

Kể từ khi có kết nối mạng trên toàn cầu (Internet) và nhất là sau khi điện thoại thông minh hay máy tính bảng được chế tạo, việc sử dụng các MXH như Facebook, Instagram, Viber, Zalo, Skype, Whatsapp, Youtube, Linked, Twitter đã không còn xa lạ với hầu hết người dùng, kể cả trẻ em, thanh thiếu niên và người lớn tuổi. Thế giới ngày càng phát triển, mạng xã hội càng giúp con người xích lại gần nhau hơn, đem lại những tính năng và lợi ích vô cùng tuyệt vời.

a) Kết nối bạn bè, gia đình, cộng đồng

Ngày nay, con người ngày càng có ít thời gian cho bản thân và mở rộng các mối quan hệ. Nhờ có mạng xã hội, thông qua văn bản, video, hình ảnh con người có thể kết nối với nhau rất thuận tiện. Ngoài ra họ cũng có thể mở rộng các mối quan hệ khác về mọi lĩnh vực mà người dùng quan tâm. Người dùng có thể kết bạn với nhiều nhóm người với những sở thích, sở trường khác nhau. Hầu hết các mạng xã hội đều yêu cầu người dùng để đưa ra một số thông tin nhất định thường bao gồm: Độ tuổi, giới tính, địa điểm, quan điểm, sở thích... Tuy nhiên, những thông tin rất cá nhân thường không được khuyến khích vì lý do an toàn. Điều này cho phép người dùng khác tìm kiếm theo một số loại tiêu chuẩn phù hợp đối với mình và duy trì một mức độ ẩn danh tương tự như hầu hết các dịch vụ hẹn hò trực tuyến.

b) Cập nhật tin tức, kiến thức, xu thế

Tin tức sẽ được cập nhật theo từng giây, trong học tập, nghiên cứu thì đây cũng là một kênh tin tức bổ ích. Theo báo cáo của Hội Liên hiệp giáo dục Mỹ (The National School Boards Association), 60% sinh viên sử dụng mạng xã hội nói chuyện về chủ đề giáo dục trực tuyến, và hơn 50% nói chuyện cụ thể về việc học ở trường. Một số mạng xã hội khác như: TermWiki, Learn Central và các trang web khác được xây dựng để thúc đẩy mối các quan hệ trong giáo dục bao gồm các Blog giáo dục, ePortfolios cũng như thông tin liên lạc như chat, bài thảo luận, và các diễn đàn học tập.

c) Cải thiện chất lượng và tốc độ của báo chí và dịch vụ công

Do tính năng cập nhật và lan rộng nhanh của MXH mà các cơ quan báo chí và thông tin đại chúng đang tích cực đăng tải cùng một lúc trên báo giấy, trên báo điện tử và trang mạng của mình để theo kịp xu thế của thời đại và giữ số lượng độc giả của mình. Các cơ quan pháp luật hay dịch vụ công cũng đang dần “lên sóng” MXH để cập nhật những tin tức và quy định mới của mình hoặc lắng nghe ý kiến phê bình góp ý của người dân nhằm giúp giảm thiểu sự quan liêu, phức tạp hay sai sót trong dịch vụ công, để tiến tới một bộ máy hành chính công thông minh và giản tiện hơn.

d) Cải thiện kỹ năng sống, kiến thức

Hiện nay trên các mạng xã hội xuất hiện ngày càng nhiều các trang dạy ngoại ngữ, nấu ăn, sửa chữa, giao tiếp, tâm lý, thể thao để xem tham khảo, tự học mà không cần đến lớp hay đóng lệ phí. Chính nhờ tham gia các cộng đồng mạng này, chúng ta đang ngày càng trở nên hoàn thiện hơn với những kỹ năng cơ bản cần thiết trong cuộc sống hiện đại như sử dụng ngoại ngữ, cách giao tiếp văn minh hay có một thể hình khỏe đẹp.

e) Kinh doanh, quảng cáo miễn phí

Rất nhiều công ty, nhà quảng cáo đã sử dụng mạng xã hội để bán hàng, quảng cáo cho sản phẩm của mình. Ở các trang mạng xã hội như Facebook, Instagram,... có rất nhiều người trẻ khởi nghiệp bằng cách bán hàng online. Mạng xã hội kết nối con người với chi phí thấp, có những chức năng phù hợp, do đó đem lại hiệu quả cao, kể cả việc tư vấn cho người dùng cũng trở nên nhanh chóng, dễ dàng. Việc kinh doanh, quảng cáo sẽ trở nên phổ biến hơn do chức năng “chia sẻ” từ đó sẽ có nhiều hơn những người dùng đọc được thông tin về sản phẩm.

f) Tiết kiệm kinh phí, thời gian, sức lao động

Nhờ MXH mà công ty, tổ chức hay hộ gia đình đã tiết kiệm được chi phí giấy, mực in, nhân công, phí điện thoại, tin nhắn. Nếu như ngày trước chúng

ta phải di chuyển rất xa, rất mất thời gian, công sức, tiền bạc thì mới có thể truyền được thông tin, thì nay nó đã quá đơn giản, chỉ cần một cú nhấp chuột, tất cả thông tin được chuyển đi nhanh chóng và dễ dàng. Một số mạng xã hội nhằm mục đích khuyến khích lối sống lành mạnh đối với người dùng. Ví dụ như: Mạng xã hội SparkPeople cung cấp cho cộng đồng các công cụ trợ đồng đẳng trong việc giảm cân, Fitocracy tập trung vào hướng dẫn người dùng trong tập thể dục hoặc cho phép người dùng chia sẻ tập luyện của mình và nhận xét về những người dùng khác.

g) Tác động chính trị, xã hội

Nếu được sử dụng đúng cách, mạng xã hội có vai trò quan trọng đối với các chính trị gia, giúp họ được nhiều người dân biết đến hơn, giúp họ tự xây dựng hình ảnh trong mắt công chúng. Từ đó, có thể thúc đẩy sự nghiệp của họ. Ví dụ điển hình cho việc sử dụng mạng xã hội thành công là Tổng thống Mỹ Donald Trump. Gần đây, ông Donald Trump đã khẳng định: “Tôi nghĩ có thể tôi đã không ngồi ở vị trí này nếu không có Twitter. Twitter là một thứ tuyệt vời đối với tôi, vì tôi có thể truyền tải suy nghĩ của mình tới công chúng. Tôi có thể sẽ không ở đây nói chuyện với tư cách tổng thống nếu không thể đưa ra những phát biểu chân thật”.

Dù vậy, MXH cũng là “con dao hai lưỡi” nếu ta sử dụng không đúng mục đích. Sử dụng MXH quá nhiều sẽ dẫn đến xao nhãng trong học tập, mất đi thời gian vận động, thể dục thể thao. Do nguồn thông tin trên mạng không có ai giám sát, kiểm duyệt nên còn tràn lan rất nhiều thông tin sai lệch, văn hóa phẩm đồi trụy, trong khi giới trẻ còn chưa đủ nhận thức để sàng lọc thông tin, dễ dẫn đến nhận thức lệch lạc, kéo theo đó là hành động sai lầm như: giết người, nghiện hút, mại dâm... Bên cạnh đó sử dụng MXH quá nhiều còn dẫn đến mất khả năng tương tác giữa mọi người, có nguy cơ mắc bệnh trầm cảm, xao nhãng những mục tiêu thật của cuộc sống. Việc bảo mật thông tin chưa thực sự tốt, nghiêm trọng nhất là nguy cơ lây lan thông tin, đặc biệt là những thông tin sai lệch trên MXH, sẽ dẫn đến những hậu quả vô cùng nghiêm trọng.

2. Sự lây lan và tác hại của thông tin sai lệch trên mạng xã hội

Trong phần này, nhóm tác giả trình bày định nghĩa thông tin sai lệch, các tính chất của thông tin sai lệch, thực trạng phát tán của thông tin sai lệch trên thế giới và thực trạng ở Việt Nam, đi kèm theo đó là các vụ việc thực tế đã và đang xảy ra trong thời gian gần đây.

2.1. Định nghĩa thông tin sai lệch và tính chất

a) Định nghĩa thông tin sai lệch

Trong thực tế trên MXH luôn tồn tại những thông tin lệch lạc, không lành mạnh gây ra ảnh hưởng tiêu cực đến người dùng bên cạnh những giá trị tích cực mà chúng mang lại. Trong thực tế, bên cạnh các thông tin bổ ích, có giá trị đối với xã hội thì còn vô số thông tin, hình ảnh có nội dung xấu độc. Tại khoản 1, điều 5 Nghị định 72/2013/NĐ-CP ngày 15/7/2013 của Chính phủ đã có quy định chi tiết về việc quản lý, cung cấp, sử dụng dịch vụ Internet và thông tin trên mạng [7]. Trong đó có nhiều hành vi bị nghiêm cấm như lợi dụng việc cung cấp, sử dụng dịch vụ Internet và thông tin trên mạng nhằm mục đích chống lại Nhà nước Cộng hòa xã hội chủ nghĩa Việt Nam; gây phương hại đến an ninh quốc gia, trật tự an toàn xã hội; phá hoại khối đại đoàn kết dân tộc; tuyên truyền chiến tranh, khủng bố; gây hận thù, mâu thuẫn giữa các dân tộc, sắc tộc, tôn giáo (điểm a). Tuyên truyền, kích động bạo lực, dâm ô, đồi trụy, tội ác, tệ nạn xã hội, mê tín dị đoan, phá hoại thuần phong, mỹ tục của dân tộc (điểm b). Tiết lộ bí mật nhà nước, bí mật quân sự, an ninh, kinh tế, đối ngoại và những bí mật khác do pháp luật quy định (điểm c). Đưa thông tin xuyên tạc, vu khống, xúc phạm uy tín của tổ chức, danh dự và nhân phẩm của cá nhân (điểm d). Quảng cáo, tuyên truyền, mua bán hàng hóa, dịch vụ bị cấm; truyền bá tác phẩm báo chí, văn học, nghệ thuật, xuất bản phẩm bị cấm (điểm đ). Giả mạo tổ chức, cá nhân và phát tán thông tin giả mạo, thông tin sai sự thật xâm hại đến quyền và lợi ích hợp pháp của tổ chức, cá nhân (điểm e).

Theo Karlova và Fisher, 2013 [8], thông tin sai lệch được hiểu là những thông tin giả mạo, không chính xác. Dựa trên mục đích của người lan truyền,

thông tin sai lệch được phân thành hai loại:

- *Thông tin sai lệch lan truyền vô ý*: Thông tin sai lệch được tạo ra và lan truyền một cách vô ý, không có chủ đích. Mọi người có xu hướng giúp lan truyền những thông tin như vậy do niềm tin với bạn bè, người thân và ảnh hưởng của họ trên MXH.
- *Thông tin sai lệch lan truyền cố ý*: Đó là những tin đồn, tin tức giả mạo, sai lệch được tạo ra và lan truyền một cách cố ý bởi người dùng với mục đích, động cơ không trong sáng.

Như vậy, có thể thấy rằng, mặc dù có những định nghĩa khác nhau về thông tin sai lệch tuy nhiên những khái niệm có những điểm tương đồng giống nhau. Đó đều là những thông tin không đảm bảo tính chính xác hoặc thông tin giả mạo, xuyên tạc vấn đề, xuyên tạc nội dung v.v... gây ảnh hưởng xấu đến cá nhân và tổ chức, đồng thời mỗi quốc gia có những quy định riêng về những hành vi bị cấm khi đưa thông tin lên mạng và đề được cụ thể hóa trong văn bản pháp luật.

Xuất phát từ những thực tế nêu trên, nhóm tác giả nhận thấy việc ngăn chặn, hạn chế kịp thời sự lan truyền của thông tin sai lệch trên MXH là vô cùng cấp thiết nhằm giảm thiểu tối đa những thiệt hại do chúng gây ra đối với người dùng, góp phần làm trong sạch môi trường mạng, nâng cao sự tin tưởng của người dùng đối với những thông tin trên MXH.

b) Tính chất

- *Thông tin xuyên tạc có thể có tốc độ lây lan nhanh, dễ dàng*: Trong thời đại truyền thông 4.0, MXH trở thành một môi trường thích hợp để thông tin xuyên tạc lây lan nhanh chóng bởi những đặc điểm của nó. MXH là những website mở, nội dung được xây dựng hoàn toàn bởi các thành viên tham gia, hơn thế nữa, MXH có sự tham gia trực tiếp của nhiều cá nhân hay các chủ thể. Qua đó, các thông tin được dư luận và các cá nhân cung cấp có thể mập mờ, không chính xác là cơ sở cho việc phát sinh thông tin xuyên tạc, sai lệch. Hơn thế nữa, người đọc thường có xu hướng quan tâm

đến các tin đồn nhiều hơn tin chính thống vì chúng có thể gây nên nhiều liên tưởng tò mò, hấp dẫn. MXH Facebook cho phép người dung đăng tin lên cùng với các chức năng bình luận, like, share. Chính vì vậy, thông tin sai lệch thông qua những đặc điểm này có thể lan truyền với tốc độ chóng mặt và ngày càng nhiều diễn biến phức tạp.

- *Khả năng lan truyền rộng, khó kiểm soát được tính xác minh của thông tin:* Ban đầu thông tin sai lệch được đăng tải lên bởi một cá nhân hoặc tổ chức nhưng chưa được xác minh hoặc ghi nhận bởi các cơ quan thẩm quyền, thông tin này thường được xuất phát từ các nhóm nhỏ và thường mang nội dung về các vấn đề chính trị, về cá nhân tổ chức khác, chúng được thảo luận, chia sẻ qua các tính năng của mạng xã hội đến nhiều người dùng để tạo cơ sở để công chúng tin tưởng hoặc bị ảnh hưởng. Khi thông tin đó đã trở nên phổ biến, có được sự quan tâm rộng rãi của công chúng, trong quá trình truyền từ người này sang người khác, các đối tượng xấu đã bóp méo dần sự thật, thêm thắt gây ra hậu quả vô cùng to lớn.

Thế nhưng, hiện nay các mạng xã hội này vẫn chưa chính xác một cơ chế để kiểm duyệt những thông tin này có phù hợp không, người dùng cũng thiếu các thông tin xác thực để kiểm chứng, thẩm định dẫn đến hiểu nhầm, hiểu sai về bản chất của vấn đề.

2.2. Thực trạng sự phát tán của thông tin sai lệch trên mạng xã hội

Không chỉ ở Việt Nam mà sự phát tán diện rộng của thông tin trên toàn thế giới đã trở thành một vấn nạn chưa từng có. Mạng xã hội trở thành một trong những ưu tiên hàng đầu của các cơ quan hành chính cũng như tư nhân áp dụng vào nhiều mục đích khác nhau. Sử dụng công cụ MXH là một trong các cách nhanh chóng và dễ dàng để tiếp nhận những ý kiến của công chúng và để cho công chúng cập nhật hoạt động của họ. Tuy nhiên, điều này đi kèm với những nguy cơ quá lạm dụng các trang MXH.

MXH ảnh hưởng rất nhiều đến các hoạt động và các mối quan hệ trong thế giới thực. Trong đó, những tin tức về giải trí được quan tâm nhất. Khi người

dùng đọc các tin tức mà họ quan tâm, họ có nhiều khả năng sẽ duy trì thảo luận quanh thông tin đó. Ngoài ra, khi nội dung thông tin liên quan đến vấn đề chính trị, người dùng có nhiều khả năng đưa ra quan điểm, ý kiến của mình về chính trị.

a) Thực trạng thế giới

Với số người sử dụng các mạng xã hội trên toàn cầu vào khoảng 3 tỷ người và không có dấu hiệu dừng lại như hiện nay, thông tin sai lệch có sức ảnh hưởng vô cùng lớn tới tình hình thế giới theo nhiều khía cạnh. Sự lây lan thông tin độc hại ảnh hưởng mạnh đến tình hình kinh tế - chính trị, không những thế các đối tượng xấu sử dụng thông tin sai lệch để lừa đảo, chiếm đoạt ảnh hưởng đến tâm lý cũng như sức khỏe của người dùng.

Ngày 23 tháng 4 năm 2013 tin tức giả mạo hãng thông tấn Associated Press tung tin Nhà Trắng bị đánh bom và cựu Tổng thống Obama bị thương nặng bởi một vụ nổ ở Nhà trắng. Ngay lập tức thông tin này làm thị trường tài chính chứng khoán của Mỹ chao đảo. Các chỉ số chứng khoán gần như sụp đổ bởi thông tin này. Chỉ số Dow Jones ngay lập tức sụt giảm đến 143 điểm gây thiệt hại 136,5 tỷ USD cho thị trường. Mặc dù vậy, thị trường chỉ rơi vào khoảng lặng hơn 1 phút trước khi AP thông báo đó là tin giả mạo do tài khoản Twitter của báo này bị tin tức chiếm quyền điều khiển và đăng tin sai sự thật [9].

Gần đây những thông tin sai lệch trên các MXH còn được cho là có ảnh hưởng không nhỏ tới cuộc bầu cử ở Pháp và ở Mỹ. Trong cuộc bầu cử tổng thống ở Pháp Facebook đã xóa 30.000 tài khoản giả mạo báo cáo tin đồn ở Pháp trước cuộc bầu cử Tổng thống vào năm 2017 [10]. Trong cuộc bầu cử ở Mỹ MXH được cho ảnh hưởng không nhỏ tới kết quả bầu cử Tổng thống năm 2016. Nhiều tài khoản giả mạo được tạo ra chia sẻ những thông tin sai lệch về sơ hở rở email của bà Hilary Clinton và các đồng sự cấp cao của bà. Các tài khoản này đã phát động một chiến dịch phản đối bà Hilary đây được cho là một trong những nguyên nhân lớn đưa đến sự thất bại của bà.

b) Thực trạng tại Việt Nam

MXH trong những năm gần đây ngày càng trở nên thịnh hành tại Việt Nam. Cũng như trên thế giới, MXH (điển hình như Facebook) được nhiều người Việt Nam coi là tin tưởng sử dụng. Chính vì vậy, các đối tượng luôn tìm cách lợi dụng điều này để có thể trục lợi cho bản thân, hay thực hiện những hoạt động chống phá, gây ảnh hưởng nghiêm trọng tới uy tín của cá nhân, tập thể, chính quyền.

Vào tháng 8 năm 2014, trên MXH lan truyền nhanh chóng nội dung tin đồn thất thiệt rằng “dịch Ebola đã bùng phát tại Hà Nội”. Trước đó, nhiều người đã truyền nhau thông tin tại Hà Nội đã có người nhiễm Ebola. Thông tin ngay sau đó đã được lan truyền một cách chóng mặt, gây nên một sự hoang mang lo lắng đến người dân. Bên cạnh đó còn có tin đồn cho rằng, nước láng giềng của Việt Nam là Campuchia đã xuất hiện ca nhiễm virus Ebola [11].

Tình trạng “ô nhiễm” thông tin xuất phát từ hoạt động phá hoại tư tưởng đồng thời phát tán các tư tưởng chống phá Đảng và nhà nước kích động biểu tình bạo loạn thông qua MXH của các thế lực thù địch, phản động chống đối ở trong và ngoài nước. Theo thống kê của cơ quan An ninh, tính đến nay có hơn 2500 trang web, blog, MXH của các cá nhân, tổ chức đang hoạt động, đăng tải các tin, bài viết, bình luận, bài phỏng vấn với mục đích tuyên truyền phá hoại tư tưởng nhằm phá hoại tư tưởng. Trong đó nổi lên một số trang như: danlambao, quanlambao, danluan. Các đối tượng quản trị những trang web này lợi dụng những điểm nóng về chính trị, xã hội và những thiếu sót trong công tác quản lý của ta để đăng tin xuyên tạc, kích động biểu tình, bạo loạn. Điển hình là hai vụ việc sự cố môi trường ở 4 tỉnh miền Trung do Formosa xả thải [12] và vụ việc khiếu kiện đất đai ở Đồng Tâm mà đỉnh điểm là vụ bắt giữa 38 chiến sĩ Cảnh sát cơ động. Những thông tin sai sự thật này gây phức tạp thêm tình hình gây khó khăn cho công tác giải quyết của Chính quyền Nhà nước.

MXH nơi để mọi người cùng nhau chia sẻ mọi điều trong cuộc sống và đồng thời cũng là nơi để mọi người cập nhật những tin tức, những hình ảnh mới, kết nối với nhau. Nhưng cũng chính từ đây, các chuyện hư cấu, tin đồn, chuyện bóp méo được đăng tải để “câu like”, gây ảnh hưởng đến cộng đồng, dư luận.

Để dễ dàng thu hút sự tham gia từ cư dân mạng, không ít các trang fanpage đã sử dụng các chiêu trò câu “like” bằng cách đăng các mẩu tin tức có nội dung đồi trụy; hoặc các ảnh về những trường hợp đau lòng, lợi dụng lòng thương của cư dân mạng để kêu gọi hành động chia sẻ thông điệp.

Những chiêu thức, trò đùa ác ý với nhiều mục đích khác nhau đã làm cho nhiều người hoang mang và tỏ ra e ngại khi tiếp nhận các thông tin trên MXH. Và không chỉ làm hoang mang dư luận, những tin đồn thất thiệt này đã trực tiếp làm ảnh hưởng đến cuộc sống của những người vô tình trở thành nạn nhân. Chỉ bằng một cú click vào xem, chia sẻ, người dùng có thể rơi vào cái bẫy khiến tài khoản cá nhân bị nguy hiểm, bị đánh cắp thông tin. Trong những đường dẫn chứa thông tin sai lệch này còn có thể kèm virus hoặc những phần mềm gián điệp nhằm lấy cắp thông tin hay chiếm quyền kiểm soát máy.

Kết luận chương

Sự ảnh hưởng rộng lớn của thông tin sai lệch đến với người dùng cũng như các công ty, doanh nghiệp ngày càng tăng mạnh nên việc tổ chức hạn chế sự ảnh hưởng của các loại thông tin này đang trở nên cấp bách hơn bao giờ hết.

Hậu quả của thông tin sai lệch trên MXH là vô cùng nghiêm trọng về mọi mặt chính trị, kinh tế, xã hội. Bản thân của cá nhân, tổ chức bị tung tin sai sự thật phải gánh chịu hậu quả, phiền toái không đáng có, thậm chí là những thiệt hại nặng nề về kinh tế, danh dự, phẩm chất. Thông tin sai lệch có thể gây ảnh hưởng nặng đến kinh tế của người dân, điển hình là phóng sự của VTV về nghề xuất khẩu lao động điều dưỡng viên tại đức có lương tháng 100 triệu [13]. Thông tin sai lệch này khiến người dùng bị nhầm lẫn dẫn đến đầu tư tiền vào sai chỗ và gây tâm lý chán nản với người đi xuất khẩu lao động. Nguy hại hơn những thông tin sai lệch về chính trị, đường lối chính sách của Đảng và Nhà nước còn gây mất lòng tin của nhân dân vào bộ máy. Những tin này là “mồi dẫn” để các thế lực thù địch tập hợp, lôi kéo lực lượng trên không gian mạng, tổ chức các cuộc biểu tình, bạo loạn, gây mất an ninh trật tự. Ở khía cạnh khác đường link chia sẻ các loại tin sốc, bịa đặt được các hacker sử dụng để phát tán mã độc là

bàn đạp cho các cuộc tấn công APT (Tấn công mạng sử dụng công nghệ cao), lừa đảo trên không gian mạng.

Trước những thực trạng to lớn hiện nay, nhóm đã mạnh dạn nghiên cứu, thực nghiệm để có thể đưa ra một giải pháp tối ưu để có thể ngăn chặn tác hại của lan truyền thông tin sai lệch trên mạng xã hội được trình bày ở chương sau.

CHƯƠNG 2

CƠ CHẾ LAN TRUYỀN THÔNG TIN SAI LỆCH

Trong thực tế trên MXH luôn tồn tại những thông tin lệch lạc, không lành mạnh gây ra ảnh hưởng tiêu cực đến người dùng trên MXH. Chương 1 đã trình bày tổng quan về các tác hại mà thông tin sai lệch gây ra đối với người dùng MXH. Chúng ta có thể tóm gọn lại những hậu quả vô cùng to lớn gây ra bởi thông tin sai lệch bằng những nội dung sau: Đối với những vấn đề mang tính xã hội, những thông tin sai lệch ảnh hưởng tiêu cực đến tâm lý, đời sống tinh thần của người dùng khi chúng được phát tán trên mạng. Nó có thể ảnh hưởng đến tinh thần, thái độ, thậm chí cả kinh tế của khu vực người dùng sinh sống. Trong hoạt động kinh doanh, những thông tin sai lệch tiêu cực về sản phẩm của một doanh nghiệp ảnh hưởng xấu đến tài chính, giá bán, doanh thu, và thậm chí là thương hiệu của doanh nghiệp đó. Đối với từng cá nhân, những thông tin sai lệch về họ có thể ảnh hưởng rất xấu, làm đảo lộn cuộc sống của họ.

Những tác hại kể trên cho thấy việc đối phó với các thông tin sai lệch là vô cùng cấp bách. Việc phát hiện nguồn thông tin sai lệch là cơ sở cho các giải pháp ngăn chặn sự phát tán của chúng. Nguồn phát tán thông tin sai lệch có thể được phát hiện thông qua khảo sát người dùng hoặc các phương pháp khai phá dữ liệu. Trong việc giám sát các thông tin sai lệch H.Zhang [14] đã đề xuất giải pháp tìm số node giám sát nhỏ nhất sao cho có thể phát hiện thông tin sai lệch với tỉ lệ τ .

Để có thể đưa ra giải pháp hiệu quả trong việc ngăn chặn sự lan truyền của thông tin sai lệch, trước tiên chúng ta phải hiểu được cơ chế thông tin sai lệch lan truyền trên MXH. Chương này phân tích quá trình lan truyền thông tin sai lệch dưới hai mô hình lan truyền: Mô hình bậc độc lập và mô hình ngưỡng tuyến tính, đây là hai mô hình phổ biến, được sử dụng rộng rãi trong các công trình nghiên cứu liên quan đến vấn đề lan truyền thông tin, lan truyền ảnh hưởng trên MXH. Đồng thời trong chương này cũng trình bày tổng quan một số hướng nghiên cứu liên quan đến bài toán hạn chế lan truyền thông tin sai lệch trên

mạng xã hội trực tuyến.

1. Mô hình lan truyền thông tin

Thông tin được phát tán trên các MXH từ người dùng này đến người dùng khác thông qua nhiều hoạt động đăng bài, chia sẻ, bình luận. Kempe [15] là người đầu tiên đưa ra các mô hình phát tán thông tin, trong đó đã đưa ra hai mô hình phát tán thông tin cơ bản là Mô hình tầng độc lập (Independent Cascade – IC) và Mô hình ngưỡng tuyến tính (Linear Threshold – LT). Hai mô hình này sau đó đã được sử dụng rộng rãi trong các bài toán liên quan đến lan truyền, phát tán thông tin.

Các mô hình phát tán thông tin là cơ sở cho việc nghiên cứu hạn chế thông tin sai lệch cũng như các tin đồn thất thiệt. Các nghiên cứu về chủ đề này những năm gần đây đều dựa trên hai mô hình IC, LT và các biến thể của chúng [16], [17], [18]. Theo đó, một mạng xã hội được biểu diễn bởi các thành phần như sau:

- V là tập hợp các đỉnh của đồ thị, $|V| = n$, biểu diễn những người dùng tồn tại trong MXH.
- $E \subset V \times V$ hợp các cạnh của đồ thị, $|E| = m$ gồm m cạnh có hướng, biểu diễn mối quan hệ giữa các cá nhân trong MXH.
- $w(u,v)$ là trọng số của cạnh (u, v) là một số thực dương biểu diễn cho các tần số tương tác, trao đổi giữa hai người dùng. $w(u, v) = 0$ nếu giữa hai đỉnh u và v không tồn tại cạnh, $w(u, v) > 0$ nếu giữa u và v tồn tại cạnh nối.

Do G là đồ thị có hướng nên cạnh (u,v) được gọi là cạnh đi ra từ u , cạnh (v,u) được gọi là cạnh đi vào đỉnh u . Ta ký hiệu $N_{out}(u)$ và $N_{in}(u)$ tương ứng là tập hợp các đỉnh hàng xóm đi ra và đi vào đỉnh u .

Quá trình lan truyền thông tin theo các bước thời gian rời rạc, với thời gian $t = 0, 1, 2, \dots$. Gọi $S_t \subset V$ là tập các đỉnh ở trạng thái *kích hoạt* tại thời điểm t . Tập các đỉnh là nguồn phát tán thông tin sai lệch ban đầu hay còn gọi là tập hạt

giống, kí hiệu là S_0

Tại mỗi bước thời gian t , đỉnh u ở trạng thái kích hoạt nếu $u \in S_0$ hoặc u nhận được thông tin sai lệch từ các đỉnh hàng xóm ở trạng thái kích hoạt và chấp nhận thông tin này để tiếp tục chia sẻ, phát tán những thông tin sai lệch đó đến những đỉnh khác trong các bước tiếp theo, quá trình kích hoạt này ở mỗi mô hình lan truyền là khác nhau, ngược lại u ở trạng thái *không kích hoạt*.

Hiện nay có nhiều mô hình lan truyền thông tin đang được nghiên cứu và sử dụng, tiêu biểu trong số đó là: Mô hình ngưỡng (Threshold Model) [16], mô hình tầng (Cascading Model) [17], mô hình lan truyền ảnh hưởng cạnh tranh (Competitive Influence Diffusion Model) [18], mô hình dịch bệnh (Epidemic Model) [21]. Trong nội dung của đề tài, nhóm tác giả sử dụng mô hình ngưỡng tuyến tính (Linear Threshold – LT) và mô hình tầng độc lập (Independent Cascade – IC) [15] để mô tả quá trình lan truyền thông tin sai lệch trên MXH.

Trong nghiên cứu này, các tác giả đã tìm hiểu các mô hình trên qua đó có thể làm rõ tính chất về hành vi lan truyền thông tin của người dùng và chọn ra mô hình phù hợp để áp dụng và cũng đưa ra các mô hình phù hợp hơn với thực tiễn. Sau đây là sự mô tả chi tiết mô hình tiêu biểu được nhiều nghiên cứu sử dụng đó là: Mô hình tầng độc lập (IC) và Mô hình ngưỡng tuyến tính (LT).

1.1. Mô hình tầng độc lập

Mô hình tầng độc lập (Independent cascade – IC) được đề xuất bởi Kempe [15] dựa trên các mô hình tương tác trong các hệ thống hạt và nghiên cứu về tiếp thị. Mô hình IC có liên quan tới mô hình dịch bệnh (Epidemic models). Đặc trưng chính của mô hình IC là quá trình lan truyền thông tin dọc theo các cạnh của đồ thị một cách độc lập với nhau.

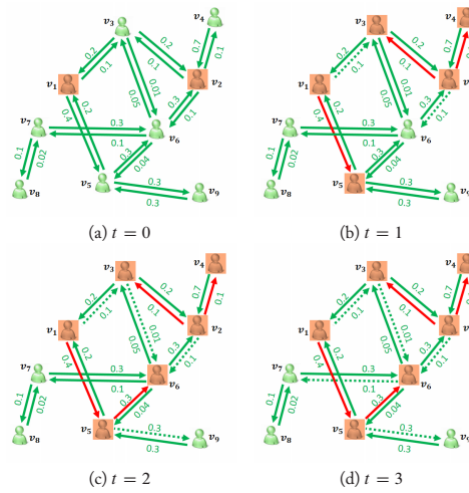
Trong mô hình IC, mỗi cạnh $(u,v) \in E$ được gán một xác suất ảnh hưởng (Influence Probability) $\rho_{uv} \in [0,1]$ biểu diễn mức độ ảnh hưởng của đỉnh u với đỉnh v . Nếu $(u,v) \notin E$, thì $\rho_{uv} = 0$.

Quá trình lan truyền thông tin trên mô hình IC diễn ra theo bước thời gian rời rạc, tạo ra tập các đỉnh *kích hoạt* theo quy tắc sau:

- Tại thời điểm $t = 0$, tập đỉnh ở trạng thái kích hoạt chính là tập nguồn phát thông tin sai lệch S_0
- Tại thời điểm $t = 1$, đầu tiên ta gán S_t bằng S_{t-1} sau đó với mỗi nút $v \notin S_{t-1}$, và với mỗi nút $u \in N_{in}(v) \cap (S_{t-1} \setminus S_{t-2})$, u thực hiện một lần thử kích hoạt bằng cách áp dụng phép thử Bernoulli (Phép tung đồng xu độc lập) với xác suất thành công là $p(u, v)$. Nếu thành công ta thêm v vào tập S_t và nói rằng u kích hoạt v tại thời điểm t . Nếu nhiều nút kích hoạt v tại thời điểm t , kết quả tương tự xảy ra, v được thêm vào tập S_t . Nói cách khác, sau khi nút u được kích hoạt tại thời điểm $t-1$, ngay lập tức trong thời điểm t , u có **một cơ hội duy nhất** để kích hoạt các hàng xóm chưa được kích hoạt v của nó với xác suất $p(u, v)$, và những sự kích hoạt này là độc lập với nhau. Nếu nút u không kích hoạt v tại thời điểm t , nó sẽ không thử kích hoạt v tại các thời điểm sau nữa. Và một khi một nút đã được kích hoạt, nó vẫn giữ trạng thái đó ở các bước sau.
- Nếu tại thời điểm t , không có nút nào được kích hoạt thêm nữa, nghĩa là $S_t = S_{t-1}$, tập các nút *kích hoạt* sẽ không còn thay đổi nữa, và quá trình truyền tin kết thúc với tập các nút bị kích hoạt cuối cùng là S_t

Hình 2.1 chỉ ra một ví dụ của quá trình lan truyền thông tin trên mô hình IC. Các đỉnh màu da cam và màu xanh tương ứng biểu diễn các đỉnh ở trạng thái kích hoạt, và không kích hoạt. Cạnh liền màu đỏ từ u đến v biểu diễn u kích hoạt thành công v , cạnh nét đứt màu xanh từ u đến v biểu diễn u kích hoạt không thành công v .

Tại thời điểm bắt đầu $t=0$, hai đỉnh v_1, v_2 ở trạng thái *kích hoạt*. Ở thời điểm $t = 1$, v_1 kích hoạt thành công v_5 nhưng thất bại với v_3 , trong khi đó v_2 kích hoạt thành công v_3 và v_4 nhưng thất bại với v_6 . Tại thời điểm $t = 2$, v_3 kích hoạt thất bại v_6 trong khi v_5 kích hoạt thành công v_6 nhưng thất bại với v_9 . Ở bước $t = 3$, v_6 kích hoạt thất bại v_7 , đến đây quá trình lan truyền thông tin kết thúc do không có đỉnh nào được kích hoạt thêm.



Hình 2.1 Một số ví dụ quá trình lan truyền thông tin trên mô hình IC

Mô hình IC phù hợp cho quá trình lan truyền thông tin hoặc virus, đó là các môi trường mà việc tiếp xúc với một nguồn là đủ để một cá nhân được kích hoạt.

1.2. Mô hình ngưỡng tuyến tính

Mô hình IC phù hợp để mô tả sự lan truyền thông tin đơn giản, ở đó một đỉnh có thể được kích hoạt từ một đỉnh duy nhất. Tuy nhiên trong thực tế có nhiều trường hợp một cá nhân cần nhiều sự tác động của các cá nhân khác để thay đổi hành vi của mình. Có thể kể đến các trường hợp như khi người dùng tiếp nhận một thông tin mới, một công nghệ mới, hay một thông tin sai lệch bởi xấu danh dự của các đồng chí lãnh đạo Đảng và Nhà nước, người dùng MXH cần được củng cố tích cực từ nhiều nguồn độc lập trong số bạn bè và người quen của họ trước khi họ thay đổi suy nghĩ của mình, chấp nhận thông tin đó.

Các nhà khoa học đã đề xuất khái niệm *hành vi ngưỡng* để mô tả các hành vi kiểu như trên. Khi một hàm tổng hợp của các người dùng đã kích hoạt trên mạng đạt đến một ngưỡng nhất định thì đối tượng sẽ được kích hoạt, xét đến hành vi ngưỡng mà mỗi cá nhân chỉ được kích hoạt khi tiếp nhận ảnh hưởng từ nhiều hơn hai nguồn thông tin.

Mô hình ngưỡng tuyến tính (Linear Threshold – LT) là mô hình khuếch tán ngẫu nhiên được đề xuất bởi Kempe [15]. Trong mô hình LT, mỗi cạnh (u,v)

$\in [0,1]$ biểu diễn mức độ ảnh hưởng của đỉnh u đến đỉnh v . Nếu $(u,v) \notin E$ thì $w(u,v)=0$. Các trọng số này được chuẩn hóa sao cho với mỗi đỉnh v , tổng trọng số tất cả các cạnh đi đến đỉnh v lớn nhất bằng 1, tức là: $\sum_u w(u,v) \leq 1$

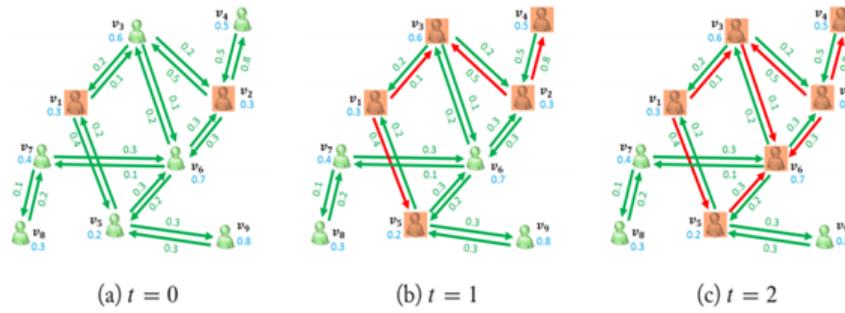
Tùy vào đặc tính của từng người dùng tương ứng, mỗi đỉnh $v \in V$ có một giá trị $\theta_v \in [0,1]$, biểu diễn ngưỡng đỉnh v bị ảnh hưởng bởi các đỉnh kích hoạt hàng xóm. Quá trình lan truyền thông tin trên mô hình LT diễn ra theo bước thời gian rời rạc, tạo ra tập các đỉnh kích hoạt theo quy tắc sau:

- Tại thời điểm $t = 0$, tập đỉnh ở trạng thái kích hoạt chính là tập nguồn phát thông tin sai lệch S_0 .
- Tại thời điểm $t + 1$, đầu tiên ta gán S_t bằng S_{t-1} . Sau đó với mỗi đỉnh chưa được kích hoạt $v \in V \setminus S_{t-1}$, nếu tổng ảnh hưởng từ những đỉnh hàng xóm kích hoạt tới v vượt ngưỡng θ_v , tức là
- Nếu tại thời điểm t , không có nút nào được kích hoạt thêm nữa, nghĩa là $S_t = S_{t-1}$, tập các nút kích hoạt sẽ không còn thay đổi nữa, và quá trình truyền tin kết thúc với tập các nút bị kích hoạt cuối cùng là S_t .

Sự ngẫu nhiên trong việc lựa chọn ngưỡng θ_0 từ 0 đến 1 phản ánh sự thiếu thông tin về ngưỡng nội bộ của mỗi cá nhân. Điều này phản ánh khá đúng với thực tế xã hội, bởi vì sự chấp nhận thông tin của mỗi người, tại những thời điểm khác nhau là khác nhau, và rất khó để nắm bắt.

Hình 2.2 chỉ ra một ví dụ quá trình lan truyền thông tin trên mô hình LT. Các đỉnh màu da cam và màu xanh tương ứng biểu diễn các đỉnh ở trạng thái *kích hoạt*, và *không kích hoạt*. Các cạnh liền màu đỏ cùng đến đỉnh v biểu diễn các cạnh này đồng thời cổ vũ kích hoạt đỉnh v và thành công.

Tại thời điểm $t=0$, tất cả các đỉnh được khởi tạo ngẫu nhiên ngưỡng $\theta_0 \in [0,1]$, hai đỉnh v_1 và v_2 là các đỉnh hạt giống. Ở thời điểm $t = 1$, v_1 và v_2 kích hoạt thành công v_3 , v_1 cũng phải kích hoạt thành công v_5 và v_2 kích hoạt thành công v_4 ; tuy nhiên v_6 lại không kích hoạt thành công vì tổng trọng số các cạnh đi đến



Hình 2.2 Ví dụ quá trình lan truyền trên mô hình LT

v_6 là 0.3, trong khi ngưỡng kích hoạt của v_6 0.7. Tại thời điểm $t = 2$, các đỉnh hàng xóm đi đến v_6 là v_2, v_3, v_5 đã được kích hoạt cho nên tổng trọng số các cạnh đi đến là 0.7 đủ để kích hoạt v_6 . Tại bước $t = 3$, quá trình lan truyền thông tin kết thúc do không có đỉnh nào được kích hoạt thêm.

2. Một số hướng nghiên cứu liên quan bài toán hạn chế lan truyền thông tin sai lệch trên mạng xã hội

Tối ưu hóa ảnh hưởng của các đối tượng trên MXH là bài toán liên quan đến lan truyền thông tin trên MXH được nghiên cứu lần đầu tiên bởi Domingos và Richardson, 2001 [20]. Sau khi Kempe, 2003 [15] lần đầu tiên xây dựng vấn đề tối ưu hóa ảnh hưởng trên MXH theo cách tối ưu hóa rời rạc, có rất nhiều nghiên cứu liên quan tập trung giải quyết bài toán tối đa hóa ảnh hưởng của thông tin trên mạng xã hội như nghiên cứu của Leskovec, 2007 [21], Goyal, 2011 [22], Wei Chen, 2014 [23].

Bên cạnh vấn đề lan truyền thông tin, lan truyền ảnh hưởng cũng có rất nhiều nghiên cứu tập trung giải quyết bài toán hạn chế thông tin sai lệch lan truyền trên các MXH trực tuyến. Một số nghiên cứu tập trung vào việc nhận dạng thông tin sai lệch và tin đồn (Rumor) dựa trên đặc trưng ngôn ngữ, cấu trúc, thời gian như nghiên cứu của Qazvinian, 2011, [24] và Kwon, 2013, [25].

Một số khác, nghiên cứu vấn đề xác định tập đỉnh là nguồn phát thông tin sai lệch ban đầu. Chẳng hạn, Dung T. Nguyen và các cộng sự, 2012, [26] đã nghiên cứu bài toán xác định k nguồn phát tán thông tin sai lệch khả nghi nhất từ tập người dùng bị kích hoạt bởi thông tin sai lệch cho trước và chứng minh

bài toán thuộc lớp NP-khó xét trên mô hình lan truyền IC, đồng thời nhóm tác giả đã đề xuất hai thuật toán dựa trên cách tiếp cận xếp hạng (Ranking) và cách tiếp cận xấp xỉ đạt tỉ lệ tối ưu $1 - \frac{1}{e} - \varepsilon$

Trong việc hạn chế phát tán thông tin sai lệch, Zhang [14] đề xuất bài toán tìm số nút nhỏ nhất trong khoảng cách δ (vùng N ()) đối với nguồn phát thông tin sai lệch để vô hiệu hóa sao cho thông tin sai lệch đến nút đích r được giới hạn. Bên cạnh đó, một số tác giả đề xuất giải pháp hạn chế sự lan truyền thông tin sai lệch trên mạng xã hội bằng cách chọn ra một số đỉnh ban đầu để tiêm thông tin tốt, từ đó lan truyền những thông tin này trên cùng mạng nhằm thuyết phục những người dùng khác tin theo, trong đó sử dụng các mô hình lan truyền thông tin khác nhau [27], [28].

Trong việc khử nhiễm đối với nguồn tin sai lệch, Nguyen [27] đề xuất bài toán tìm tập người dùng hạt giống sao cho tỷ lệ khử nhiễm sau thời gian T đối với nguồn thông tin sai lệch I trong mạng là $\beta \in (0,1)$. Tức là khử nhiễm đối với các nguồn phát này sao cho số người dùng được khử nhiễm là $\beta \cdot |V|$. Ceren [29] đưa ra bài toán phản bác lại thông tin sai lệch bằng cách chọn k người dùng để thuyết phục họ nhận thức được các thông tin để phản bác lại, triệt tiêu các thông tin sai lệch. Trong nghiên cứu này, nhóm tác giả đã xây dựng mô hình Oblivious Independent Campaign, họ chứng minh đây là bài toán NP-Khó và hàm mục tiêu là hàm đơn điệu tăng và submodular.

Liên quan gần nhất đến vấn đề nghiên cứu trong đề tài của nhóm tác giả là công trình nghiên cứu của H. Zhang, 2016 [30]. Trong nghiên cứu của mình, H. Zhang đề xuất hai bài toán: Bài toán phát hiện thông tin sai lệch yêu cầu xác định k vị trí đặt giám sát (Monitor) trên MXH sao cho cực đại hóa xác suất phát hiện thông tin sai lệch và Bài toán đặt giám sát yêu cầu tìm ra tập đỉnh có kích thước nhỏ nhất để đặt giám sát sao cho xác suất thông tin sai lệch kích hoạt thành công đỉnh r nhỏ hơn một ngưỡng cho trước.

CHƯƠNG 3

GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

1. Đặt vấn đề, định nghĩa bài toán

Tự do ngôn luận, tự do báo chí là một trong những quyền căn bản của công dân đã được Hiến pháp ghi nhận và đã được Đảng và Nhà nước nhất quán. Tuy nhiên thực tế hiện nay có rất nhiều cá nhân, tổ chức đã và đang lợi dụng các quyền này để xâm phạm lợi ích của Nhà nước và lợi ích chính đáng của công dân. Với sự phát triển mạnh mẽ của Internet và mạng xã hội trực tuyến, việc đăng tải, tuyên truyền các thông tin sai lệch mang tính chất xuyên tạc, chống đối đường lối chính sách của Đảng, pháp luật của Nhà nước, hoạt động của bộ máy chính quyền diễn ra thường xuyên hơn với âm mưu, quy mô và tổ chức ngày càng rộng lớn, chặt chẽ.

Như thường lệ, mỗi khi cơ quan chức năng Việt Nam tiến hành truy tố, xét xử những cá nhân có hành vi vi phạm pháp luật như: Hoạt động nhằm lật đổ chính quyền nhân dân (theo Điều 109 Bộ luật Hình sự năm 2015, sửa đổi bổ sung năm 2017); tuyên truyền chống Nhà nước Cộng hòa XHCN Việt Nam (Điều 117) hay lợi dụng các quyền tự do, dân chủ, xâm phạm lợi ích của Nhà nước, quyền, lợi ích hợp pháp của tổ chức, công dân (Điều 331) thì ngay lập tức một số tổ chức, cá nhân chống đối trong và ngoài nước lại ráo riết tuyên truyền xuyên tạc, chống phá Việt Nam trên lĩnh vực dân chủ, nhân quyền. Những đối tượng này đã lợi dụng quyền tự do ngôn luận để viết và đăng tải các nội dung sai lệch, không có căn cứ, tuyên truyền xuyên tạc đường lối, chính sách của Đảng, pháp luật của Nhà nước, bôi nhọ các cán bộ, Đảng viên làm ảnh hưởng đến uy tín của cơ quan, tổ chức, đi ngược lại với lợi ích quốc gia, dân tộc.

Xuất phát từ những thực tế nêu trên, nhóm tác giả nhận thấy việc nghiên cứu đề ra giải pháp ngăn chặn kịp thời sự lan truyền của thông tin sai lệch trên MXH là một thách thức lớn cần giải quyết kịp thời. Tuy nhiên trong quá trình nghiên cứu, nhóm tác giả đã gặp phải một số vấn đề khó khăn.

Theo Điều 22, Luật An ninh Quốc gia nước Cộng hòa xã hội chủ nghĩa Việt Nam năm 2004, Công An nhân dân mà một trong các cơ quan chuyên trách được giao thực hiện nhiệm vụ bảo vệ An ninh quốc gia. Theo Điều 10, Luật Viễn thông năm 2009 quy định Cơ quan quản lý chuyên ngành về viễn thông là cơ quan thuộc Bộ Thông tin và Truyền thông. Tuy nhiên trong thực tế, các cơ quan chức năng của Nhà nước ta chưa có quyền quản lý đối với các nội dung trên mạng xã hội, được quản lý bởi các tập đoàn Facebook, Twitter, Microsoft. Do đó, việc xóa bỏ các nội dung, thông tin sai lệch lan truyền trên mạng xã hội, hay vô hiệu hóa người dùng đăng tải thông tin sai lệch là không khả thi.

Một vấn đề đặt ra khác là, các thông tin sai lệch phán tán trên mạng xã hội xuất phát từ các nguồn phát tán với tốc độ nhanh, quy mô rộng lớn. Đã có nhiều nghiên cứu đưa ra nhằm xác định các nguồn phát tán này, tuy nhiên hiệu quả và độ chính xác vẫn còn nhiều hạn chế. Trong thực tế, các nguồn phát tán thông tin sai lệch, các thông tin chống đối và xuyên tạc chính sách, đường lối của Đảng, pháp luật của Nhà nước thường là các đối tượng chủ mưu, cầm đầu, nắm vai trò quan trọng trong các tổ chức phản động. Đây là những đối tượng cơ hội chính trị, đối tượng bất mãn, đối tượng bảo thủ, ngoan cố và rất khó để thuyết phục các đối tượng đó gỡ bỏ các bài đăng, thông tin sai lệch trên mạng xã hội.

Trong nghiên cứu này, nhóm tác giả đã dùng phương pháp giám sát, điều tra để tìm ra những đối tượng có tính chất trên, coi họ là nguồn phát tán thông tin sai lệch. Sau đó, nhóm tác giả đã nghiên cứu và đề xuất giải pháp chọn ra tập người dùng để vô hiệu hóa họ trong việc tham gia vào quá trình phát tán thông tin sai lệch. Trên mạng xã hội, hoạt động này tương đương với các biện pháp như sau:

- Thuyết phục người dùng không chấp nhận và lan truyền những thông tin sai lệch đến những người dùng khác, tức là làm cho người dùng này không tham gia vào quá trình phát tán thông tin sai lệch. Kỹ thuật này, ở trong các nghiên cứu gần đây còn được gọi là đặt giám sát (Monitor) người dùng [31], [32], [33].

- Bằng sự can thiệp và giúp đỡ của nhà mạng có thể loại bỏ, cách ly thông tin sai lệch đến với những người dùng này. Năm 2016, Twitter đã loại bỏ 125.000 tài khoản được nghi ngờ có liên quan đến các hoạt động khủng bố [34]. Năm 2017, Facebook đã xóa 30.000 tài khoản giả mạo lan truyền các tin đồn trước cuộc bầu cử tổng thống Pháp năm 2017 [35].

Dưới những điều kiện khác nhau, như tiềm lực tài chính, khả năng thuyết phục, phạm vi lan truyền thông tin khác nhau, cùng với đó là các mục tiêu khác nhau, có thể kể đến ảnh hưởng lan truyền của thông tin sai lệch, tỉ lệ người dùng nhiễm thông tin sai lệch và với các mô hình lan truyền thông tin khác nhau. Nhóm tác giả đã tiến hành thu thập dữ liệu trực tiếp từ mạng xã hội Facebook, với phạm vi xung quanh các đối tượng chống đối, đối tượng cầm đầu trong thời gian gần đây, sau đó mô hình hóa các dữ liệu thu được, và nghiên cứu lý thuyết, đề xuất hai giải pháp ngăn chặn hiệu quả, tiến hành thực nghiệm và so sánh với các giải pháp đã công bố trước đây.

Kết quả cuối cùng của giải pháp đưa ra một danh sách người dùng tiềm năng, theo đánh giá về lý thuyết và thực nghiệm cho thấy, nếu dùng các biện pháp ngăn chặn nêu trên để vô hiệu hóa những người dùng này, ảnh hưởng của những thông tin sai lệch đối với mạng xã hội là nhỏ hơn những giải pháp đã được công bố trước đây, và thời gian thực hiện các giải pháp của nhóm tác giả là tương đối nhanh.

2. Tổng thể về giải pháp

Giải pháp ngăn chặn thông tin sai lệch trên MXH được nhóm tác giả thực hiện qua 4 bước. Đầu tiên là quá trình giám sát, điều tra để thu thập được tài khoản mạng xã hội của những đối tượng chủ mưu, cầm đầu các tổ chức phản động, các đối tượng có uy tín lớn. Sau đó nhóm tác giả sử dụng công cụ kiểm thử Selenium để thu thập mạng xã hội xung quanh các đối tượng này. Bước tiếp theo, nhóm tác giả mô hình hóa toán học mạng xã hội thực thu được ở bước trên, và cuối cùng đề xuất, áp dụng hai giải pháp hiệu quả, để đưa ra tập người dùng tiềm năng có thể áp dụng các biện pháp vô hiệu hóa để ảnh hưởng của thông tin

sai lệch trên mạng là nhỏ nhất.

MXH cho phép các nhà phát triển (Developers) có thể lấy được thông tin dữ liệu, tuy nhiên điều đó phải được sự cho phép của người dùng thông qua các mã truy cập Access Token (Access Token là một đoạn mã do MXH sinh ra ngẫu nhiên khi người dùng đồng ý cho ứng dụng thứ 3 thực hiện các thao tác đối với tài khoản của người dùng). Điều này là một vấn đề khó khăn khi chúng ta cần khảo sát, lấy thông tin đối với một số lượng lớn người dùng nhằm xây dựng được đồ thị quan hệ giữa các người dùng trên MXH.

Nhằm giải quyết vấn đề trên, nhóm tác giả đề xuất một hướng tiếp cận mới là sử dụng công cụ kiểm thử phần mềm tự động mã nguồn mở Selenium cho việc kiểm thử ứng dụng Web. Cụ thể ở đây nhóm tác giả thư viện mã nguồn mở của Selenium trên ngôn ngữ Python, kiểm thử tự động với trình duyệt Chromium.

2.1. Công cụ kiểm thử Selenium

Selenium là một công cụ kiểm thử phần mềm tự động, được phát triển bởi ThoughtWorks từ năm 2004 với tên ban đầu là JavaScriptTestRunner. Đến năm 2007, tác giả Jason Huggins rời ThoughtWorks và gia nhập Selenium Team, một phần của Google và phát triển thành Selenium như hiện nay.

Selenium cung cấp công cụ phát lại (Playback) cho việc tạo ra các bài kiểm thử mà không cần phải học một ngôn ngữ kịch bản khác (Selenium IDE). Nó cũng cung cấp một ngôn ngữ thử nghiệm tên miền cụ thể (Selenese) để tạo ra các bài kiểm thử bằng một số ngôn ngữ lập trình phổ biến bao gồm C#, Groovy, Java, Perl, Python, PHP, Ruby, Scala. Các thử nghiệm sau đó có thể chạy trên hầu hết các trình duyệt Web hiện đại. Selenium triển khai trên các nền tảng Windows, Linux và cả MacOS. Selenium là phần mềm mã nguồn mở, được phát hành theo giấy phép Apache 2.0.

Selenium bao gồm một số thành phần với mỗi thành phần tham gia vào một vai trò cụ thể trong việc hỗ trợ phát triển tự động kiểm tra ứng dụng Web.

a) IDE Selenium

Là một môi trường phát triển tích hợp hoàn chỉnh (IDE) cho các bài kiểm tra Selenium. Nó được thực hiện như một Tiện ích bổ sung của Firefox và cho phép ghi lại, chỉnh sửa và gỡ lỗi. Nó trước đây được gọi là Selenium Recorder. Selenium-IDE ban đầu được tạo ra bởi Shinya Kasatani và được tặng cho dự án Selenium vào năm 2006. Nó ít được duy trì và tương thích với Selenium RC. Các tập lệnh có thể tự động được ghi lại và chỉnh sửa theo cách thủ công, cung cấp hỗ trợ tự động hoàn thành và khả năng di chuyển các lệnh nhanh chóng. Kịch bản được ghi lại trong Selenese, một ngôn ngữ kịch bản thử nghiệm đặc biệt cho Selenium. Selenese cung cấp các lệnh để thực hiện các hành động trong một trình duyệt (bấm vào một liên kết, chọn một tùy chọn) và để lấy dữ liệu từ các trang kết quả.

b) API Client Selenium

Để thay thế cho các bài kiểm tra viết bằng Selen, các bài kiểm tra cũng có thể được viết bằng nhiều ngôn ngữ lập trình khác nhau. Những thử nghiệm này sau đó giao tiếp với Selenium bằng cách gọi các phương thức trong API ứng dụng khách Selenium. Selenium hiện cung cấp các API ứng dụng khách cho Java, C#, Ruby, JavaScript và Python. Với Selenium 2, một API ứng dụng khách mới đã được giới thiệu (với WebDriver là thành phần trung tâm của nó). Tuy nhiên, API cũ (sử dụng lớp Selenium) vẫn được hỗ trợ.

c) Selenium WebDriver

Selenium WebDriver là sự kế thừa của Selenium RC. Selenium WebDriver chấp nhận các lệnh (được gửi bằng Selenese hoặc thông qua API ứng dụng khách) và gửi chúng tới trình duyệt. Điều này được thực hiện thông qua trình điều khiển trình duyệt cụ thể cho trình duyệt, trình điều khiển sẽ gửi lệnh tới trình duyệt và truy xuất kết quả. Selenium WebDriver không cần một máy chủ đặc biệt để thực hiện các thử nghiệm. Thay vào đó, WebDriver trực tiếp khởi động một cá thể trình duyệt và điều khiển nó. Tuy nhiên, Selenium Grid có thể

được sử dụng với WebDriver để thực hiện các kiểm tra trên các hệ thống từ xa. Nếu có thể, WebDriver sử dụng chức năng cấp hệ điều hành gốc hơn là các lệnh JavaScript dựa trên trình duyệt để điều khiển trình duyệt. Điều này bỏ qua các vấn đề với sự khác biệt tinh tế giữa lệnh gốc và JavaScript, bao gồm cả các hạn chế bảo mật.

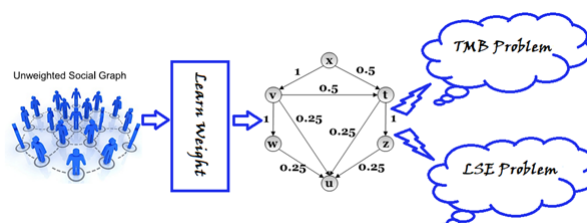
d) Selenium remote control

Selenium Remote Control (RC) là một máy chủ, được viết bằng Java, chấp nhận các lệnh cho trình duyệt thông qua HTTP. RC giúp có thể viết các bài kiểm tra tự động cho một ứng dụng web bằng bất kỳ ngôn ngữ lập trình nào, cho phép tích hợp tốt hơn Selenium trong các khung kiểm thử đơn vị hiện có. Để làm cho các bài kiểm tra viết dễ dàng hơn, dự án Selenium hiện cung cấp các trình điều khiển máy khách cho PHP, Python, Ruby, .NET, Perl và Java.

e) Selenium Grid

Selenium Grid là một máy chủ cho phép kiểm tra sử dụng các cá thể trình duyệt web chạy trên các máy từ xa. Với Selenium Grid, một máy chủ hoạt động như một hub. Kiểm tra liên hệ với trung tâm để có được quyền truy cập vào các phiên bản trình duyệt. Trung tâm có một danh sách các máy chủ cung cấp quyền truy cập vào các cá thể của trình duyệt (các nút WebDriver) và cho phép các thử nghiệm sử dụng các cá thể này. Selenium Grid cho phép chạy thử nghiệm song song trên nhiều máy và quản lý các phiên bản trình duyệt khác nhau và cấu hình trình duyệt tập trung (thay vì trong từng thử nghiệm riêng lẻ).

2.2. Kỹ thuật lấy dữ liệu trên MXH



Hình 3.1 Quá trình lấy dữ liệu

Quá trình lấy dữ liệu, áp dụng để xử lý được mô tả thông qua mô hình ở Hình 3.1.

a) Thu thập dữ liệu từ MXH Facebook

Các bước cụ thể của giai đoạn thu thập dữ liệu trên MXH Facebook được mô tả cụ thể như sau:

- Đầu tiên, nhóm tác giả tạo một tài khoản Facebook rồi sử dụng Công cụ kiểm thử tự động Selenium để tự động đăng nhập vào tài khoản Facebook đó.
- Bằng các truy vấn đặc biệt, nhóm tác giả lấy được ID (Identity Number) của tài khoản trên.
- Thông qua phương pháp giám sát và điều tra, thông qua những tin tức, vụ việc trong thời gian gần đây. Nhóm tác giả khảo sát dữ liệu với tập nguồn ban đầu bao gồm 6 tài khoản Facebook, đây là các đối tượng chủ mưu, cầm đầu, nắm vai trò quan trọng trong các tổ chức phản động, thường xuyên có các bài viết, bình luận mang tính chất chống đối, và đặc biệt các đối tượng này có uy tín khá lớn, có lượng người theo dõi, bạn bè trên Facebook là vô cùng đông đảo.
 - + <https://www.facebook.com/gioan.namphong>: Linh mục giáo xứ Thái Hà Nguyễn Nam Phong, một nhân vật khá nổi tiếng trong làng chống chính quyền trong nước và thường xuyên ra nước ngoài để "liên lạc" với tổ chức khủng bố Việt Tân.
 - + <https://www.facebook.com/minhnhat.paultran>: Paul Trần Minh Nhật là giáo dân thuộc xứ Ngọc Long, xã Công Thành, Yên Thành, Nghệ An, từng bị bắt và kết án 4 năm tù về tội “hoạt động nhằm lật đổ chính quyền nhân dân”. Hiện tại, Minh Nhật đang là cộng tác viên của trang “Tin mừng cho người nghèo”, mang danh nghĩa rao giảng tin mừng của Chúa, nhưng lại nhuộm màu sắc chính trị, luôn ủng hộ các đối tượng vi phạm pháp luật.

- + <https://www.facebook.com/ThaiDung2016>: Gioan Thái Văn Dung là sinh viên tốt nghiệp ngành tin học, đã tham gia quản lý cửa hàng Internet, Do hiểu biết được CNTT mà biết thêm về xã hội, tích cực truyền bá thông tin, hoạt động mạng, tham gia biểu tình chống TQ xâm lược.
- + <https://www.facebook.com/pham.doan.trang>: Nhà báo Phạm Đoan Trang, từng công tác tại Báo Pháp Luật thành phố Hồ Chí Minh. Tác giả của cuốn sách “Chính trị Bình Dân” có những nội dung nhạy cảm, mang tính chất kích động, chống phá chính quyền, Đảng và Nhà Nước.
- + <https://www.facebook.com/jbnguyenhuuvinh>: Anh Ba Sàm, tên thật là Nguyễn Hữu Vinh là một blogger, từng là công an và đảng viên Đảng Cộng sản Việt Nam, từng công tác ở Ủy ban Việt kiều Trung ương. Ông bị Chính phủ CHXHCN Việt Nam bắt giữ và bị cáo buộc và phạt tù 5 năm do có hành vi đăng tải các bài viết trên mạng Internet vi phạm Điều 258 Bộ Luật Hình sự năm 2015 sửa đổi bổ sung năm 2017 về Tội lợi dụng các quyền tự do dân chủ xâm phạm lợi ích của Nhà nước, quyền, lợi ích hợp pháp của tổ chức, công dân.
- + <https://www.facebook.com/profile.php?id=100015485029386>: Nguyễn Trọng đang sinh sống tại California, Hoa Kỳ. Lợi dụng quyền tự do dân chủ, thường xuyên có những bài viết xuyên tạc, chống phá đường lối chính sách của Đảng, pháp luật của Nhà nước.
- Tiếp theo, nhóm tác giả tự động trập cập vào trang web có địa chỉ www.facebook.com hoặc www.facebook.com/name/friends là trang bạn bè của những người dùng đó.
- Theo thuật toán sắp xếp của Facebook, những người dùng có liên quan nhất và hay tương tác nhất với người dùng đang theo dõi sẽ hiện ra đầu tiên trong trang bạn bè nói trên. Nhóm tác giả sử dụng thư viện BeautifulSoup để lấy mã HTML của trang này, tìm và lấy các đường dẫn chứa địa chỉ

Facebook bạn bè của người dùng đang theo dõi, lấy ID của họ.

- Nhóm tác giả thu thập theo phương pháp loang theo chiều rộng: Bắt đầu từ ID Facebook của 6 đối tượng kể trên, từ đó tìm tiếp thông tin về khoảng 20 người bạn xuất hiện đầu tiên trên trang Facebook bạn bè của mỗi người dùng đó. Với mỗi người dùng thu thập được, nhóm tác giả lại tiếp tục tiến hành thu thập như trên, và cứ tiếp tục tìm kiếm đến khi thu thập đủ số lượng người dùng cần thiết.

Cách làm này có ưu điểm là những người dùng thu thập được có nhiều mối liên hệ với nhau, đồ thị MXH thu thập được có nhiều liên kết, mang lại sự trực quan hơn cho cấu trúc đồ thị mạng. Ngoài ra còn có quá trình tiền xử lý, bước này sẽ hạn chế, loại bỏ được những đỉnh cô lập, hay những đỉnh khai thác được ít thông tin (Không công khai danh sách bạn bè).

Bên cạnh những ưu điểm trên, thì cách làm này cũng có nhược điểm là không khai thác được hết thông tin của người dùng, bởi nhiều người dùng không để chế độ công khai bạn bè. Trong thực tế sẽ để lọt nhiều đối tượng nguy hiểm, ẩn chứa nhiều nguy cơ tiềm tàng.

b) Mô hình hóa dữ liệu thu được

Ở bước này, từ những dữ liệu thu thập được về danh sách ID người dùng, mối quan hệ bạn bè giữa họ. Nhóm tác giả tiền xử lý, đánh số lại các đỉnh, xây dựng đồ thị mô tả MXH với các đỉnh đại diện cho các người dùng, các cạnh biểu diễn mối quan hệ bạn bè giữa những người đó, trọng số (nếu có) đại diện cho mức độ tương tác giữa người dùng với nhau. Kết quả đầu ra của thuật toán thu thập thông tin trên MXH Facebook bao gồm 2 file dữ liệu:

- ID.txt chứa ánh xạ từ ID của người dùng sang chỉ số đỉnh được đánh số lại. Qua quá trình thu thập dữ liệu, chuẩn hóa, nhóm tác giả thu được 542 ID người dùng, được đánh số lại từ 0 – 541, xem bảng 3.1
- Network.txt chứa mô tả cụ thể của đồ thị MXH đã thu được. File dữ liệu gồm nhiều dòng, mỗi dòng chứa 2 số là chỉ số của 2 người dùng có quan

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

0	100003708850657
1	100009945675640
2	100010197936720
3	641613321
4	100002541019308
5	100015485029386
6	100004745583805
...	...
539	100004236103486
540	100004142778267
541	100006259445960

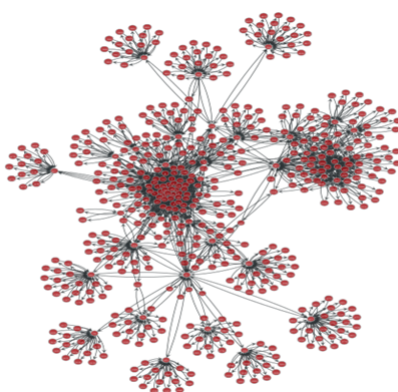
Bảng 3.1 Ví dụ ánh xạ ID người dùng sang chỉ số được đánh số.

hệ với nhau.

0	7
0	8
0	9
0	4
...	...
99	400
99	437
99	498

Bảng 3.2 Mô tả cụ thể của đồ thị MXH thu được

Từ dữ liệu thu được, nhóm tác giả mô phỏng lại và thu được hình ảnh tổng quan về cấu trúc mạng như sau :



Hình 3.2 Mô phỏng cấu trúc toàn mạng dữ liệu thu được

2.3. Giải pháp hạn chế: Limiting the spread of epidemics within time constraint on online social network (Ngăn chặn khả năng lan truyền của thông tin sai lệch trong thời gian giới hạn trên một mạng xã hội)

Giải pháp này được đề xuất và nghiên cứu sau khi xem xét thực tế quá trình lan truyền của một thông tin sai lệch trên mạng xã hội. Thông tin sai lệch luôn bắt đầu từ một số người dùng nhất định (có thể coi như đỉnh “nguồn” phát tán thông tin sai lệch). Những người dùng này thường là những tên có tư tưởng lệch lạc, phản động, và thông tin sai lệch thường là những thông tin bị bóp méo, những sai lầm của cán bộ bị thôi phồng, hoặc những lời lẽ bịa đặt vô căn cứ. Từ những người dùng bắt đầu này, thông tin sẽ được biết đến bởi những người xung quanh có liên kết với chúng, thường là những người dùng có quan hệ bạn bè với chúng. Thông tin này ban đầu có thể không tác động tới người dùng xung quanh, tuy nhiên nếu nó xuất hiện thêm nhiều lần từ nhiều nguồn khác nhau thì họ có thể sẽ tiếp nhận thông tin đó, và từ đó trở thành một nguồn phát thông tin sai lệch mới. Nếu mọi việc cứ phát triển thì thông tin này dần dần sẽ lan truyền sang toàn bộ một cộng đồng, hoặc thậm chí là toàn mạng xã hội (Quá trình lan truyền này được biểu diễn thông qua mô hình lan truyền được định nghĩa bên dưới). Do đó, cần có phương pháp giúp kiểm soát tình hình, ngăn chặn thông tin sai lệch lan truyền quá rộng, và đây cũng là lí do mà giải pháp này được nghiên cứu và công bố. Giải pháp được thiết kế để có thể chọn một số người dùng đặc

biệt. Những người dùng được chọn này sau đó sẽ được các cơ quan chức năng có thẩm quyền tiến hành các phương pháp phù hợp để ngăn chặn khả năng phát tán thông tin của họ. Từ đó, thông tin sai lệch sẽ không có khả năng truyền thông qua những người dùng này, và dẫn đến khả năng lan truyền sang những người dùng tiếp theo bị cản trở. Giải pháp này đã được chứng minh rằng nó có thể lựa chọn ra các đỉnh hiệu quả nhất để ngăn chặn, do đó vừa đem lại hiệu quả cao mà không hao tốn quá nhiều tài nguyên vào việc ngăn chặn các đỉnh kém hiệu quả khác.

a) Mô hình lan truyền thông tin Ngưỡng tuyến tính với bước thời gian rời rạc (Time Constraint Deterministic Linear Threshold - T-DLT).

Như đã được trình bày ở trên, có hai mô hình phổ biến trong bài toán lan truyền thông tin nói chung là mô hình IC và mô hình LT. Tuy nhiên, ở giải pháp này chúng tôi đề xuất một mở rộng mô hình DLT [36], vốn cũng là một trường hợp biến thể của mô hình LT, được gọi là T-DLT. Sở dĩ chúng tôi sử dụng mô hình này thay cho mô hình LT và IC bởi với bài toán này, mô hình T-DLT cho phép chúng tôi giảm thiểu tính phức tạp của cơ chế lan truyền, đồng thời vẫn đảm bảo kết quả không bị ảnh hưởng so với hai mô hình điển hình trên. Chi tiết mô hình T-DLT có thể được mô tả như sau:

- *Ký pháp đồ thị:* Đặt $G=(V,E,w)$ là biểu diễn của MXH với tập đỉnh V , tập cạnh có hướng E , với $|V|=n$ và $|E|=m$. Mỗi đỉnh đại diện cho một người dùng trong MXH, mỗi cạnh $e=(u,v)$ trong tập E tương ứng đại diện cho mối quan hệ giữa người dùng u và người dùng v . Chúng tôi ký hiệu đỉnh vào liền kề và đỉnh ra liền kề của đỉnh u lần lượt là $N_{in}(u)$ và $N_{out}(u)$, $d_{in}(u)$ và $d_{out}(u)$ lần lượt là bậc vào và bậc ra của đỉnh u . Đặt $I \subset V$ là tập các đỉnh bị lây nhiễm ban đầu. Mỗi cạnh có hướng (u,v) sẽ có trọng số $w(u,v)$, tượng trưng cho mức độ ảnh hưởng thông tin của v tới u , thỏa mãn $\sum_{(u,v)} w(u,v) \leq 1$.
- *Trạng thái của đỉnh:* Quá trình phát tán thông tin sai lệch từ tập đỉnh ban đầu I tới các đỉnh còn lại trong MXH tiến triển theo từng bước thời gian

rời rạc $t=1,2,\dots,d$. Mỗi đỉnh $v \in V$ có hai trạng thái là *bị lây nhiễm* (hay “kích hoạt”) và *không bị lây nhiễm* (hay “không kích hoạt”).

- *Ngưỡng lây nhiễm* (hay *ngưỡng kích hoạt*): Mỗi đỉnh v sẽ có một ngưỡng lây nhiễm cho trước $\theta(v) \in [0,1]$. Giá trị này đại diện cho trọng số vào của đỉnh v cần chuyển thành bị lây nhiễm để v trở thành bị lây nhiễm.
- *Quá trình lan truyền*: Quá trình lây lan thông tin sai lệch được mô phỏng tuần tự theo từng bước thời gian rời rạc (còn gọi là “vòng”, “bước”) $t=0,1,2,\dots,d$. Đặt $I(t)$ là tập các đỉnh bị lây nhiễm sau t bước, quá trình lan truyền được mô phỏng như sau:

- + Tại bước $t=0$, tất cả các đỉnh trong tập I đều bị lây nhiễm, $I(0)=I$.
- + Tại bước $t \geq 1$, tất cả các đỉnh không bị lây nhiễm v sẽ chuyển thành bị lây nhiễm nếu tổng số các đỉnh vào gần kề chạm ngưỡng lây nhiễm của nó, $\sum_{(u,v) \in E} w(u,v) \geq \theta(v)$.
- + Một đỉnh bị lây nhiễm sẽ giữ nguyên trạng thái bị lây nhiễm cho tới hết quá trình lan truyền. Quá trình lan truyền chấm dứt khi $t = d$.

Trong mô hình LT, giá trị của ngưỡng $\theta(v)$ được đặt một cách ngẫu nhiên trong khoảng $[0,1]$ và sẽ được chỉnh sửa dựa vào thông tin có được thêm trong quá trình lan truyền. Do đó, mô hình này thuộc dạng mô hình ngẫu nhiên. Khác với mô hình LT, trong mô hình T-DLT ngưỡng lây nhiễm $\theta(v)$ được cho trước. Trường hợp này, giá trị ngưỡng có thể được đặt dựa theo các khảo sát thực tế hoặc các phương pháp khai phá dữ liệu.

b) Định nghĩa bài toán

Đối với giải pháp này, việc ngăn chặn thông tin sai lệch được tiến hành bằng cách loại bỏ một số người dùng ra khỏi mạng để thông tin không thể truyền qua được, sao cho phạm vi lan truyền của thông tin là nhỏ nhất có thể. Nếu như coi mạng xã hội là một đồ thị, thì ta có thể hiểu rằng những đỉnh “bị lây nhiễm” là những người dùng đã bị ảnh hưởng bởi thông tin sai lệch và thông tin sai lệch đó có thể truyền từ người dùng này sang những người dùng thân cận, những đỉnh

“không bị lây nhiễm” là những đỉnh không bị ảnh hưởng bởi thông tin sai lệch, còn những đỉnh “cứu được” là những đỉnh mà có trạng thái chuyển từ “bị lây nhiễm” sang “không bị lây nhiễm” khi chiến thuật loại bỏ đỉnh được áp dụng. Trong thực tế, việc loại bỏ một người dùng ra khỏi mạng xã hội có thể được thực hiện bằng cách thuyết phục người dùng đó thoát khỏi sự ảnh hưởng của thông tin sai lệch, hoặc ngăn chặn kết nối mạng của người dùng đó, hoặc xóa bỏ tài khoản của người dùng ra khỏi mạng xã hội, v.v. . . Tuy nhiên, những cách làm này đều là những cách làm yêu cầu có sự phối hợp từ nhiều bên liên quan, do đó nếu tiến hành loại bỏ quá nhiều sẽ dẫn đến chi phí phát sinh về thời gian và vật chất rất lớn, thậm chí sẽ gây ra trường hợp loại bỏ những thông tin lan truyền quá rộng dẫn đến vô ích. Vì vậy, vấn đề đặt ra là phải làm sao chọn ra được một số lượng K đỉnh nào đó để khi thực hiện chiến thuật loại bỏ, ta thu được số đỉnh cứu nhiều nhất. Số lượng K sẽ được tính toán một cách cân đối, cẩn thận nhất về chi phí để không làm mất tác dụng của chiến thuật loại bỏ. Chiến thuật loại bỏ K đỉnh này có thể được mô hình hóa và phát biểu khoa học như sau:

Kí hiệu $f_d(I)$ là tập hợp các đỉnh đã bị lây nhiễm sau d vòng trên đồ thị $G = (V, E)$ trong mô hình T-DLT, $f_d(I, A)$ là tập các đỉnh bị lây nhiễm sau khi loại bỏ một tập các đỉnh $A \subseteq V$ từ G (số đỉnh bị lây nhiễm trong đồ thị sót lại). Khi đó, số đỉnh cứu được khi loại bỏ tập A là: $h_d(A) = |f_d(I, \emptyset)| - |f_d(I, A)|$

Trong mô hình T-DLT, ta xây dựng một bài toán tối ưu tổ hợp là: *Hạn chế sự lây lan của thông tin sai lệch* (Limiting the Spread of Epidemics - LSE), có mục tiêu tìm kiếm một tập đỉnh có kích cỡ tối đa k đỉnh sao cho khi loại bỏ tập đỉnh đó, số đỉnh cứu được sẽ đạt cực đại.

Định nghĩa (Bài toán LSE): Cho đồ thị vô hướng $G=(V, E)$ biểu diễn một MXH trong mô hình T-DLT. Một tập các đỉnh bị lây nhiễm ban đầu $I \subset V$, số vòng lan truyền d và số đỉnh tối đa loại bỏ được là $k > 0$. Tìm một tập k đỉnh $A \subseteq V$ sao cho khi loại bỏ khỏi mạng thì số đỉnh cứu được sau d vòng là $h_d(A)$ đạt cực đại.

Kí hiệu $G_d = (V_d, E_d)$ đồ thị con của đồ thị $G=(V, E)$ trong đó V_d là tập các đỉnh mà khoảng cách giữa một đỉnh bất kì trong đó với một đỉnh trong tập I tối

đa là d , và tập E_d là tập các cạnh nối từ các đỉnh thuộc tập I sang các đỉnh khác có khoảng cách tối đa là d , đồng thời $n_d = |V_d|$, $m_d = |E_d|$. Ta thấy rằng sự lan truyền thông tin sai lệch chỉ xảy ra bên trong đồ thị G_d . Do đó, để đơn giản hóa, thay vì xét đến tất cả các đỉnh thuộc G , ta sẽ chỉ tìm kiếm kết quả trong G_d .

c) Độ phức tạp và tính xấp xỉ của bài toán LSE

Trong mục này, chúng tôi chỉ ra tính NP-Khó của bài toán LSE bằng cách tương đương hóa bài toán LSE với bài toán Phủ đỉnh (Set Cover). Chúng tôi cũng sẽ chỉ ra tính khó xấp xỉ của LSE, tức là việc xấp xỉ bài toán LSE với tỉ lệ $n^{1-\varepsilon}$, với $0 < \varepsilon < 1$, là bài toán NP-Khó.

Định lí: LSE là bài toán NP-Khó trong mô hình T-DLT.

Chứng minh: Để chứng minh LSE là bài toán NP-Khó, ta giảm thể bài toán về bài toán Phủ đỉnh (Set cover)

Bài toán Phủ đỉnh (Set Cover - SC): Cho một số tự nhiên dương t , một tập phổ quát $\vartheta = e_1, e_2, \dots, e_M$ và tập con $S = s_1, s_2, \dots, s_N$ ta có thể giả sử rằng $t < M < N$. Bài toán Phủ đỉnh yêu cầu rằng: Có tồn tại hay không 1 lượng t tập con, sao cho chúng là ϑ .

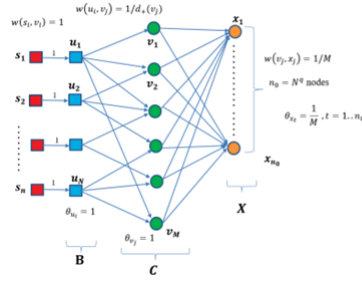
Bài toán SC đã được chứng minh là bài toán NP-Khó khi kích thước các tập được giới hạn là 3 và mỗi phần tử xuất hiện trong đúng hai tập con. Để giảm thể bài toán SC về bài toán LSE, đầu tiên ta xây dựng một biểu diễn J_{LSE} của LSE từ biểu diễn J_{SC} của bài toán SC. Sau đó chúng tôi chỉ ra rằng nếu J_{SC} có tồn tại lời giải S kích thước t , thì J_{LSE} có lời giải A , $|A| \leq k$ với $h_d(A) \geq k + M$ và ngược lại.

Xây dựng: Cho một biểu diễn $J_{SC} = (\vartheta, S, t)$ của bài toán SC, ta xây dựng biểu diễn $J_{LSE} = (G, I, d, k)$ của bài toán LSE như theo hình 3.3

- Tập các đỉnh và các cạnh với mỗi $S_i \in S$, ta xây dựng một đỉnh u_i . Ta có thêm một cạnh có hướng (s_i, u_i) . Với mỗi phần tử $e_j \in \vartheta$ ta thêm một đỉnh v_j và thêm một cạnh có hướng (u, v) với mỗi $e_j \in S_i$. Để thuận tiện, ta đặt $B = u_1, u_2, \dots, u_N$, $C = v_1, v_2, \dots, v_M$.

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

- Ngưỡng lây nhiễm và trọng số: ta đặt $w(u,v) = 1$, $w(u_i, v_i) = \frac{1}{d_{in}(v_j)}$, $\theta_{u_i} = \theta_{v_i} = 1$.
- Sau cùng, ta cho $k = t$, $d = 2$.



Hình 3.3 Giảm thể từ bài toán SC về bài toán LSE.

Phân tích: Qua bước xây dựng, ta thấy rằng $|f_d(I, \emptyset)| = M + N$. Nếu như có đỉnh bất kì $u_i \in B$ là đỉnh kề vào và $v_j \in C$ là đỉnh chưa bị lây nhiễm thì:

$$\sum_{C \in \text{OnhngxmtiOblynhimu}} w(u, v_j) \leq 1 - \frac{1}{d_{in}(v_j)} < 1 = \theta_{v_j}$$

Do đó v_j là đỉnh không bị lây nhiễm. Nếu không, các đỉnh trong C là các đỉnh không bị lây nhiễm.

(\rightarrow) Giả sử S' là lời giải của biểu diễn J_{SC} , có nghĩa là $|S'| = t = k$ và nó phủ t phần tử ϑ . Nếu ta chọn tập A chứa đỉnh u_i tương ứng với $S_i \in S'$, với mọi đỉnh $v_i \in B$ kề với ít nhất một đỉnh trong A . Bằng bước phân tích trên, mọi đỉnh trong C là đỉnh không bị lây nhiễm. Ta có $h_d(A) = t + M = k + M$

(\leftarrow) Ngược lại, nếu J_{LSE} có lời giải A , $|A| \leq k$ với $h_d(I, A) \geq k + M$. Nếu A chứa t_1 ($1 \leq t_1 \leq k$) đỉnh trong C , $h_d(I, A) \leq k - t_1 + M < k + M$. Do đó A không chứa bất kì đỉnh nào thuộc C . Ở đây, $A \subset B$. Kết hợp với $h_d(I, A) \geq k + M$, mọi đỉnh thuộc C là đỉnh không bị lây nhiễm. Do đó, mọi đỉnh $u_i \in A$ kề ít nhất với một đỉnh thuộc C . Vì vậy, qua bước xây dựng ta thấy rằng, $S' = S_i | u_i \in A$ là lời giải cho bài toán J_1 . Định lý được chứng minh.

Bằng cách thay đổi một chút phương pháp chứng minh bên trên, ta có thể chứng minh thêm được một định lý về tính khó xấp xỉ của bài toán LSE.

Định lý: Đánh giá xấp xỉ bài toán LSE với tỉ lệ $n^{1-\varepsilon}$ trong mô hình T-DLT với mọi hằng số $0 < \varepsilon < 1$ là bài toán NP-khó .

Chứng minh: Để chứng minh kết quả này, ta sử dụng phương pháp gap-introduction reduction [37] để chứng minh tính khó xấp xỉ của bài toán LSE. Sử dụng phép giản thể trong thời gian đa thức từ bài toán SC thành bài toán LSE , chúng tôi sẽ chỉ ra nếu tồn tại một thuật toán có thời gian đa thức có thể xấp xỉ được bài toán trên với tỉ lệ $1-\varepsilon$, thì sẽ tồn tại một thuật toán có thời gian đa thức để giải bài toán gốc.

Xây dựng: Cho một biểu diễn của bài toán SC, $J_{SC} = (\vartheta, S, t)$, ta xây dựng một biểu diễn $J_{LSE} = (G, I, d, k)$

- Tập các đỉnh và các cạnh: Với mỗi đỉnh $v_j \in C$, ta thêm $n_0 = N^q$ đỉnh $X = x_1, x_2, \dots, x_{n_0}$ với một hằng số q đủ lớn và thêm một cạnh có hướng (v_j, x_l) , $l = 1 \dots n_0$.
- Ngưỡng lây nhiễm và trọng số: Gán $w(v_j, x_l) = \frac{1}{M}$; $\theta_{x_l} = \frac{1}{M}$, $l = 1 \dots n_0$.

Giả sử J_{SC} có tập phủ đỉnh S' kích cỡ t , ta chọn tập $A = u_i | S_i \in S$. Qua bước phân tích, mọi đỉnh C đều là các đỉnh không bị lây nhiễm. Điều này dẫn tới mọi đỉnh trong X đều là các đỉnh không bị lây nhiễm. Bởi bước xây dựng này, ta có: $|f_d(I, \emptyset)| = N + M + N^q$, $|f_d(I, A)| = N - k$, do đó:

$$h(A) = |f_d(I, \emptyset)| - |f_d(I, A)| = M + N^q + k$$

Trong trường hợp J_{SC} không có tập phủ đỉnh kích thước t , nó có ít nhất một đỉnh v_j bị lây nhiễm. Điều này dẫn tới việc mọi trong X đều bị lây nhiễm. Do đó, $|f_d(I, A)| > N - k + N_q$ dẫn tới $h(A) = M + k < 2N$.

Lúc này giá trị của thuật toán có thể xấp xỉ bài toán LSE trong tỉ lệ $n^{1-\varepsilon}$, ta chứng minh được rằng lời giải của bài toán SC có thể được tìm trong thời gian đa thức. Với bất kì biểu diễn J_{SC} nào, ta xây dựng một biểu diễn J_{LSE} theo cách xây dựng trên trong thời gian đa thức của m và n .

Trong trường hợp J_{SC} có tập đỉnh phủ kích cỡ t , theo cách xây dựng của chúng tôi, lời giải tối ưu A_{opt} của J_{LSE} có $h(A_{opt}) = k + M + N_q$. Thuật toán xấp xỉ lời giải tối ưu với tỉ lệ $n^{1-\varepsilon}$ ($n = 2N + M + N_q$ là số đỉnh của đồ thị đầu vào) do đó có thể tìm kiếm lời giải.

$$h(A(J_{LSE})) \geq \frac{1}{n^{1-\varepsilon}} h(A_{opt}) > \frac{1}{n^{1-\varepsilon}} (M + N_q + k) = \frac{n^\varepsilon}{n} (M + N_q + k) > \frac{(M + N_q + k)^\varepsilon}{M + N_q + k} N_q > \frac{N^{q \cdot \varepsilon}}{4N^q} N^q = \frac{1}{4} N^{q \cdot \varepsilon}.$$

Cho q đủ lớn để $q > \frac{\ln(8N)}{\varepsilon \cdot \ln(N)}$, ta được:

$$h(A(J_{LSE})) > 2N.$$

Nếu J_{LSE} không có tập phủ đỉnh kích cỡ t , thì tập tối ưu A_{opt} của J_{LSE} có

$$h(A(J_{LSE})) < 2N.$$

Điều này cho thấy J_{LSE} có một tập phủ đỉnh kích cỡ t khi và chỉ khi $h(A(J_{LSE})) < 2N$. Do đó ta có thể dùng A để giải quyết bài toán SC trong thời gian đa thức, hay $P = NP$. Điều này mâu thuẫn với giả thiết $P \neq NP$.

d) Thuật toán nhanh và hiệu quả để giới hạn sự lây nhiễm thông tin (Fast And Effective Limiting Epidemics – FLE).

Trong phần này, chúng tôi thiết kế một thuật toán heuristics có độ phức tạp thấp và kết quả rất khả quan. Trước khi bắt đầu, chúng tôi có định nghĩa một vài kí hiệu sẽ được sử dụng trong quá trình đề xuất thuật toán:

- $t(u)$: vòng nhỏ nhất sao cho đỉnh u chuyển từ đỉnh không bị lây nhiễm sang đỉnh bị lây nhiễm.
- $a_{in}(u) = \sum_{I_{t(u)-1} \cap N_{in}(u)} w(u, v)$: tổng trọng số các cạnh vào từ các đỉnh vào hàng xóm đã bị lây nhiễm trước vòng $t(u)$.
- $\infty(u) = \sum_{v \in U_{i=t(u)+1}^d} I_i \cap N(u) w(u, v)$, $i = t(u) + 1, \dots, d$: tổng trọng số các cạnh ra từ u tới các đỉnh bị lây nhiễm b tại vòng i .
- $\beta(u)$: số lượng các đỉnh ra, sao cho chúng chuyển từ bị lây nhiễm thành không bị lây nhiễm sau khi loại bỏ đỉnh u ra khỏi đồ thị.

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

Một cách trực quan, số lượng gia tăng các đỉnh được cứu $\delta(A, u)$ có thể xấp xỉ theo $\beta(u)$. Bên cạnh đó, để tăng cường tính hiệu quả của thuật toán đề xuất, chúng tôi kết hợp $\alpha(u)v\beta(u)$ để tính toán khả năng lan truyền tin sai lệch của đỉnh u . Ý tưởng chính của thuật toán là chọn ra các đỉnh một cách lần lượt theo đánh giá của hai hàm $\alpha v\beta$. Ban đầu, chúng tôi khởi tạo tập đỉnh được chọn $A=\emptyset$ và đặt tập các đỉnh được xét chọn U bằng với V_d . Trong mỗi bước, chúng tôi chọn ra đỉnh có trị số $\beta(u)$ lớn nhất trong đồ thị còn lại. Trường hợp tất cả các đỉnh đều có giá trị $\beta(u)$ là 0, chúng tôi chọn đỉnh có giá trị $\alpha(u)$ cực đại. Thuật toán được mô tả ở hình 3.4 sau.

```
Data: Graph  $G = (V, E, \theta)$ , set of initial infected nodes  $I$ ,  
propagation hop  $d$ .  
Result: set of nodes  $A$   
1.  $S \leftarrow \emptyset$ ;  
2. Calculate  $G' \leftarrow G_d(I)$ ;  $U \leftarrow V_d$   
3. for  $i = 1$  to  $k$  do  
4.   Calculate  $\alpha(u), \beta(u)$  on  $G'$  (Algorithm 3)  
5.    $u_{max} \leftarrow 0$   
6.   if  $\beta(v) = 0, \forall v \in U$  then  
7.      $u_{max} \leftarrow \arg \max_{v \in U} \alpha(v)$   
8.   else  
9.      $u_{max} \leftarrow \arg \max_{v \in U} \beta(v)$   
10.  end  
11.   $S \leftarrow S \cup \{u_{max}\}$   
12.   $U \leftarrow U \setminus \{u_{max}\}$   
13.  Remove  $u_{max}$  and all edges that adjacent with  $u_{max}$  from  $G'$   
14. end  
15. return  $S$ ;
```

Hình 3.4 Thuật toán nhanh và hiệu quả để giới hạn sự lây nhiễm thông tin

Khó khăn trong thuật toán chủ yếu đến từ bước tính toán các hàm $\alpha v\beta$. Để giải quyết việc này, chúng tôi bắt đầu với ý tưởng tìm kiếm theo chiều rộng (Breath First Search - BFS) để tính toán $f_d(I)$ và dần dần cập nhật hàm $\alpha_{in}()$. Sau đó, chúng tôi sử dụng nó để tính toán các hàm $\alpha v\beta$ theo định nghĩa của chúng. Thuật toán chi tiết được thể hiện trong hình 3.5

Chứng minh: Để đánh giá độ phức tạp của thuật toán này, chúng ta cần đánh giá trước độ phức tạp của thuật toán. Ta dễ thấy rằng việc tính toán $f_d(I)$ (dòng 5 - 19) có cách thức hoạt động giống BFS. Do đó, độ phức tạp của nó là $O(m_d v n_d)$. Với bước tính toán $\alpha(u)v\beta(u)$ (dòng 21 - 31), ta cần thăm tất cả các đỉnh trong $f_d(I)$ và tập các cạnh ra liên kề, do đó độ phức tạp của nó là $O(m_d v n_d)$. Do đó, độ phức tạp chung của thuật toán là $O(m_d v n_d)$.

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

Data: Graph $G = (V, E, \theta)$, set of infected nodes I , propagation hop d

Result: $f_d(I), a_+(u), t(u), \alpha(u), \beta(u), \forall u \in G$

```

1.  $t(u) \leftarrow +\infty, a_+(u) \leftarrow 0, \alpha(u) \leftarrow 0, \beta(u) \leftarrow 0, \forall u \in V_d$ 
2.  $t(s) \leftarrow 0, \forall s \in I$ 
3. Queue  $Q \leftarrow \emptyset$ 
4.  $\backslash\backslash$  Calculate  $a_+(u), t(u), f_d(I)$ 
5. while  $Q \neq \emptyset$  do
6.    $u \leftarrow Q.pop()$ 
7.    $f_d(I) \leftarrow u$ 
8.   if  $t(u) < d$  then
9.     foreach  $v \in N_-(u)$  do
10.      if  $t(u) < t(v)$  then
11.         $a_+(v) = a_+(u) + w(u, v)$ 
12.        if  $(a_+(v) \geq \theta_v)$  and  $(t(u) + 1 < t(v))$  then
13.           $t(v) \leftarrow t(u) + 1$ 
14.           $Q.push(v)$ 
15.        end
16.      end
17.    end
18.  end
19. end
20.  $\backslash\backslash$  Calculate  $\alpha(u), \beta(u)$ 
21. foreach  $u \in f_d(I)$  do
22.    $\alpha(u) \leftarrow 0; \beta(u) \leftarrow 0$ 
23.   foreach  $v \in N_-(u) \cap f_d(I)$  do
24.    if  $t(u) < t(v)$  then
25.       $\alpha(u) \leftarrow \alpha(u) + w(u, v)$ 
26.      if  $a_+(v) - w(u, v) < \theta_v$  then
27.         $\beta(u) \leftarrow \beta(u) + 1$ 
28.      end
29.    end
30.  end
31. end
32. return  $A$ ;

```

Hình 3.5 Thuật toán tính $\alpha(u)\beta(u)$ có độ phức tạp $O(k(m_d + n_d))$

2.4. Giải pháp hạn chế Targeted Misinformation Blocking (Xác định và ngăn chặn thông tin sai lệch trên mạng xã hội)

Với cách tiếp cận như cách Twitter đã làm năm 2016 là loại bỏ 125.000 tài khoản được nghi ngờ có liên quan đến các hoạt động khủng bố, hay như Facebook đã xóa 30.000 tài khoản giả mạo lan truyền các tin đồn trước cuộc bầu cử tổng thống Pháp năm 2017, giải pháp hạn chế Targeted Misinformation Blocking được đưa ra nhằm mục đích tìm những tài khoản cần loại bỏ khỏi mạng để làm hạn chế phát tán thông tin sai lệch.

a) Định nghĩa bài toán

Kí hiệu $G=(V,E,w)$ là một đồ thị có hướng biểu diễn cho một mạng xã hội trên mô hình LT với V biểu diễn tập người dùng, E biểu diễn mối quan hệ giữa các cặp người người dùng, w là tập trọng số của các cạnh biểu diễn tần số tương tác giữa các người dùng. Gọi $N_{out}(v)$ và $N_{in}(v)$ lần lượt là tập đỉnh đi ra từ v và đi

vào v . Mỗi cạnh có hướng $(u, v) \in E$ được gán một trọng số $w(u, v) \in [0, 1]$ đảm bảo $\sum_{u \in N_{out}(v)} w(u, v) \leq 1$.

Đưa ra một tập nguồn $S = s_1, s_2, \dots, s_n, s_i \in V$ là tập tiềm năng phát tán thông tin sai lệch (tương tự như tập hạt giống trong bài IM [?]).

Trong mô hình LT, mỗi đỉnh $v \in V$, có hai trạng thái active và inactive. Quá trình lan truyền trên mô hình LT được diễn tả như sau: Đầu tiên, mỗi đỉnh v được gán một trọng số ngẫu nhiên $\theta_v \in [0, 1]$ thể hiện giá trị active của đỉnh v phải đạt được để có thể active v . Tiếp đến quá trình lan truyền diễn ra trong các vòng $t = 1, 2, 3, \dots$ như sau:

- Ở vòng 1, tất cả các đỉnh $s_i \in S$ ở trạng thái active và các đỉnh còn lại ở trạng thái inactive.
- Ở vòng $t \geq 1$, một đỉnh v ở trạng thái inactive được active nếu thỏa mãn điều kiện $\sum_{u \in N_{in}(v), v \text{ active}} w(u, v) \geq \theta_v$.
- Khi một đỉnh ở trạng thái active thì sẽ luôn duy trì trạng thái đó trong suốt quá trình lan truyền thông tin. Quá trình lan truyền kết thúc khi không có đỉnh nào được active thêm.

Kí hiệu $\sigma_S(G)$ là ảnh hưởng lan truyền của tập S trên đồ thị G dưới mô hình LT hay là số lượng đỉnh được active bởi nguồn S . Theo nghiên cứu của Kempe và các cộng sự đã chỉ ra rằng mô hình LT là tương đương với khả năng tiếp cận trong một đồ thị ngẫu nhiên g hay còn gọi là đồ thị live-edge (đồ thị mẫu). Đồ thị live-edge được xây dựng như sau:

- Với mỗi $v \in V$, chọn tối đa một cạnh đi vào nó một cách ngẫu nhiên, xác suất (u, v) được chọn là $w(u, v)$.
- Xác suất không có cạnh nào được chọn là $1 - \sum_{u \in N_{in}(v)} w(u, v)$.

Cạnh được chọn gọi là cạnh sống (live-edge) và tất cả các cạnh khác gọi là cạnh bị chặn (blocked-edge). Chúng ta có :

$$\sigma_S(G) = \sum_{g \in G^*} Pr[g] R(g, S)$$

Trong đó G^* là tập tất cả các đồ thị sinh ra bởi G theo mô hình live-edge với xác suất $\Pr[g]$ và $R(g, S)$ là tập có thể đi đến từ S trong g .

Ảnh hưởng lan truyền của nguồn S khi xóa bỏ các đỉnh trong tập A trên đồ thị $G[V \setminus A]$ kí hiệu là $\sigma_S(G \setminus A)$.

Phát biểu bài toán (Bài toán TMB): Gọi $G(V, E, w)$ là một đồ thị có hướng biểu diễn một mạng xã hội. Đưa ra một tập nguồn $S = s_1, s_2, \dots, s_k, S \in V$ là tập tiềm năng phát tán thông tin sai lệch và một số nguyên $\gamma \leq |V|$ thể hiện cho số lượng người dùng chúng ta muốn đảm bảo thông tin sai lệch không lan truyền đến được họ.

Yêu cầu : Tìm tập đỉnh $A \in V \setminus S$ có ít đỉnh nhất sao cho khi xóa các đỉnh trong tập A khỏi đồ thị G thì đảm bảo hàm mục tiêu $h(A) = \sigma_S(G) - \sigma_S(G \setminus A) \geq \gamma$, tức là đảm bảo ít nhất γ người không tiếp xúc thông tin sai lệch.

b) Thuật toán đề xuất

Thuật toán tham lam (Greedy algorithm) Trong phần này chúng tôi giới thiệu một thuật toán xấp xỉ đảm bảo tỉ lệ $1 + \ln(\gamma/\epsilon)$ dựa vào tính chất submodular và tính đơn điệu của hàm $h(A)$.

Data: Graph $G = (V, E, w)$, $S = \{s_1, s_2, \dots, s_k\}$, threshold $\gamma < |V|$, parameter $\epsilon \in (0, \gamma)$
Result: set of nodes A

1. $A \leftarrow \emptyset$;
2. **while** $h(A) > \gamma - \epsilon$ **do**
3. $u = \arg \max_{u \in V \setminus (A \cup S)} h(A \cup \{u\}) - h(A)$; $A \leftarrow A \cup \{u\}$;
4. **end**
5. **return** A ;

Thuật toán tham lam Greedy.

Thuật toán TMB (TMB algorithm) Thuật toán tham lam đưa ra một lời giải tốt xong lại không đảm bảo yêu cầu về thời gian với dữ liệu lớn do đó trong phần này chúng tôi đưa ra một giải pháp đảm bảo yêu cầu cả về độ tốt cũng như thời gian tính toán. Dưới đây là các bước chính trong thuật toán chúng tôi đề xuất:

- Đầu tiên sáp nhập các đỉnh trong tập nguồn S thành một nút I bằng cách áp dụng thuật toán 2 được đề xuất bởi Zhang [?]. Chúng ta thu được một

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

đồ thị mới và nguồn phát tán thông tin sai lệch bây giờ chỉ là một đỉnh I duy nhất.

```
Data:  $G = (V, E, w), S$ .
Result:  $G', w', I$ 
1  $G' \leftarrow G$ 
2 Add node  $I$  to  $G'$ 
3 for  $s \in S$  do
4   if there exists edge  $(s, v)$  then
5     if  $(I, v) \notin G'$  then
6       Add edge  $(I, v)$  to  $G'$ 
7        $w'(I, v) = w(s, v)$ 
8     else
9        $w'(I, v) = w'(I, v) + w(s, v)$ 
10    end
11    Remove  $(s, v)$  from  $G'$ 
12  end
13 end
14 Remove all node in  $S$  from  $G'$ 
15 Return  $G', I$ 
```

Thuật toán hợp nhất nguồn

- Tiếp đến chúng tôi tạo ra $\mu (\mu \in N^*)$ đồ thị mẫu g từ G bằng cách: mỗi đỉnh $v \in V$ chọn nhiều nhất một đỉnh kề đi đến nó một cách ngẫu nhiên sao cho xác suất chọn cạnh (u, v) là $w(u, v)$ và xác suất không chọn cạnh nào là $1 - \sum_u w(u, v)$. Những cạnh được chọn được gọi là cạnh sống (live-edge) và tất cả các cạnh còn lại được gọi là cạnh bị chặn (blocked-edge). Chi tiết cách lấy đồ thị mẫu được mô tả như sau:

```
1  $G' \leftarrow \emptyset$ 
2 for  $v \in V_G$  do
3    $\gamma_{\text{active}} \leftarrow \text{random}(0, 1)$ 
4    $\text{sum\_weight} \leftarrow 0$ 
5   for  $u \in \text{in\_neighbors}(v)$  do
6      $\text{sum\_weight} += w(u, v)$ 
7     if  $\text{sum\_weight} \geq \gamma_{\text{active}}$  then
8        $G'.\text{add\_edge}(u, v)$ 
9       Break
10    end
11  end
12 end
13 return  $G'$ 
```

Thuật toán lấy đồ thị mẫu theo mô hình LT

- Trên cây T_i , ảnh hưởng địa phương của nút v là số nút con của cây gốc v kể cả v trên cây T_i . Áp dụng thuật toán DFS để dàng tìm được số nút con của v . Ảnh hưởng giảm của nguồn S khi loại đỉnh u khỏi G bằng lượng ảnh hưởng giảm trung bình khi loại u trên tất cả các cây T_i . Sau đó áp dụng phương pháp Lazy forward của Leskovec [?] để tìm kiếm lời giải. Chi tiết cách làm được mô tả trong thuật toán dưới đây.

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

Algorithm 4: Scalable TMB (STMB) Algorithm

Data: Graph $G = (V, E, w)$, $S = \{s_1, s_2, \dots, s_q\}$, threshold $\gamma > 0$
Result: set of nodes A

1. $A \leftarrow \emptyset$; $(G', I) \leftarrow \text{Merge}(G, S)$.
2. Remove all node, I can't reach in G .
3. Generate η live-edge graphs and set η tree $\mathcal{L} = \{T_1^1, T_1^2, \dots, T_1^{|\eta|}\}$
4. For each $T_l \in \mathcal{L}$, calculate $r(u, T_l)$ for all $u \in T_l$ (by using DFS algorithm).
5. **for** $u \in V$ **do**
6. $u.\delta(u) \leftarrow \frac{1}{\eta} \sum_{T_l \in \mathcal{L}} r(u, T_l)$; $u.cur \leftarrow 1$
7. Insert element u into Q with $u.\delta(u)$ as the key
8. **end**
9. $h_{max} \leftarrow 0$; iteration $\leftarrow 1$
10. **while** $h_{max} < \gamma - \epsilon$ **do**
11. $u_{max} \leftarrow \text{dequeue } Q$
12. **if** $u_{max}.cur = \text{iteration}$ **then**
13. $A \leftarrow A \cup \{u_{max}\}$; iteration $\leftarrow \text{iteration} + 1$
14. **foreach** $T_l \in \mathcal{L}$ **do**
15. **if** $u_{max} \in T_l$, remove node u_{max} and update $r(v, T_l)$, $\forall v \in T_l$.
16. **end**
17. $h_{max} \leftarrow h_{max} + u_{max}.\delta(u_{max})$
18. **else**
19. $u_{max}.\delta(u_{max}) \leftarrow \frac{1}{\eta} (\sum_{T_l \in \mathcal{L}} r(I, T_l) - \sum_{T_l \in \mathcal{L}} r(I, T_l \setminus u_{max}))$
20. $u_{max}.cur = \text{iteration}$; re-insert u_{max} into Q
21. **end**
22. **end**
23. **return** A ;

Thuật toán xác định và ngăn chặn thông tin sai lệch STMB

Độ phức tạp của thuật toán:

Để đánh giá độ phức tạp của thuật toán trước tiên chúng ta đi phân tích độ phức tạp của các thuật toán con trong nó.

- Thuật toán Merge source node có độ phức tạp là $O(k + |N_{in}(S)|)$.
- Quá trình tạo ra μ đồ thị mẫu dòng 3 có độ phức tạp $O(\mu(m + n))$.
- Do T_l là một cây tính toán $r(T_l, u) \forall u \in T_l$ trong $O(\mu n)$
- Phần Lazy Forward có độ phức tạp là $O(qmn)$ với q là số vòng lặp từ dòng 10 đến 23.

Vậy nên toàn bộ thuật toán TMB có độ phức tạp là $O(\mu(m + qn))$.

2.5. Áp dụng các giải pháp đề xuất trên dữ liệu thực

Trong phần này nhóm tác giả đưa ra kết quả thực nghiệm của giải hai giải pháp chỉ ra trong mục trước trên dữ liệu có sẵn và dữ liệu thu thập được, từ đó để thấy được hiệu quả của phương pháp nhóm đề xuất.

a) Giải pháp *Limiting the spread of epidemics within time constraint on online social network*

- Dữ liệu có sẵn

Trong phần này, chúng tôi thực nghiệm trên một số MXH trực tuyến để kiểm chứng tính hiệu quả của thuật toán khi so sánh nó với các thuật toán cơ sở

Chương 3. GIẢI PHÁP HẠN CHẾ THIẾT HẠI DO THÔNG TIN SAI LỆCH GÂY RA TRÊN MẠNG XÃ HỘI

Mạng	Số đỉnh	Số cạnh	Loại	Bậc trung bình
Gnutella	6,301	20,777	Có hướng	3.29
Wikipedia vote	7,115	103,689	Có hướng	14.57
Amazon	262,111	1,234,877	Có hướng	4.71
Google web	875,713	5,105,039	Có hướng	5.83

Bảng 3.3 Các bộ dữ liệu

được sử dụng trong các bài toán lan truyền thông tin [28], [?], [?]. Chúng tôi so sánh đánh giá trên ba yếu tố chính: chất lượng lời giải, tính mở rộng, tác động của số vòng lan truyền d trên một vài các MXH trực tuyến thực tiễn. Các thuật toán cơ sở sẽ sử dụng bao gồm:

- Random: Thuật toán chọn ngẫu nhiên k đỉnh trong $N_d(S)$.
- Maxdegree (Maxdeg): Thuật toán heuristic chọn lần lượt các đỉnh có bậc cao nhất trong tập V_d cho đến khi đủ k đỉnh.
- Greedy: Thuật toán heuristic chọn lần lượt k đỉnh, trong mỗi lần chọn sẽ chọn ra đỉnh sao cho khi loại bỏ đỉnh đó sẽ cứu được nhiều đỉnh nhất có thể.

Chúng tôi thực nghiệm sử dụng một hệ thống máy có cấu hình sau: Intel(R) Core(TM) i5-6200U CPU @ 2.30 GHz (up to 2.40 GHz), bộ nhớ RAM 4GB, ngôn ngữ lập trình C++.

- *Dữ liệu:*

Chúng tôi thực nghiệm thuật toán trên nhiều bộ dữ liệu thực tế khác nhau. Bên cạnh việc lựa chọn các bộ dữ liệu để đảm bảo sự đa dạng về kích thước, chúng tôi cũng lựa chọn nhiều miền dữ liệu khác nhau, bảng ... thể hiện các bộ dữ liệu nhóm tác giả sử dụng.

Các thiết lập Chúng tôi sử dụng các thiết lập dưới đây cho việc đánh giá thực nghiệm:

- Trọng số của cạnh: $w(u,v) =$

TÀI LIỆU THAM KHẢO

- [1] Bộ Thông tin và Truyền thông đưa ra biện pháp ngăn chặn thông tin xấu độc trên MXH. Truy xuất từ: <https://news.zing.vn/thong-tin-tieu-cuc-chu-yeu-tren-mang-xa-hoi-cua-nuoc-ngoai-post796383>.
- [2] N.A Christakis, J. H. Fowler (2010), "Social network sensors for early detection of contagious outbreak", PloS one. doi:10.1371/journal.pone.0012948.
- [3] Elias Boutros Khalil, Bistra Dilkina, and Le Song. 2014. Scalable diffusion-aware optimization of network topology. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA, 1226–1235.
- [4] Ceren Budak, Divyakant Agrawal, Amr El Abbadi, Limiting the Spread of Misinformation in Social Networks.
- [5] CV Pham, MT Thai, HV Duong, BQ Bui, HX Hoang. Maximizing misinformation restriction within time and budget constraints. Journal of Combinatorial Optimization, 1-39.
- [6] P.Cui, S. Jin, L.Yu, F.Wang, W.Zhu, S. Yang (2013), " Cascading outbreak prediction in network: a data-driven approach", In: Proceeding of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, pp. 901-909.
- [7] Quy định việc cung cấp, sử dụng Internet của chính phủ. Truy xuất từ: <https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Nghi-dinh-72-2013-ND-CP-quan-ly-cung-cap-su-dung-dich-vu-Internet-va-thong-tin-tren-mang-201110.aspx>.
- [8] Karlova, Natascha, Fisher Karen. A Social Diffusion Model of Misinformation and Disinformation for Understanding Human Information Behaviour. Information Research, 2013, pp. 1-17.

- [9] Tin tặc giả mạo đăng tin sai lệch trên hãng thông tấn Associated Press. Truy xuất từ: <https://www.usatoday.com/story/theoval/2013/04/23/obama-carney-associated-press-hack-white-house/2106757>
- [10] Ảnh hưởng của MXH với bầu cử Pháp. Truy xuất từ: <https://baomoi.com/facebook-va-phep-thu-mang-ten-bau-cu-phap/c/22045463.epi>.
- [11] Tin đồn dịch Ebola bùng phát tại Hà Nội. Truy xuất từ: <http://baophapluat.vn/tin-nong/lam-ro-danh-tinh-2-nguoi-tung-tin-don-that-thiet-ve-benh-dich-ebola-194152.html>
- [12] Tin đồn công ty Formosa xả thải chưa qua xử lí. Truy xuất từ: <http://m.vietnamnet.vn/vn/thoi-su/moi-truong/bo-tn-mt-bac-bo-tin-don-formosa-phat-thai-dioxin-370436.html>
- [13] Phóng sự về nghề điều dưỡng viên xuất khẩu lao động tại Đức. Truy xuất từ: <https://baomoi.com/du-hoc-sinh-duc-phan-doi-phong-su-cua-vtv-ve-nghe-dieu-duong-vien-luong-thang-100-trieu/c/21365466.epi>
- [14] H Zhang, M. Alim, M. T. Thai, and H. Nguyen, Monitor Placement to Timely Detect Misinformation in Online Social Networks, in Proceedings of the 2015 IEEE International Conference on Communications (ICC), 2015.
- [15] D. Kempe, J. Kleinberg, and E. Tardos. 2003. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. Washington, DC, USA, 137-146. <https://doi.org/10.1145/956750.956769>.
- [16] D. Kempe, J. Kleinberg, and E. Tardos. Influential Nodes in a Diffusion Model for Social Networks. In ICALP, 2005, pp. 1127-1138.

- [17] J. Goldenberg, B. Libai, and E. Muller. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, 2001, pp. 211-223
- [18] T. Carnes, R. Nagarajan, S. M. Wild, and A. V. Zuylen. Maximizing Influence in a Competitive Social Network: a Follower's Perspective. In *Proceedings of the Ninth International Conference on Electronic Commerce*, 2007, pp. 351-360.
- [19] J. Leskovec, M. Mcglohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading Behavior in Large Blog Graphs. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, 2007, pp. 551-556.
- [20] Pedro Domingos and Matthew Richardson. Mining the Network Value of Customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 57-66.
- [21] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. Van-Briesen, and N. S. Glance.: Cost-effective outbreak detection in networks. In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 420-429, ACM (2007).
- [22] A. Goyal, W. Lu, L. V.S. Lakshmanan. Celf++: Optimizing the Greedy Algorithm for Onfluence Maximization in Social Networks. In *Proc. WWW*, 2011, pp. 47-48.
- [23] Wei Chen, Laks V.S. Lakshmanan, and Carlos Castillo. Information and Influence Propagation in Social Networks. A Publication in the Morgan & Claypool Publishers series Synthesis Lectures on Data Management, 2014.
- [24] V. Qazvinian, E. Rosengren, D. R. Radev, and Qiaozhu Mei. Rumor has it: Identifying Misinformation in Microblogs. In *Proc. EMNLP*, 2011, pp. 1589-1599.

- [25] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang. 2013. Prominent Features of Rumor Propagation in Online Social Media. In Proc. ICDM, 2013, pp. 1103-1108.
- [26] D. T. Nguyen, N. P. Nguyen, and M. T. Thai. Sources of Misinformation in Online Social Networks: Who to Suspect. In Proceedings of the IEEE Military Communications Conference, 2012.
- [27] Nam P. Nguyen, Guanhua Yan, My T. Thai (2013), "Analysis of misinformation containment in online social networks", Computer Networks.
- [28] D. T. Nguyen, N. P. Nguyen, and M. T. Thai. Sources of Misinformation in Online Social Networks: Who to Suspect. In Proceedings of the IEEE Military Communications Conference, 2012.
- [29] Ceren Budak, Divyakant Agrawal, Amr El Abbadi, Limiting the Spread of Misinformation in Social Networks.
- [30] Huiling Zhang, Md Abdul Alim, Xiang Li, My T. Thai, and Hien T. Nguyen. 2016. Misinformation in online social networks: Detect them all with a limited budget. ACM Trans. Inf. Syst. 34, 3, Article 18 (April 2016), 24 pages. DOI:<http://dx.doi.org/10.1145/2885494>.
- [31] Zhang Y, Prakash B (2015) Data-aware vaccine allocation over large networks. ACM Trans Knowl Discov Data. <https://doi.org/10.1145/2803176>.
- [32] Zhang H, Alim M, Li X, My TT, Nguyen H (2016a) Misinformation in online social networks: catch them all with limited budget. ACM Trans Inf Syst 34(3):18.
- [33] Zhang Y, Adigay A, Saha S, Vullikanti A, Prakash A (2016b) Near-optimal algorithms for controlling propagation at group scale on networks. IEEE Trans Knowl Data Eng. <https://doi.org/10.1109/TKDE.2016.2605088>
- [34] Yadron D (2017) Twitter deletes 125,000 Isis accounts and expands anti-terror teams. The Guardian Web.

- <https://www.theguardian.com/technology/2016/feb/05/twitter-deletes-isis-accounts-terrorismonline>. Accessed 24 June 2017.
- [35] Kottasov I (2017) Facebook targets 30,000 fake accounts in France. CNN media Web. <http://money.cnn.com/2017/04/14/media/facebook-fake-news-france-election/index.html>. Accessed 24 June 2017.
- [36] Zaixin Lu, Wei Zhang, Weili Wu, Joonmo Kim, and Bin Fu. 2011. The complexity of influence maximization problem in the deterministic linear threshold model. *Journal of Combinatorial Optimization* 24, 3 (April 2011), 374–378. <https://doi.org/10.1007/s10878-011-9393-3>.
- [37] Vijay V. Vazirani. 2001. *Approximation Algorithms*. Springer, Verlag New York, Inc. New York, NY, USA.