**Các thuật toán thích nghi để phát hiện cấu trúc cộng đồng trong mạng xã hội năng động**

Nam P. Nguyen, Thang N. Dinh, Ying Xuan, My T. Thai

Department of Computer and Information Science and Engineering, University of Florida

Email: {fnanguyen, tdinh, yxuan, mythaig}@cise.uﬂ.edu

***Abstracts(tóm tắt)* – Mạng xã hội có tính chất rất đặc biệt: cấu trúc cộng đồng. Sự hiểu biết về cấu trúc cộng đồng mạng là một lợi thế lớn. Nó không chỉ cung cấp những thông tin hữu ích trong phát triển thêm về chiến lược nhận thức xã hội (social-aware) cho các vấn đề mạng xã hội mà còn hứa hẹn một loạt ứng dụng kích hoạt bằng mạng điện thoại, chẳng hạn như định tuyến trong mạng di động Ad Hoc (MANETs) và luồn sâu vào các chính sách ngăn chặn của mạng di động. Thật tiếc, sự hiểu biết về cấu trúc này là một thách thức lớn, đặc biệt là trong các mạng xã hội năng động nơi mà các hoạt động và tương tác xã hội đang phát triển một cách nhanh chóng. Chúng ta có thể xác định cấu trúc cộng đồng mạng nhanh chóng và hiểu quả không? Chúng ta có thể cập nhật cấu trúc mạng dựa trên những thông tin đã được biết trước thay vì tính toán lại từ đầu không? Trong bài báo này, chúng tôi trình bày về Quick Community Adaptation(QCA) – Sự thích ứng cộng đồng nhanh chóng, một sự thích ứng dựa trên định nghĩa và chỉ ra cấu trúc cộng đồng của mạng xã hội trực tuyến năng động. Cách tiếp cận của chúng tôi không chỉ nhanh chóng mà còn cập nhật hiệu quả cộng đồng mạng, thông qua một loạt các thay đổi, bằng cách chỉ sử dụng các cấu trúc xác định từ các ảnh chụp (snapshots) mạng trước đó, nhưng lại có khả năng đưa ra sự thay đổi của cấu trúc cộng đồng theo thời gian. Để làm rõ tính hiệu quả của thuật toán, chúng tôi kiểm tra tổng quát QCA trên mạng xã hội thực tế bao gồm email, arXiv và Facebook. Cuối cùng chúng tôi chứng minh khả năng áp dụng của thuật toán qua một ứng dụng thực tế trong MANETs. Kết quả so sánh cho thấy rằng: chiến lược định tuyến social-aware sử dụng QCA như một phát hiện cộng đồng mạng tốt hơn phương pháp có sẵn hiện nay.**

# **Giới thiệu**

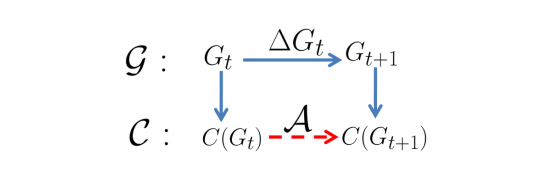
Nhiều mạng xã hội thể hiện đặc điểm của sự bao hàm cấu trúc cộng đồng[1][2], có nghĩa là : chúng tự phân chia thành các nhóm đỉnh với các kết nối dày hơn bên trong mỗi nhóm và ít kết nối hơn giữa các nhóm giao nhau, nơi mà đỉnh và các kết nối đại diện cho mạng người dùng(network users) và tương tác xã hội của họ. Các thành viên trong mỗi cộng đồng mạng xã hội thường chia sẻ những điểm chung như ảnh, phim, âm nhạc hoặc các chủ đề tranh luận và do đó, họ có xu huớng giao tiếp thường xuyên với nhau hơn là giao tiếp với các thành viên bên ngoài cộng đồng của họ. Việc phát hiện các cộng đồng trong mạng lưới là việc tập hợp các đỉnh của mạng lại thành các nhóm bằng cách các nút (nodes) trong mỗi nhóm được kết nối nhiều bên trong và ít bên ngoài.

Cần lưu ý để phân biệt giữa phát hiện cộng đồng mạng và phân nhóm đồ thị. Hai vấn đề này cùng có một mục tiêu là phân vùng các nút mạng (network nodes) thành các nhóm; tuy nhiên, số lượng các nhóm được định nghĩa trước hoặc đưa ra một phần của đầu vào phân nhóm đồ thị trong khi các cộng đồng thường chưa biết trong việc phát hiện các cộng đồng. Phát hiện các cộng đồng trong mạng lưới cung cấp cho chúng ta sự hiểu biết hữu ích về cấu trúc bên trong cũng như các nguyên tắc tổ chức của nó. Hơn nữa, hiểu biết về cấu trúc của cộng đồng mạng có thể cung cấp cho chúng ta những điểm hữu ích hơn về một số phần chưa được khám phá của mạng, do đó giúp ngăn ngừa các mối nguy hiểm tiềm năng như virus hoặc sự lan truyền mầm bệnh. Nghiên cứu dựa trên phát hiện cộng đồng trong mạng tĩnh (static networks) cs thể được tìm thấy trong một cuộc khảo sát [3], cũng như công việc của [4][5][6][7] và sự tham khảo trong đó.

Tuy nhiên, các mạng xã hội ở thế giới thực không phải lúc nào cũng tĩnh (static). Trong thực tế, phần lớn các mạng xã hội (như Facebook, Bebo và Twitter) phát triển, mở rộng về kích thước và không gian như sự gia tăng người dùng của họ, do đó tự giúp đỡ bản thân họ trong lĩnh vực mạng lưới năng động. Một mạng lưới năng động là một loại tiến hóa đặc biệt của mạng lưới phức tạp trong đó thường xuyên thay đổi theo thời gian. Trong khả năng phán đoán của một mạng xã hội trực tuyến, ví dụ như Facebook, Twitter hoặc Flickr, các thay đổi thay đổi thường được giới thiệu bởi những người dùng gia nhập vào hoặc thu hồi từ một hoặc nhiều nhóm hoặc các cộng đồng, qua bạn bè và bạn bè kết nối cùng nhau hoặc qua người mới kết bạn với những người khác. Bất kì những khả năng này có vẻ như ảnh hưởng nhỏ tới cấu trúc địa phương (local structure) của một mạng lưới, sự năng dộng của mạng lưới trong thời gian dài, mặt khác, có thể dẫn tới một sự thay đổi đáng kể trong cấu trúc cộng đồng mạng, vì vậy đặt ra nhu cầu xác định lại. Tuy nhiên, sự thay đổi nhanh chóng và không thể xác định trước của mạng xã hội năng động khiến nó trở thành một vấn để vô cùng phức tạp nhưng đầy thách thức.

Mặc dù, người ta có thể chạy bất kì phương thức phát hiện cộng đồng tĩnh (static community) nào, [4][5][6][8], để tìm thấy những cấu trúc cộng đồng mới bất cứ khi nào mạng lưới được cập nhật, tuy nhiên có thể gặp phải một số nhược điểm không thể bỏ qua: (1) thời gian hoạt động dài của phương thức tĩnh cụ thể trên mạng lưới lớn, (2) the trap of local optima và (3) hầu như cùng phản ứng lại những thay đổi nhỏ với một phần của mạng lưới (the almost same reaction to a small change to some local part of the network). Cách tốt hơn, hiệu quả và ít tốn thời gian hơn để thực hiện công việc này là cập nhật thích nghi (adaptively update) cộng đồng mạng từ những cấu trúc đã biết trước, điều này giúp tránh tranh cãi của việc tính toán lại từ đầu. Phương pháp giải quyết thích nghi này là phần trọng tâm chính của nghiên cứu trong bài báo này. Trong Hình 1, chúng tôi tổng quát hóa ý tưởng sự thích nghi cấu trúc mạng lưới cộng đồng năng động.

Phát hiện cấu trúc cộng đồng của một mạng xã hội năng động là một lợi ích đáng kể. Để đưa ra ý nghĩa của nó, hãy để ý đến định tuyến vấn đề trong mạng lưới (communication network) nơi các nút (nodes) và các liên kết (links) trình bày về giao tiếp người và di động (mobile) tương ứng.



Hình 1: Mạng tiến triển từ thời gian *t* tới *t*+1 dưới sự thay đổi Δ*Gt*. Thuật toán thích nghi A nhanh chóng tìm được cấu trúc cộng đồng mới C (Gt + 1) dựa trên cấu trúc trước đó C (Gt) cùng với sự thay đổi Δ*Gt*.

Do tính di động của các nút (nodes) và tính không ổn định của các liên kết (links) của mạng lưới, việc thiết kế mô hình định tuyến hiệu quả là cực kì thách thức. Tuy nhiên, từ khi con người có xu hướng hình thành các nhóm giao tiếp tự nhiên, tồn tại các nhóm nút kết nối bên trong dày hơn bên ngoài nằm dưới MANET (Mobile Ad-hoc Network) như một sự phản chiếu, vì thế, hình thành cấu trúc cộng đồng trên MANET này. Một thuật toán định tuyến hiệu quả, ngay khi nó phát hiện ra cấu trúc cộng đồng mạng, có thể định tuyến hoặc gửi những tin nhắn ngay lập tức tới các nút cộng đồng tương đương (hoặc các liên quan) như là đích đến. Làm theo các này, chúng ta có thể tránh những tin nhắn thừa (unnecessary mesages) chuyển tiếp thông qua các nút trong các cộng đồng mạng khác nhau, vì vậy có thể giảm số bản sao tin nhắn cũng như giảm chi phí thông tin, điều này là cần thiết trong MANETs.

Các đóng góp của bài báo này:

* Chúng tôi đề xuất QCA – thuật toán nhanh và thích nghi cho việc xác định một cách hiệu quả cấu trúc cộng đồng trong mạng xã hội năng động. Cách tiếp cận này có tính toán đến phát hiện cấu trúc mạng trước đây và chỉ quá trình thay đổi mạng lưới, do đó làm giảm đáng kể chi phí và thời gian xử lý.
* Chúng tôi nghiên cứu tính năng động của mạng xã hội và chứng minh kết quả lý thuyết liên quan, đó là những tính năng cơ bản của phương pháp.
* Chúng tôi đánh giá bao quát các thuật toán trên các mạng xã hội năng động thực tế khác nhau bao gồm: Enron email network, ArXiv citation network (mạng trích dẫn ArXiv) và Facebook. Kết quả thực nghiệm cho thấy phương pháp của chúng tôi không chỉ được cạnh tranh mà còn phát hiện cấu trúc cộng đồng chất lượng cao, một cách kịp thời.
* Như một ứng dụng, chúng tôi dùng QCA như cốt lõi của xác định cộng đồng trong định tuyến chiến lược MANETs. Kết quả mô phỏng cho thấy rằng QCA làm tốt hơn các phương pháp có sẵn hiện tại và xác nhận việc áp dụng phương pháp đề xuất trong điện toán di động (mobile computing).

Phần còn lại của bài báo được tổ chứ như sau: Phần *II*: Giới thiệu sơ bộ và định nghĩa vấn đề. Phần *III*: Đưa ra đầy đủ mô tả các thuật toán và lý thuyết liên quan. Phần *IV*: Cho thấy kết quả thử nghiệm của cách tiếp cận trên các bộ dữ liệu khác nhau trong thực tế. Phần *V*: Giới thiệu một ứng dụng thực tế trong cách tiếp cận MANETs. Phần *VI*: Thảo luận các công việc liên quan và cuối cùng là tổng kết công việc trong phần *VII*.

# Sơ lược