**Nhận dạng cộng đồng trong mạng xã hội đa phương tiện**

**Trừu tượng: Nhận dạng cộng đồng đã nhanh chóng nổi lên như một vấn đề quan trọng trong nhiều năm qua. Mặc dù rất nhiều phương pháp cho vấn đề này đã được đề xuất, xong không có phương pháp nào giải quyết được trực tiếp vấn đề cho các mạng xã hội trực tuyến đa phương tiện, thứ mà một người dùng có thể có nhiều tài khoản với mạng xã hội khác nhau. Trong bài báo này , chúng tôi sẽ giới thiệu và so sánh hai cách tiếp cận là *hợp nhất* và *phương pháp nối* để nhận dạng cộng đồng trong các mạng xã hội trực tuyến đa phương tiện. Hơn thế nữa, chúng tôi phát triển cho mỗi cách một thuật toán NMF riêng. Với mục đích kiểm tra, chúng tôi mở rộng điểm chuẩn LFR đưa kết quả cho mạng xã hội trực tuyến đa phương tiện. Những thí nghiệm chuyên sâu của chúng tôi đưa ra các phương thức cải thiện tốt dựa trên cách tiếp cận sơ khai là tìm kiếm kết cấu cộng đồng trong mỗi mạng riêng.**

1. Giới thiệu

Mạng xã hội trực tuyến đã trở nên phổ biến trong thập kỷ qua. Nhiều mạng xã hội nổi tiếng hiện nay có hàng triêu người sử dụng như twitter, google+ hay hàng tỷ người sử dụng như facebook. Mặc dù về bản chất chúng khác nhau , xong vẫn thể hiện một vài đặc tính chung như là thế giới thu nhỏ, miễn phí và các đặc điểm nổi bật được biết đến như là kết cấu cộng đồng.

Những cộng đồng được định nghĩ trực quan như là nhóm các đỉnh mà có nhiều kết nối đến với nhau hơn phần còn lại của mạng. Ví dụ như, một cộng đồng trong Facebook tương ứng với nhóm người dùng cùng chung sở thích , như nấu ăn , thời trang , âm nhạc vv. Do đó, mục đích của phát hiện cộng đồng được hiểu là chia các mạng thành nhóm các đỉnh. Như vậy, nó mượn lại chính nó để làm đa dạng sự ứng dụng như là các chiến lược chuyển tiếp và định tuyến trong mạng truyền thông. Những cấu trúc cho chúng ta cái nhìn chuẩn xác về cách chức năng của mạng tác động lẫn nhau. Một số lượng lớn phương pháp được đưa ra cho việc nhận diện cộng đồng, tuy nhiên, không có nghiên cứu nào về vấn đề mạng xã hội trực tuyến đa phương tiện, nơi một người dùng có thể tham gia vào nhiều mạng cùng lúc.

Trong mạng xã hội đa phương tiện, để người dùng tham gia nhiều mạng yêu cầu chúng tôi phân tích đồng thời tất cả các mạng. Nhiều kết nối trong một lớp có thể cho thấy mối quan hệ tiềm ẩn với các lớp khác, do đó,cần cung cấp thêm thông tin để làm rõ cấu trúc cơ bản của những mạng này. Như hình minh họa Fig.1 thể hiện nhóm người dùng Facebook, Twitter, LinkedIn, bằng trực quan, chúng ta có 2 tập hợp là {1,2} và {3,4,5,6,7}. Tuy nhiên, nếu tìm kiếm cấu trúc cộng đồng trong mỗi lớp riêng thì chúng ta sẽ thu được một vài cấu trúc cộng đồng khác nhau, thứ mà thường không phải là từ trực giác của chúng ta.

Nhận diện cộng đồng trong mạng xã hội trực tuyến đa phương tiện cũng đem đến một vài thách thức. Thứ nhất, mạng xã hội trực tuyến đa phương tiện thường không đồng nhất, chúng có thể có hướng và vô hướng, quá tải và không quá tải hoặc có nhiều tỷ lệ mật độ khác nhau. Ngoài ra, sự đa dạng hình học còn làm cho bài toán càng phức tạp thêm.

Mặc dù có nhiều nghiên cứu dành cho nhận diện cấu trúc cộng đồng xong cấu trúc cộng đồng trong mạng xã hội trực tuyến đa phương tiện thì chưa được giải quyết. Những nghiên cứu gần nhất như nhận diện cấu trúc cộng đồng trong mạng đa phương tiện thì không thể ứng dụng trực tiếp cho những mạng xã hội đa phương tiện được. Lý do nằm ở chỗ các mạng xã hội trực tuyến đa phương tiện chỉ có duy nhất một kiểu thực thể, mỗi thực thể đại diện cho một số lớp và những cách tiếp cận hiện có lại bỏ qua hiện tượng quan trọng này.

Trong bài báo này, chúng tôi chỉ ra vấn đề nhận dạng kiến trúc cộng đồng trong mạng xã hội trực tuyến đa phương tiện, nơi mà các mạng có thể có hướng hoặc vô hướng, bị quá tải hoặc không . Những đóng góp chúng của chúng tôi là:

1. Chúng tôi giới thiệu và so sánh 2 lớp tiếp cận. Lớp thứ nhất tên là tiếp cận thống nhất, tìm kiếm một cấu trúc cộng đồng thích hợp bằng việc kết hợp nhiều tài khoản của cùng người dùng trong các mạng. Lớp thứ 2 là tìm kiếm các cấu trúc cộng đồng thích hợp nhất trong mạng sử dụng kĩ thuật khớp nối. Chúng tôi cũng phát triển phương thức NMF chuyên biệt cho mỗi lớp.
2. Chúng tôi mở rộng điểm chuẩn LFR[7] để tạo một điểm chuẩn mới cho việc nhận dạng cộng đồng trong mạng xã hội trức tuyến đa phương tiện. Phần mới mở rộng có khả năng tạo ra những lớp với phân chia các bậc điểm khác nhau và phần ở giữa liên kết trong và ngoài cộng đồng.
3. Chúng tôi thực hiện nhiều thí nghiệm chuyên sâu từ nguồn dữ liệu được tổng hợp. Nhiều kết quả thu được cho thấy cách tiếp cận của chúng tôi là tốt hơn cách ban đầu trong việc tìm ra cấu trúc cộng đồng tại mỗi mạng riêng rẽ.

**Đánh giá công việc**. Nhận diện cấu trúc cộng đồng thu hút được sự chú ý lớn từ ngành khoa học mạng. Nguồn cảm hứng trong công việc của Newman và Girvan , các thuật toán khác nhau đã được phát triển cho nhiều vấn đề như clique-based, degree-based, matrixpertubation-based [8 ] theo quan điểm về mạng xã hội. Louvain và Infomap là các phương pháp hàng đầu được ưa chuộng theo tiều chuẩn trong [7]. Khoa học cộng đồng đã đang phát triển các công cụ về mạng thời gian, mặc cho vẫn còn nhiều việc phải làm. Sự gia tăng một lượng lớn các nghiên cứu với mức độ chuyên môn khác nhau đã gây sự chú ý đến việc nghiên cứu lớp mạng đa phương tiện.

Thừa số hóa ma trận không âm(NMF) đã được giới thiệu lần đẩu bởi Paatero và Tapper, phổ biến bới Lee và Se-ung. Ý tưởng chính đó là lấy gần đúng một ma trận không âm V bằng tích của hai ma trận không âm khác *W và H*. Do tính không âm của thừa số nên những đã triển khai một số lượng lớn nghiên cứu …, khai phá văn bản, phân tích dữ liệu quang phổ, giảm nhiễu tiếng nói, tin sinh học và nhiều hơn nữa. Hiện nay, trong khoa học xã hội, Lin el al. giơi thiệu MetaFac[8] sử dụng mối quan hệ giữa siêu đồ thi và thừa số ten xơ. Sau đó,Wang el al. đã sử dụng thuật toan NMF và đưa ra 3 cách giải NMF cho mạng vô hướng, có hướng và phức tạp với nhiều thuộc tính khác nhau(người dùng và phim ảnh). Không may khi phương pháp này không tương thích mạng xã hội trực tuyến đa phương tiện vì chúng có cả kiểu thực thể đơn và đa trong nhiều mạng khác nhau.

Ngày nay, một vài vấn đền về mạng xã hội đơn được tổng quát hóa thành mạng đa thiết lập. Brokda et al. giới thiệu 2 thuật toán để giải quyết đường đi ngắn nhất trong mạng xã hội nhiều lớp. Quá trình truyền tin thông qua nhiều mạng được xem xét tai [16]. Kazienko et al. đã nghiên cứu mô hình mạng xã hội đa phương tiện và ứng dụng trong các hệ thống xã hội [17]. Trong [18], các tác giả đã nghiên cứu việc kết hợp người dùng giữa các vấn đề mạng xã hội. Tuy nhiên, nhận dạng cấu trúc cộng đồng trong mạng xã hội trưc tuyến đa phương tiện không được nghiên cứu một cách hệ thống.

**Cách tổ chức.** Phần còn lại của bài báo bao gồm. phần II định nghĩa vấn đề và giới thiệu 2 lớp phương pháp. Phần III và IV đưa ra chi tiết 2 lớp tiếp cận bao gồm công thức , bổ sung các luật và chứng minh hội tụ. Kết quả thực nghiệm có ở mục V và một số kết luận nhận xét trong mục VI.

II . XÂY DỰNG BÀI TOÁN

Chúng tôi mô hình hóa mạng xã hội trực tuyến đa phương tiện như là tập các đồ thị *G* .*G* bảo gồm lớp mạng đơn *p*. Lớp i viết tắt là *Gi = (Vi, Ei)* trong đó *Vi và Ei* là tập các đỉnh và các cạnh tương ứng với mỗi lớp. Chú ý là mỗi đỉnh có thể xuất hiện trong một hoặc nhiều lớp. Chúng tôi định nghĩa tập *V = Upi =1 Vi và n = |V| - độ lớn tập V*. Hiện tại , ta mô tả mỗi lớp trong ma trận Ai là một n \* n ma trận liền kề của *Gi*. . Một lớp 3 mạng xã hội trực tuyến được minh họa tại hình 1.

Giả sử tồn tại *k* cộng đồng trong lớp *i*. Chúng tôi mô hình hóa sự tương tác (Ai)uv giữa các điểm u và v trong lớp i bằng một hỗn hợp mô hình kết hợp do tất cả k cộng đồng. Có nghĩa là chúng ta ước chừng (Ai)uv sử dụng (Ai)uv = với pml là ước lượng tỉ trọng giữa các cộng đồng m và l, pm -> u và pl ->v là xác suất mà một tương tác với cộng đồng m và l liên quan đến nút u và v , tương ứng. Ta có Ai  = XiSiXti  trong đó Xi là ma trận không âm với (Xi)um = pm->u , Si cũng là một ma trận không âm với (Si)ml = PmL . Mục tiêu của chúng tôi đó là tìm ra được cấu trúc cộng đồng, cái có thể tương ứng với một n\*k ma trận Xi cho mỗi lớp i lớp mà mỗi hàng thể hiện mối quan hệ chung của một người dùng. (Xi)um phản án khả năng tham gia của người dùng u đến cộng đồng m. Sự miêu tả này có thể được sử dụng cho độ chập hoặc phân chia cấu trúc cộng đồng. Ở phần dưới, phân chia cấu trúc cộng đồng là nội dụng chính của bài bao.

**Giả thuyết chính.** Nếu điểm u , v cùng một nhóm lớp, chúng cũng có khả năng thuộc về một nhóm của các lớp khác. Dựa trên lý luận chặt chẽ chúng tôi thực thi giả định này và chia làm 2 lớp tiếp cận:

*Tiếp cận thống nhất:* Chúng tôi buộc các trường hợp người dùng trong những lớp khác nhau được gộp chung một cộng đồng bằng việc tập hợp tất cả các lớp bên trong một mạng đơn, nơi mà nhiều trường hợp xuất hiện như là một đỉnh trong mạng kết hợp.

*Tiếp cận khớp nối:* Chúng tôi nới lỏng việc thực thi bằng cách sử dụng sơ đồ khớp nối. Thay vì buộc các trường hợp của một đỉnh, chúng tôi đề nghị chúng được ở cùng một cộng đồng bằng việc tạo ra một cạnh nối giữa các cặp trường hợp nhau.

III Tiếp cận thống nhất

Trong phần này, chúng tôi giải thích cách tiếp cận thống nhất để tìm kiếm một cấu trúc cho tất cả các lớp. Chúng tôi xem xét theo 2 hướng: 1) Chuyển nhiều lớp thành một lớp mạng và ứng dụng các thuật toán hiện có. 2) Phỏng theo thuật toán NMF trên các mạng ban đầu.

1. *Mạng kết hợp*

Ứng dụng các thuật toán cho mạng xã hội trực tuyến đa phương tiện, chúng tôi cần: 1) Kết hợp tất cả các lớp trong một mạng đơn Gc,2) Ứng dụng những thuật toán cấu trúc cộng đồng… Louvain[9], Infomap[10], để tìm cấu trúc cộng đồng trong Gc 3) Gc đã tìm được cấu trúc cộng đồng thì quay trở lại mỗi lớp tìm cấu trúc cộng đồng của chúng

Một lớp mạng đa chiều G được định nghĩa trong phần xây dựng bài toán, mạng kết hợp[12] được ký hiệu là *Gc = (V,Ec) với Ec = (E1 U E2 U … Ep0 và E1, E2, …* là cá cạnh của lớp.

Hiện tại chúng tôi có thể đảm bảo sử dụng thuật toán cho các mạng đơn trên mạng kết hơp. Tuy vậy, sự kết hợp bộc lộ một vài thiếu sót của nó ..,những loại cạnh trong các lớp có thể khác nhau, một vài lớp thì quá tải , một vài thì không. Đặc điêm đó làm chúng khó kết hợp các lớp. Do đó, chúng tôi giới thiệu thuật toán NMF dựa trên lớp mạng đa phương tiện ban đầu.

1. *Thuật toán NMF dựa trên các mạng ban đầu*

Chúng tôi đưa ra thuật toán NMF cho cả mạng có hướng và vô hướng.

1. Mạng có hướng: Chúng tôi cố gắng tìm một thành phần trong ma trận X để kết

hợp với cấu trúc các mạng đã nhận. Đặc biệt, chúng tôi muốn cực tiểu tổng khác nhau giữa XSXT và các ma trận Ai, I = 1…p. Ở đây S cho ta kết nối giữa các cộng đồng. Sau đó, vấn đề nhận diện cộng đồng có thể đúc kết thành một bài toán thừa số hóa ma trận không âm. Do đó, chúng tôi có được hàm sau:



ở đây d(A||B) đo sự khác biệt giữa 2 ma trận. Công thức đó được gọi là bình phương khoảng cách Eclip [13]



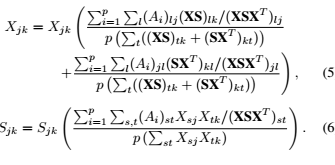
Tương tự, khoảng cách thứ 2 tên là khoảng cách Kullback-Leibler [13] của A từ B được định nghĩa:



1. Sử dụng khoảng cách KL: Giá của hàm sử dụng khoảng cách như sau:



Chúng tôi lấy được các quy tắc cập nhật theo bộ khung trong [13]



Thuật toán 1 : thuật toán NMF cho các mạng có hướng sử dụng khoảng cách KL

**Đầu vào**: Ma trận {Ai| I = 1…p}, tối đa T lần lặp

**Đầu ra:** Thành viên của ma trận X  
gán Xij , Sij cho giá trị ngẫu nhiên {0,1}

**Lặp lại**

Cập nhật Xjk theo cthuc 5

Cập nhật Sij theo công thức 6

**cho đến khi** Hội tụ hoặc sau T lần lặp

Với mỗi hàng I, argmaxj {Xij} là cộng đồng mà đỉn I được gán.

Trả về danh sách cộng đồng tương ứng với các đỉnh.

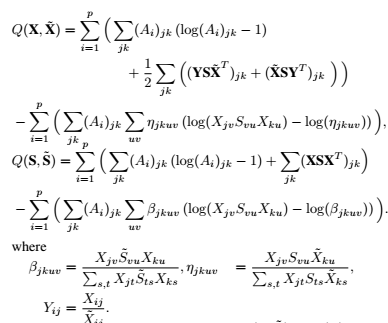
Alg.1 mô ta thuật taosn cho mạng có hướng sử dụng khoảng cách KL trong tiếp cận thống nhất. Phân khúc chính là việc cập nhật phương pháp, nơi mà Xjk và Sij nhận cập nhận trong mỗi lần lặp cho đến khi hội tụ hoặc sau T lần cập nhật. Hàng I của ma trận X cho thấy sự tham gia của người dùng I trong tất cả cộng đồng. Do đó, chúng tôi gán người dùng i cho cộng động đáp ứng giá trị lớn nhất của Xi nếu như có một vài cộng đồng như vậy thì chọn cái đầu tiên.

Để tìm kiếm k cộng động, chúng tôi dùng phương pháp phổ biến nhất trong [19]. Chúng tôi chọn k là nơi mô đun hàm Q nhận giá trị tối đa.

Lý thuyết 1: Giá trị của hàm đối tượng trong eq.4 là không tăng và đã hội tụ đến mức cực tiểu theo quy tắc được cập nhật ở Eq.5 và 6

Chứng minh: Chúng tôi sẽ chỉ ra rằng X và S hội tụ và điểm hội tụ là 1 cực tiểu địa phương

**Hội tụ:** Để chứng minh hội tụ, chúng ta cần tìm hàm bổ trợ cho X và S để đưa ra cập nhật cho các quy tắc. Chúng ta định nghĩa các hàm bổ trợ là Q(X,X~) và Q(S,S~) như sau:



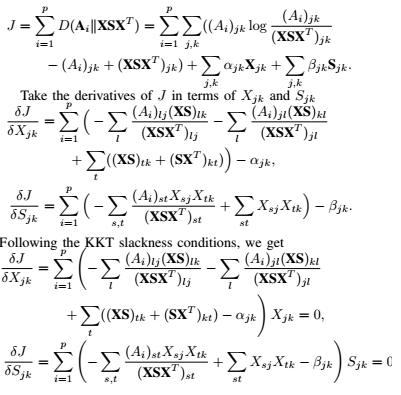
Sau đó, chúng tôi chỉ cần kiểm tra cho *Q*(**X***,* **X** ˜) *> F*(**X**) và *Q*(**S***,* **S** ˜) *> F*(**S**). Phép cộng thứ 2 của các bất đẳng thức là tương đương với



Bất đẳng thức Jensen [13] và hàm logarit lồi. Từ đó, chúng ta có thể kiểm tra *Q*(**S***,* **S** ˜) > *F*(**S**) . Chúng ta cũng có  và làm thỏa mãn bất đẳng thức *Q*(**S***,* **S** ˜) *> F*(**S**). Sau đó , lấy đạo hàm của *Q*(**X***,* **X** ˜) and *Q*(**S***,* **S** ˜), chúng tôi nhận được các luật cập nhật.

**Cực tiểu địa phương:** Chúng ta cần điểm bên ngoài mà việc cập nhật các luật đáp ứng điều kiện KKT [13].

Giới thiệu về phép nhân Lagrangian α*jk* and *βjk* làm mất hàm L, ta có



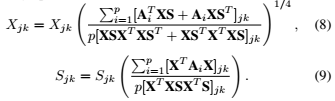
Sau đó, chúng ta có thể nìn thấy rằng việc cập nhật các luật đáp ứng các điều kiện ở trên hoặc X hoặc S sẽ hội tụ đến một tối thiệu địa phương. Từ ma trận Ai, S , và X tất cả đều không âm trong suốt quá trình cập nhật, cuối cùng X và S cũng sẽ không âm.

1. *Sử dụng khoảng cách Eclip*

Chúng ta cũng có thể sử dụng khoảng cách Eclip và có được giá của hàm tương ứng



ĐỊnh lý 2: Giá trị của hàm đối tượng trong ct.7 là không tăng và hội tụ tại một cực tiểu địa phương theo công thức sau



Chúng tôi không chứng minh định lý 2 do không gian không cho phép.

1. *Mạng vô hướng:* bài toán trong mạng vô hướng đó là thực ra là một trường hợp đặc biệt của bài toán có hướng. Ma trận kề cho mỗi lớp là cân đối do đó thừa số Ai = XXT và sau đó công thức cho kết quả :
2. *Phiên bản khoảng cách-*KL:



Với các quy tắc cập nhật chỉ cho ma trận X



1. Phiên bản khoảng cách Eclip:



Thỏa mãn quý tắc cập nhật:



1. Kĩ thuật khớp nối

Để đề nghị các trường hợp của một đỉnh trong mạng đa phương tiện cùng cồng đồng, chúng tôi tạo các cạnh nối giữa chúng và xây dựng mạng đã nối. chúng tôi nghiên cứu bốn lược đồ khớp nối cơ bản [12] , cụ thể là đường chéo, ngôi sao, đầy đủ và phân loại. Trong bài báo[20], các tác giả ứng dụng 2 biến thể của ngôi sao và tổng hợp các khớp nối trong trường hợp bài toán có ít giá ảnh hưởng nhất.Chúng lấy tên là biểu đồ khớp nối losses và lossy cho 2 biển thể, khuôn mẫu được xây dựng bằng cách tạo ra các đỉnh vào như là một lớp trung gian tương tự để ghép sa, trong khi tập hợp được sử dụng cho sau này.Tuy nhiên, họ đã thực hiện một số thay đổi để thích ứng với sự khuếch tán quy trình, nghĩa là xác định trọng lượng và ngưỡng.

*Nối đường chéo*[12]: Cho 2 lớp Gi và Gi+1 ,nếu 2 đỉnh u € Gi và v € Gi+1 thuộc về một thực thể, thì tồn tại một cạnh nối (u,v).

*Nối phân loại[12]* Với mọi cặp lớp Gi và Gj, nếu có 2 đỉnh u € Gi và v € Gi+1 đều thuộc 1 thực thể, thì tồn tại một cặp đỉnh (u,v).

*Nối sao[12]:* Chúng tôi thếm lớp trung gian khác là Gp+1 = (V, Ei ) trong đó Ei  là rỗng và chúng tôi nối mỗi đỉnh trong Gp+1 đến tất cả các đỉnh thuộc cùng thực thể trong mọi lớp khác.

*Nối đầy đủ[12]* Với 2 lớp liền kề Gi = (V,Ei) và Gi+1= (V, Ei+1) , nếu có một đỉnh (u,v) € (Ei U Ei+1), chúng ta có các đỉnh nối(ui, vi+1) trong đó *ui € Gi, vi+1* *€*   
*Gi+1* and (*ui+1, vi*) và *ui+1* *€ Gi*+1*, vi € Gi*.

Với kiến thức về khớp nối, bên cạnh các ma trận Ai với i = 1..p , chúng tôi giới thiệu thêm ma trận Aij đại diện cho nối kết hợp giữa các lớp i và j.

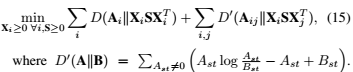
1. Các mạng có hướng

Để tìm cấu trúc cộng đồng trong mạng xã hộ trực tuyến đa phương tiện sử dụng phương pháp khớp nối, chúng tôi thực hiện: 1) xây dựng mạng được nối bằng một biểu đồ khớp nối , sau đó 2) ứng dụng một thuật toán nhận dạng cấu trúc cộng đồng cho mạng lưới kết nối à 3) trích xuất ra cấu trúc cộng đồng.

Sau khi xấy dựng mạng lưới kết nối, chúng ta có thể áp dụng thuật toán hiện có cho mạng đơn mà lợi thế rõ ràng là yêu cầu ít hơn. Chúng ta hãy lấy NMF một ví dụ với giá của hàm :

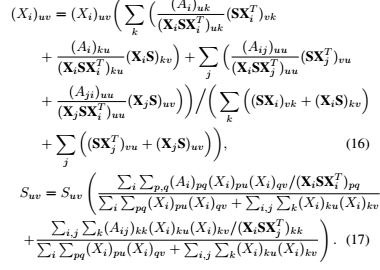


Với A là (n \* p ) \* (n \* p) ma trận kề của mạng nối. Tuy vậy, chúng tôi thấy rằng ma trận A rất thưa bởi vì nó chỉ bao gồm Ai và Ạij . Do đó, Do đó, chúng ta có thể tận dụng lợi thế của cấu trúc đó và có được bài toán NMF sau đây theo khoảng cách KL.



Phép cộng đầu tiên thể hiện cho mỗi lớp riêng , trong khi cái thứ hai đưa ra việc giải thích các khớp nối.

Định lý 3: Giá trị của hàm đối tượng trong eq.15 là không tăng và hội tụ tại một cực tiểu địa phương theo quy tắc cập nhật sau:



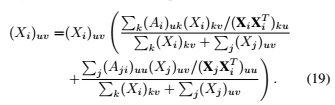
Việc chứng minh cho định lý này khá giống với cách tiếp cận thống nhất nên xin được lược qua.

1. Mạng Vô hướng

Bài toán về mạng vô hướng có thể biểu diễn như sau :



Nó cúng chính là mọt bài toán dạng đặc biển của mạng có hướng khi S là một dạng ma trận . Do đó chúng tôi coi nó theo cùng một cách và có được các quy luật dưới đâu:



V. KẾT QUẢ THÍ NGHIỆM

Phần này chúng tôi so sáng các thuật toán và đồ thì khớp nối . Đặc biệt là chúng tôi sẽ giới thiệu các thiết lập điểm chuẩn LFR để tạo mạng xã hội trực tuyến đa phương tiện. Sau đó chúng tôi sẽ chạy thuật toán NMF và 2 thuật toán tốt nhát cho mạng đơn là Infomap[10] và Louvain[9] trên mạng kết nối khi sự khác nhau các phần mức độ ngoài cộng đồng trên tổng số mức độ mỗi đỉnh và đồng thời thay đổi độ trung bình đỉnh trong mỗi lớp. Chúng tôi cũng kiểm tra thuật toán cho mỗi lớp.

1. *Mở rộng điểm chuẩn LFR*  
   Điểm chuẩn LFR [7] đã được Lancichinetti giới thiệu el .at vào năm 2008 mà … Các thử tục điểm chuẩ ban đầu gồm 3 bước.

1 Gán số bậc cho mỗi đỉnh sao cho tuân theo luật hàm mũ với số mũ cùng cấp

2 Gán các đỉnh đến nhiều cộng đồng, nghĩa là số đỉnh của cộng đồng tuân theo luật phân phối năng lượng với một số mũ nhất định.Ở cùng một thời gian, phương thức xác định chế độ ngoài cộng đồng và trong cộng đồng của mỗi đỉnh để đáp ứng yêu cầu của phân số µ.

3. Vẽ ngẫu nhiên các cạnh với chế độ được xác định rõ.

Không may, điểm chuẩn LFP thì không có khả năng sinh ra mạng xã hội trực tuyến đa phương tiện. LFR cho ra những mạng đơn với tổng cấu trúc cộng đồng khác trong mỗi lần nó chạy, thâm chí với thông số như nhau. Do đó, chúng tôi tạo một vài sự thay đổi để hỗ trợ nhiều tính năng trong khi vẫn đảm bảo được đặc tính quan trọng của luật năng lượng.

Bước mà chúng ta có thể thay thế điểm chuẩn LFR là sau bước 2 khi chúng tôi gán đỉnh đến các cộng đồng. Để thay đổi độ đo trung bình đỉnh với mỗi lớp, chúng tôi nhân các đỉnh .. với tỉ lệ độ trung bình mong muốn để tính độ trung bình từ phương pháp. Như vậy, chúng tôi có thể đưa ra kết quả cho nhiều lớp cùng cấu trúc cộng đồng và độ trung bình đỉnh khác nhau trong mỗi lớp.

*B. Tập dữ liệu và thiết lập*

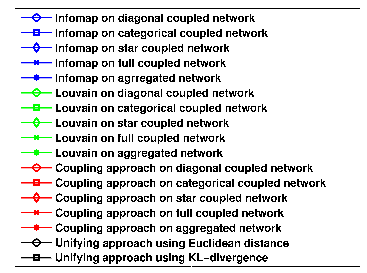
Chúng tôi tạo 4 kiểu mạng những cái được xác định bằng các tính chất có hướng hay quá tải. Với mỗi kiểu mạng, chúng tôi cho ra 5 mạng 3 lớp với 1000 đỉnh, khi độ trung bình đỉnh trong các lớp là (5, 5, 5); (15,15 ,15); (20, 20, 20); (25, 25, 25); (15, 20, 25) tương ứng.

Trong thực tế, chúng tôi không thể biết chính xác liệu hai tài khoản mạng xã hội trức tuyến khác nhau thuộc về cùng môt người dùng. Do đó, nối hai lớp, chúng tôi có thể chỉ kết nối p% so với tất cả đỉnh. Với tập dữ liệu kiểm tra, chúng tôi đưa ra các mạng được nối khi p = 100% và p = 20%.

Tất cả thí nghiệm được tiến hành trên mạng không quá tải vô hướng, kết quả cho 3 loại mạng khác đều giống và đưa vào tài liệu hỗ trợ. Chúng tôi sử dụng hệ thống Linux chạy trên một máy Intel CPU Core Dual 3 GHz, 4 GB RAM .

*C . Kết quả thực nghiệm*

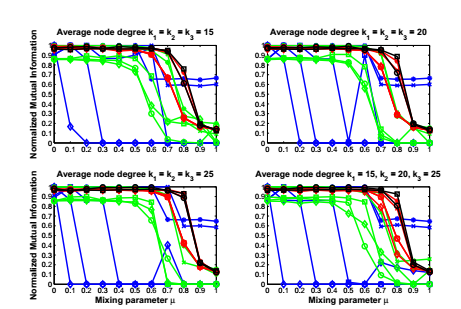
` Chúng tôi sử dụng chú giải sau



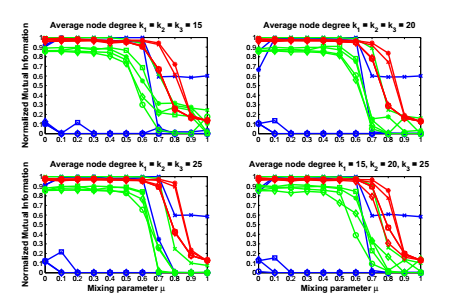
**So sánh các thuật toán** . Hình 2 va 3 thể hiện điểm MNI cho tất cả thuật toán với các độ trung bình đỉnh khác nhau và trộn thông số µ trong mạng không quá tải vô hướng.Phù hợp nhất với cả 3 thí nghiệm , thuật toán NMF luôn luôn nhận NMI cao nhất và vẫn ổn định trong tất cả các cấu hình của mạng. Infomap chủ yếu dựa vào kiểu khớp nối, .. cũng thực hiện tốt như thuật toán NMF trên những mạng kết hợp và mạng nối đầy đủ nhưng rất kém ở các mạng nối khác. Trong khi, Kết quả của Louvain nằm ở giữa 2 phương thức khác trên tất cả bộ dự liệu.

**So sánh giữa khác biều đồ khớp nối:** Quan sát vào cá biểu đồ nối đã sử dụng, chúng tôi thấy rằng những mạng kết hợp và nối đầy đủ hỗ trợ tốt cho tất cả thuật toán , cao nhất phép đo NMI. Các mạng nối đường chéo, nối phân loại và nối sao chỉ phù hợp với thuật toán NMF khi có hành vi giống nhau như chạy trong mạng kết hợp hay nối đầy đủ. Tuy nhiên , Các kết quả các mạng này với phương pháp Infomap không nhanh lắm thậm chí với giá trị thông số trộn là rất nhỏ.

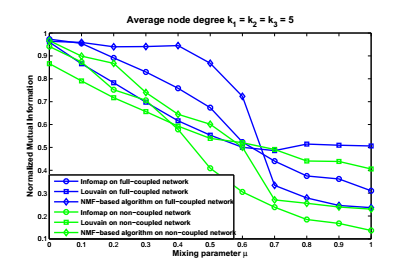
**So sánh Khớp nối và không khớp nối.** Với không không nối , chúng tôi thích tiếp cận bằng việc tìm cấu trúc cộng đồng trong mỗi lớp tương ứng. Chúng tôi chạy 3 thuật toán trên các mạng mà không nối và nối đầy đủ, kết quả ở hình 4. Ta có thể dễ dàng nhìn thấy, với cùng thuật toán , việc chạy trên mạng được nối cho kết quả tốt hơn, đặc biệt là thuật toán NMF. Từ các quan điểm chung, thuật toán NMF cho thấy độ giới hạn độ chính xác của điểm NMI trước khi µ đạt 0.7. Trong khi hình.5 cho thấy các kết quả khi chúng tối thử µ = 0.3 và thay đổi p. Ta nhìn thấy rằng thuật toán NMF dựa trên mạng nối tốt hơn các mạng khác, với Infomap và Louvain , chạy trên các mạng nối tốt hơn không nối khi p >= 0.3



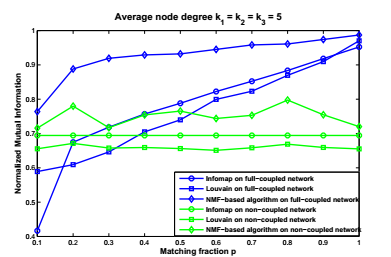
Hình 2: Điểm NMI trên mạng với p = 100%



Hình 3. Điểm NMI trên mạng với p = 20%



Hình.4. Giá trị của nhận dạng với thông số trộn khác nhau (p = 20%)



Hình.5 Giá trị nhận dạng với

Tóm lại, những thuật toán NMF cho kết quả tốt nhất và có thể sử dụng trong tất cả các lớp của mạng xã hội trực tuyến đa phương tiện. Louvain đạt được thống kết tốt và độ chính xác trung bình khi được so sánh với NMF. Mặc dù Infomap làm việc rất tốt với mạng nối đầy đủ và kết hợp nhưng nó không cho không phù với mạng nối đường chéo, sao và phân loại.

VI. NHẬN XÉT KẾ LUẬN

Trong phần này, chúng tôi nghiên cứu vấn đề nhận diện cộng đồng trong mạng xã hội trực tiếp đa phương tiện. Chúng tôi giới thiệu và so sánh hai lớp tiếp cận,là khớp nối và thống nhất, các chúng tôi phát triển một dựa trên thuật toán NMF cho mỗi các tiếp cận. Những thí nghiệm chuyên sâu cho thấy rằng thuật toán NMF thích hợp và nhận các kết quả tốt hơn so với Infoma và Louvain trong điểm chuẩn của chúng tôi. Mặc dù Infomap và Louvain chỉ thực thi tốt trên mạng nối đầy đủ và kết nối xong chúng chạy nhanh hơn thuật toán NMF.

