

MÔ HÌNH LỰC CHO BIỂU DIỄN ĐỒ THỊ PHÂN NHÓM

Trương Quốc Định¹, Taoufiq Dkaki²

¹ Khoa Công nghệ thông tin & Truyền thông, Trường Đại học Cần Thơ

² Institut de Recherche en Informatique de Toulouse

tqding@cit.ctu.edu.vn, dkaki@irit.fr

TÓM TẮT - Đồ thị (graph) là cấu trúc cho phép mô hình hóa nhiều loại dữ liệu phức tạp thuộc nhiều lĩnh vực trong thế giới thực. Bên cạnh đó, đồ thị còn là cấu trúc được sử dụng chủ yếu cho việc biểu diễn thông tin. Khi biểu diễn một lượng lớn thông tin thì việc xác định được các nhóm dữ liệu cũng như mối liên hệ giữa các nhóm là một mục tiêu quan trọng cần đạt được. Trong bài báo này, chúng tôi đề xuất một giải pháp vẽ đồ thị giúp hiển thị một cách rõ nét cấu trúc phân nhóm của dữ liệu cũng như sự liên kết giữa các nhóm. Trong phạm vi nghiên cứu của bài báo này chúng tôi chỉ tập trung vào khía cạnh hiển thị thông tin và giả sử rằng dữ liệu đã được phân nhóm theo một tiêu chí nào đó. Chúng tôi đề xuất giải pháp vẽ đồ thị dựa trên mô hình lực (energy-based model) trong đó các nhóm sẽ được hiển thị trong các vùng riêng biệt và không trùng lặp. Các vùng hiển thị riêng biệt không trùng lặp này có thể do người dùng tự định nghĩa hoặc do giải thuật tự tính toán. Trong cả hai trường hợp, giải pháp do chúng tôi đề xuất đều làm nổi bật được cấu trúc phân nhóm cũng như cấu trúc tổng thể của dữ liệu.

Từ khóa - Đồ thị, đồ thị phân nhóm, vẽ đồ thị.

I. GIỚI THIỆU

Vẽ đồ thị tự động là lĩnh vực nghiên cứu sôi động kể từ nhiều thập niên trở lại đây và trở nên quan trọng hơn rất nhiều khi cấu trúc đồ thị ngày càng được ứng dụng nhiều trong thực tế bởi lẽ nó có thể mô hình hóa cho đa dạng các loại dữ liệu phức tạp. Thật vậy, cấu trúc đồ thị đã chứng minh được tầm quan trọng của mình trong rất nhiều lĩnh vực như: mạng xã hội [1], kỹ nghệ phần mềm [2], thiết kế mạch điện [3], thiết kế cơ sở dữ liệu [4] ... Một cách tổng quát hơn, cấu trúc đồ thị có thể mô hình hóa các loại dữ liệu thể hiện dưới dạng tập các đối tượng và quan hệ giữa các đối tượng đó.

Rất nhiều chiến lược được đề xuất cho việc vẽ đồ thị tự động. Các giải thuật phổ biến nhất đều dựa trên giải pháp tương đối đơn giản đó là mô hình lực có hướng (forced-directed model) [5, 6] và cho kết quả tốt (ít sự giao cắt giữa các cung, cấu trúc cân đối) đối với các đồ thị có kích thước nhỏ (vài trăm đỉnh). Một số giải thuật khác [7, 8] dựa trên quá trình tính toán nhiều pha đã chứng tỏ khả năng thích ứng với các đồ thị có kích thước lớn (vài nghìn đỉnh). Các giải thuật này khá thành công trong việc hiển thị cấu trúc nhóm của đồ thị khi mà các nhóm này được sinh ra một cách “tự nhiên” dựa trên cấu trúc nội tại của đồ thị. Tuy nhiên các giải thuật được xây dựng dựa trên các giải pháp vừa nêu “luôn không thành công” trong việc hiển thị cấu trúc nhóm được xây dựng dựa trên thuộc tính siêu dữ liệu của các đỉnh.

Một yếu tố quan trọng trong việc hiển thị một đồ thị phân nhóm (clustered graph) đó là các đỉnh của cùng một nhóm phải gần nhau và tách biệt với các đỉnh thuộc vào các lớp khác. Vùng bao phủ bởi các nhóm phải không trùng nhau và là các vùng bao lồi (convex hull). Chúng ta có thể tách biệt hai trường hợp: vùng bao phủ của mỗi nhóm được định nghĩa sẵn, vùng bao phủ của mỗi nhóm được tính toán bởi giải thuật. Thật vậy, trong đa phần các trường hợp, vùng bao phủ của mỗi nhóm đã được định nghĩa trước. Ví dụ như trong lĩnh vực thiết kế mạch điện, số lượng lớn các linh kiện thuộc về một board mạch sẽ phải được bố trí trong một vùng không gian giới hạn.

Trong bài báo này, chúng tôi đề xuất một mô hình lực cho phép vẽ đồ thị phân nhóm đảm bảo các điều kiện rằng mỗi nhóm sẽ được hiển thị bên trong một vùng bao lồi phân biệt với vùng hiển thị của các nhóm khác cũng như tối ưu hóa hình thức hiển thị của từng nhóm. Trong phạm vi nghiên cứu này, chúng tôi giả sử rằng các nhóm đã được tính toán và biết trước, có thể là được tính toán bởi một giải thuật phân nhóm đồ thị nào đó (dựa theo cấu trúc liên kết giữa các đỉnh) hoặc tính toán dựa trên dữ liệu thuộc tính gắn kết với các đỉnh của đồ thị. Mô hình chúng tôi đề xuất cũng phải cho kết quả tốt trong cả hai trường hợp: có ràng buộc hoặc không có ràng buộc về vùng hiển thị cho mỗi nhóm. Mô hình mà chúng tôi đề xuất sẽ phải tạo ra được một “bức vẽ” sao cho người dùng có thể sử dụng nó để xem xét và khám phá các yếu tố tiềm ẩn bên trong dữ liệu thông qua việc nhận rõ được các liên kết liên nhóm và các liên kết nội tại trong mỗi nhóm. Mô hình chúng tôi đề xuất là một sự điều chỉnh của mô hình được đề xuất bởi [5] và tiếp nối công trình nghiên cứu được giới thiệu trong [9]. Một trong số những ưu điểm nổi bật của mô hình do chúng tôi đề xuất so với những mô hình khác đó là sự thành công trong việc “tách rời” các nhóm ngay cả trong trường hợp số lượng cung gắn kết các đỉnh trong một nhóm là rất thấp.

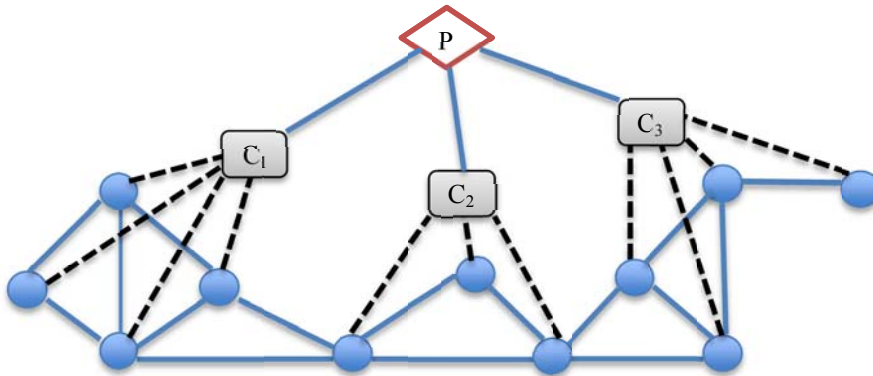
Phần còn lại của bài báo được tổ chức như trình bày sau đây. Trong phần II chúng tôi sẽ nêu định nghĩa về cấu trúc đồ thị phân nhóm và tổng quan các vấn đề về vẽ đồ thị tự động. Mô hình lực cho biểu diễn đồ thị phân nhóm được trình bày trong phần III. Phần IV sẽ mô tả kết quả đạt được của mô hình đề xuất thông qua 3 ví dụ áp dụng. Trong trường hợp có ràng buộc về vùng hiển thị của mỗi nhóm, trước tiên chúng tôi sử dụng dữ liệu về mạng trích dẫn (citation network) rút trích từ [10], thực hiện phân tích Hub và Authority và dùng giá trị Authority của mỗi đỉnh để vẽ đồ thị theo kiến trúc phân tầng. Tiếp theo chúng tôi sử dụng dữ liệu liên quan đến khoảng 1000 hợp đồng [11] mua bán ruộng đất giữa nông dân và các lãnh chúa trong vùng Lot, một vùng nhỏ ở miền Tây Nam nước Pháp và hiển thị chúng

trên bản đồ Kohonen. Cả 2 ví dụ trên đều cho thấy tính hiệu quả của mô hình mà chúng tôi đề xuất. Phần V sẽ thảo luận một số ưu nhược điểm của mô hình đề xuất cùng định hướng nghiên cứu trong tương lai.

II. TỔNG QUAN

A. Cấu trúc đồ thị phân nhóm

Đồ thị G là cặp (V, E) , trong đó V là tập hữu hạn các đỉnh (nút) và $E \subseteq V \times V$ là tập hữu hạn các cung. Đồ thị phân nhóm CG là một bộ ba (V, E, P) , trong đó V là tập hữu hạn các đỉnh (nút), $E \subseteq V \times V$ là tập hữu hạn các cung và P là phân nhóm xác lập trên V . Số phần tử của P chính là số nhóm của G . Hình 1 minh họa cấu trúc đồ thị phân nhóm với số nhóm là 3.



Hình 1. Minh họa đồ thị phân nhóm

B. Các công trình có liên quan

Vẽ đồ thị tự động là vấn đề nghiên cứu của cả lĩnh vực toán học và khoa học máy tính. Trong gần 2 thập kỷ trở lại đây đã có nhiều công trình nghiên cứu được tiến hành. Một vài trong số đó phù hợp với cấu trúc chung của đồ thị như: giải pháp vẽ trực giao (orthogonal drawing) [12], giải pháp vẽ theo đường tròn (circular layout) [13] trong khi một số chỉ phù hợp với cấu trúc đặc biệt như cấu trúc cây [14].

Eades [15] phát triển một trong số các chiến lược vẽ đồ thị được sử dụng nhiều nhất trong cộng đồng với tên gọi mô hình lực. Mô hình này xem đồ thị như là một cấu trúc cơ học trong đó các đỉnh của đồ thị được xem như các viên bi thép và các cung được xem như các lò xo kết nối các viên bi. Các đỉnh sẽ di chuyển theo lực hút/đẩy phát sinh bởi các lò xo và vị trí của các đỉnh tại trạng thái cân bằng (có mức năng lượng nhỏ nhất) chính là hình vẽ của đồ thị. Dựa trên chiến lược này, Kamada và Kawai [6] cũng như Fruchterman and Reingold [5] đã đề xuất 2 giải pháp được sử dụng nhiều nhất trong cộng đồng. Kamada và Kawai đề xuất lực hút của các lò xo sẽ phải tỷ lệ với khoảng cách đồ thị (độ dài đường đi ngắn nhất) giữa hai đỉnh. Trong khi đó Fruchterman and Reingold bổ sung khái niệm năng lượng điện vào mỗi đỉnh cho phép hai đỉnh sẽ đẩy nhau thay vì chỉ hút nhau khi có cung nối. Bên cạnh đó nhóm tác giả cũng bổ sung khái niệm “cooling temperature” nhằm giảm khoảng cách dịch chuyển tối đa sau mỗi lần lặp dựa trên giả thuyết cách bố trí các đỉnh ở lần lặp sau đã tốt hơn cách bố trí ở lần lặp trước.

Các giải pháp trên, tuy nhiên, không có được sự ổn định trong việc tạo ra các hình vẽ của cùng một đồ thị khi mà ở giai đoạn khởi tạo vị trí của các đỉnh được khởi tạo một cách ngẫu nhiên. Gajer [16] đề xuất giải pháp bao gồm nhiều pha dựa trên phép phân tích tập con độc lập lớn nhất (maximal independent set filtration) trong đó tập nhỏ nhất chỉ bao gồm ba phần tử. Ở mỗi pha chỉ có các đỉnh không thuộc tập các đỉnh đã xác định vị trí ở pha liền trước được khởi tạo vị trí và dịch chuyển. Chính chiến lược này đã đẩy nhanh tốc độ thực thi của giải thuật cũng như đảm bảo được tính ổn định của giải thuật khi mà chỉ có ba đỉnh (pha đầu tiên) được khởi tạo vị trí ngẫu nhiên ban đầu.

Trong ngữ cảnh đồ thị phân nhóm, không có quá nhiều công trình nghiên cứu được thực hiện. Các công trình này có thể được xếp vào ba nhóm tiếp cận đặc trưng được đề xuất bởi [17], [18] và [19]. Ho [17] đề xuất phương pháp vẽ đồ thị phân nhóm trong không gian ba chiều dựa trên hướng tiếp cận chia để trị. Mỗi nhóm sẽ được vẽ riêng biệt trong không gian hai chiều và biểu diễn ba chiều của đồ thị phân nhóm sẽ là sự tổ hợp các hình vẽ không gian hai chiều của mỗi nhóm. Chuang [18] đề xuất giải pháp bổ sung một đỉnh giả cho mỗi nhóm cùng lực hút mạnh từ đỉnh này đến các đỉnh khác trong nhóm để thực hiện việc “níu giữ” các đỉnh thuộc cùng một nhóm ở gần nhau. Giải pháp đề xuất được áp dụng cho trường hợp vùng bao hiển thị của mỗi nhóm được người dùng định nghĩa trước. Không như hai nhóm giải pháp vừa trình bày, Balzer [19] đề xuất cách tiếp cận phân cấp cho vấn đề vẽ đồ thị phân nhóm. Cấu trúc phân nhóm của đồ thị (có thể là cấu trúc phân cấp - cây) sẽ được biểu diễn. Tại mỗi thời điểm người dùng có thể chọn “khám phá” cấu trúc bên trong của nhóm tương ứng được chọn.

Giải pháp cho bài toán vẽ đồ thị phân nhóm mà chúng tôi đề xuất phân biệt với các giải pháp trước đó ở hai khía cạnh: cách thức định nghĩa các nhóm cũng như cách thức tác động của lực lên sự dịch chuyển của các đỉnh. Tác giả của các công trình nêu trên định nghĩa nhóm là tập hợp các đỉnh với rất nhiều cung kết nối các đỉnh trong nhóm và rất ít các cung kết nối đến các đỉnh ngoài nhóm. Tuy nhiên điều này không phải lúc nào cũng đúng trong thực tế vì các đỉnh

có thể được nhóm lại theo một ràng buộc bất kỳ nào đó chứ không phải chỉ dựa trên bản chất mối liên hệ giữa các đỉnh. Ví dụ như đối với mạng trích dẫn (citation network), các nhóm có thể được xây dựng dựa trên sự tương đồng về mặt địa lý giữa các đỉnh (tác giả của các bài báo thuộc cùng một trường, một quốc gia ...). Và trong trường hợp như thế thì các giải pháp đề xuất ở trên không cho kết quả mỹ mãn. Chúng tôi cũng tin rằng, một bản vẽ “đẹp” cho đồ thị phân nhóm cần đáp ứng các yêu cầu sau:

- Mỗi một nhóm phải được bố trí trong một vùng bao lồi
- Vùng bao lồi của các nhóm không bao phủ nhau
- Giảm đến mức tối thiểu sự giao cắt giữa các cung kết nối các đỉnh trong cùng một nhóm
- Giảm thiểu sự trùng lặp về vị trí giữa các đỉnh

III. MÔ HÌNH ĐỀ XUẤT

Mô hình mà chúng tôi đề xuất là sự điều chỉnh của mô hình lực đề xuất bởi [5] và cũng bao gồm nhiều lần lặp với hai bước chính ở mỗi lần lặp: tính toán lực hút và lực đẩy cho mỗi cặp đỉnh, tính khoảng cách và hướng dịch chuyển của mỗi đỉnh dựa trên khoảng cách dịch chuyển tối đa. Chúng tôi cũng sử dụng khái niệm “cooling temperature” đã trình bày trong [5] để giảm khoảng cách dịch chuyển tối đa qua mỗi lần lặp. Ý tưởng của việc giảm giá trị khoảng cách dịch chuyển này đó là “bản vẽ” sẽ tốt hơn sau mỗi lần lặp vì thế việc dịch chuyển sẽ phải nhỏ dần sau mỗi lần lặp.

Điểm điều chỉnh quan trọng so với [5] đó là chúng tôi phân biệt 2 loại lực: lực tác động bởi đỉnh thuộc cùng một nhóm và lực tác động bởi đỉnh khác nhóm đồng thời bổ sung lực hút giữa các đỉnh không có cung nối của cùng một nhóm (cung giả).

- Nội lực: lực hút/đẩy giữa các đỉnh thuộc cùng một nhóm
- Ngoại lực: lực hút/đẩy giữa các đỉnh khác nhóm
- Lực giả: lực hút giữa các đỉnh không có cung nối thuộc cùng một nhóm.

Các lực được định nghĩa thông qua khái niệm khoảng cách tối ưu. Khoảng cách tối ưu là khoảng cách giữa các đỉnh không có cung nối và được định nghĩa như sau

$$optDist = \sqrt{\frac{\text{kích thước vùng hiển thị}}{\text{số đỉnh của đồ thị}}}$$

Nếu gọi f_a , f_r lần lượt là lực hút và lực đẩy và d là khoảng cách euclide giữa hai đỉnh, khi đó f_a và f_r có thể định nghĩa như sau:

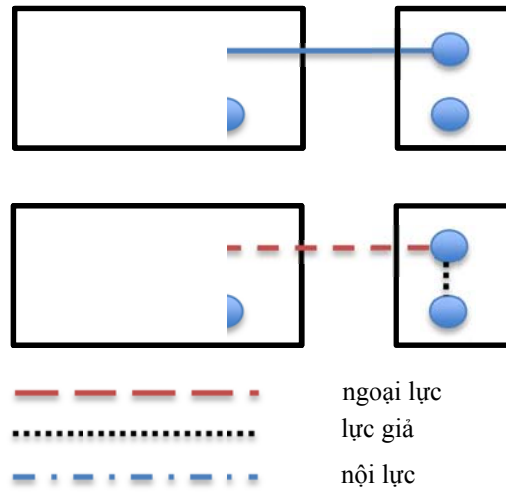
$$f_a(d) = d^2/optDist$$

$$f_r(d) = -optDist^2/d$$

Trong mô hình chúng tôi đề xuất 2 loại lực (nội lực và ngoại lực) vì thế cần định nghĩa 2 loại khoảng cách tối ưu: $optDist$ – khoảng cách tối ưu giữa các đỉnh không thuộc cùng nhóm và $optDistCluster$ – khoảng cách tối ưu giữa các đỉnh thuộc cùng một nhóm. Và chúng tôi cũng tin rằng khoảng cách giữa các đỉnh thuộc cùng nhóm phải nhỏ hơn khoảng cách giữa các đỉnh thuộc các nhóm khác nhau ($optDist = a * optDistCluster$ với $a > 1$) vì kích thước vùng hiển thị của mỗi nhóm sẽ nhỏ hơn kích thước vùng hiển thị toàn bộ đồ thị. Khi chọn $optDist < optDistCluster$ chúng tôi giả thuyết rằng các đỉnh bên ngoài nhóm sẽ ít có tác động đến sự dịch chuyển của một đỉnh hơn là các đỉnh thuộc cùng nhóm. Một vấn đề nữa cũng cần được quan tâm đó là các lực giả (lực hút) sẽ phải có cường độ nhỏ hơn lực hút nội lực.

A. Vùng hiển thị cho nhóm không được định nghĩa trước

Như đã đề cập ở phần trên, trong trường hợp này chúng tôi bổ sung thêm các cung giả gắn kết các đỉnh thuộc cùng một nhóm nhưng không có cung kết nối. Việc bổ sung các cung giả này cho phép níu giữ các đỉnh thuộc cùng một nhóm ở gần nhau. Tuy nhiên, bên cạnh việc thể hiện được cấu trúc phân nhóm của đồ thị thì cấu trúc nội tại của mỗi nhóm cũng phải được quan tâm. Thật vậy cách biểu diễn mỗi nhóm trong đồ thị phân nhóm sẽ là tối ưu nếu như nó được biểu diễn như là một đồ thị riêng lẻ. Để đảm bảo được điều này thì các lực giả cần có trọng số thấp hơn so với các lực thật. Nói một cách khác, việc bổ sung các cung giả sẽ phải không gây quá nhiều tác động đến cấu trúc thật của nhóm. Trong mô hình của chúng tôi, cường độ của các lực giả sẽ được tính toán dựa trên tỷ lệ cung giả so với cung thật của đồ thị. Hình 2 mô tả các loại lực sẽ được tính toán trong quá trình dịch chuyển các đỉnh. Phần trên của hình minh họa đồ thị cần được vẽ với 2 nhóm và các cung nối như hình. Phần hình bên dưới minh họa cho các loại lực hút sẽ tác động lên từng đỉnh khi áp dụng mô hình mà chúng tôi đề xuất. Giữa 2 đỉnh luôn tồn tại lực đẩy cho dù là có hay là không có cung nối, để đơn giản các lực này sẽ không minh họa trong hình 2.



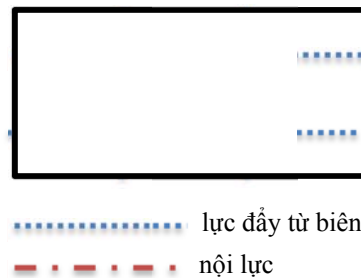
Hình 2. Mô hình lực cho vùng hiển thị không được định nghĩa

B. Vùng hiển thị được định nghĩa trước

Trong trường hợp này thì vùng hiển thị (vùng bao lồi) của mỗi nhóm được định nghĩa trước. Đây là một trường hợp ứng dụng mà hầu như chưa có công trình nghiên cứu nào trước đây đề cập đến. Với trường hợp này thì việc bổ sung các cung giả nhằm tạo nên các lực hút giả cho các đỉnh thuộc cùng một nhóm nhưng không có cung nối là không còn cần thiết. Thay vào đó sẽ có các lực đẩy giữa đường biên của vùng hiển thị với các đỉnh bên trong nhóm tương ứng. Lực đẩy này sẽ mạnh khi khoảng cách từ đỉnh đến đường biên là nhỏ và ngược lại. Tuy nhiên tại cùng một thời điểm, do có nhiều lực đồng thời tác động lên trên một đỉnh nên sẽ dẫn đến trường hợp lực đủ lớn để đẩy đỉnh ra khỏi vùng hiển thị. Để tránh trường hợp này, khi khoảng cách từ đỉnh đến đường biên bằng hoặc nhỏ hơn khoảng cách tối thiểu (khoảng cách dịch chuyển tối đa tại mỗi bước lặp) thì lực đẩy sẽ được nâng lên cực đại. Nếu gọi f_r là lực đẩy của đường biên và d là khoảng cách từ đỉnh đến đường biên thì f_r được định nghĩa như sau:

- $= \infty$ nếu $d < \text{optDistCluster}$
- $= \text{optDistCluster}/k^3$

Hình 3 minh họa việc bổ sung các lực đẩy từ biên



Hình 3. Minh họa tác dụng lực đẩy với đường biên

C. Giải thuật

Tham số đầu vào

Đồ thị phân nhóm CG trong đó các đỉnh của đồ thị đã được phân thành các nhóm khác nhau (bắt buộc).
 Vùng hiển thị (vùng hình chữ nhật xác định bởi đỉnh trên bên trái và chiều dài, chiều rộng) của mỗi nhóm (tùy chọn).

Kết quả

Hình vẽ của đồ thị (tọa độ x, y của mỗi đỉnh của đồ thị)

Giải thuật

→ trường hợp vùng hiển thị không được định nghĩa
 Khởi tạo ngẫu nhiên vị trí của các đỉnh
 → trường hợp vùng hiển thị được định nghĩa
 Khởi tạo vị trí các đỉnh bên trong vùng hiển thị của lớp tương ứng
while temperature > 1 **begin**
 for $v \in V$ **do begin**
 → trường hợp vùng hiển thị được định nghĩa

```

    Tính lực đẩy từ biên lên đỉnh v
    for u ≠ v in V do begin
        tính lực đẩy tác dụng trên v và u
    end
end
for (u, v) ∈ E do begin
    tính lực hút tác dụng lên u và v
end
end
→ trường hợp vùng hiển thị không được định nghĩa
for (u, v) ∉ E và u, v thuộc cùng nhóm do begin
    tính lực hút giả tác dụng lên u và v
end
end
for v ∈ V do begin
    tổng hợp các lực lên v
    dịch chuyển v dựa trên hàm cooling temperature
end
end
giảm giá trị temperature
end

```

Cũng giống như mô hình gốc, mô hình chúng tôi đề xuất là một quá trình lặp bao gồm 2 bước. Trước tiên chúng tôi tính toán tác động của toàn bộ các lực lên trên mỗi đỉnh. Tiếp theo là tính toán khoảng cách dịch chuyển của mỗi đỉnh theo hàm cooling temperature. Hàm cooling temperature xác định khoảng cách dịch chuyển tối đa ở mỗi bước lặp, ban đầu khoảng cách tối đa này có thể được xác lập giá trị bằng với khoảng cách tối ưu giữa các đỉnh, sau mỗi bước lặp khoảng cách dịch chuyển tối đa có thể giảm đi 1. Giá trị temperature đơn giản có thể là số lần lặp của giải thuật.

Quá trình dịch chuyển khởi nguồn từ cấu hình khởi tạo ngẫu nhiên lúc đầu. Đối với trường hợp vùng hiển thị của mỗi nhóm không xác định thì các đỉnh được khởi tạo vị trí ngẫu nhiên trong phạm vi khung vẽ. Ngược lại trong trường hợp vùng hiển thị của mỗi nhóm được định nghĩa trước thì các đỉnh phải được khởi tạo vị trí bên trong vùng hiển thị tương ứng của mỗi nhóm. Ở bước tính toán lực tác động lên mỗi đỉnh, nếu là đỉnh thuộc cùng một nhóm thì sử dụng khoảng cách tối ưu optDistCluster, nếu là trường hợp thuộc hai nhóm khác nhau thì sử dụng khoảng cách tối ưu optDist.

IV. KẾT QUẢ VÀ ĐÁNH GIÁ

Trong phần này, chúng tôi giới thiệu và thảo luận kết quả thông qua ba ví dụ cụ thể để từ đó khẳng định tính khả thi của mô hình mà chúng tôi đề xuất.

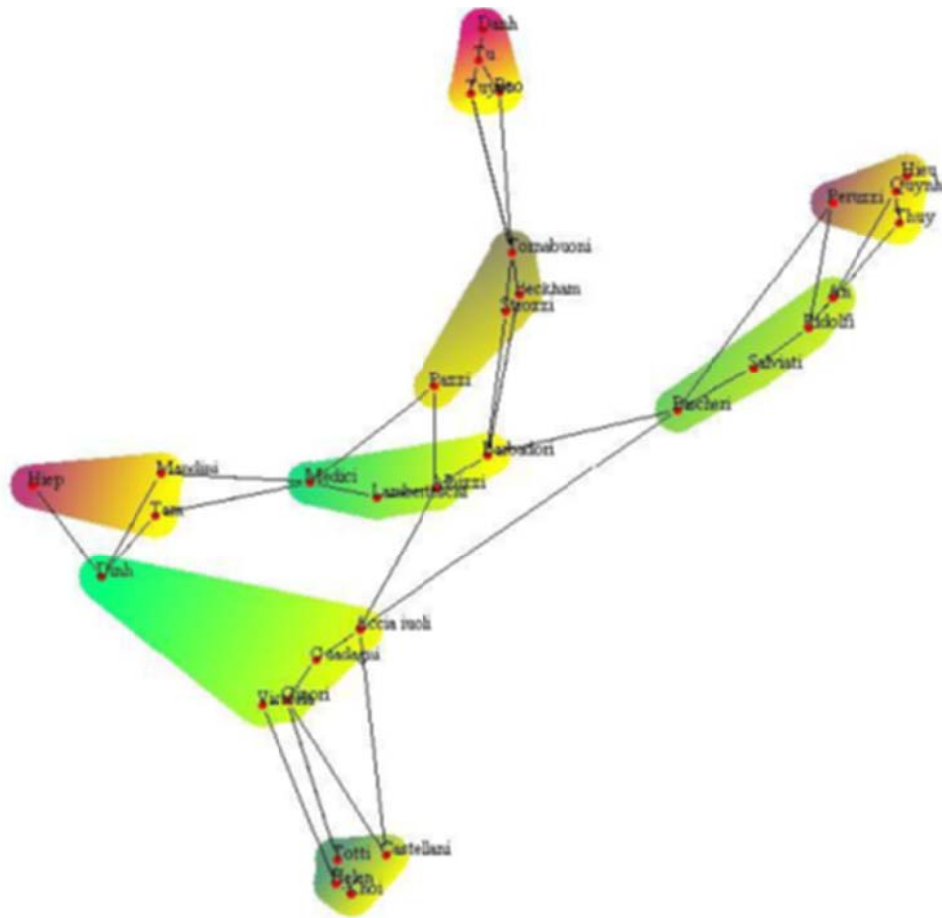
A. Ví dụ 1 - Mạng cộng tác

Ví dụ đầu tiên chúng tôi sử dụng đó là một mạng cộng tác bao gồm 32 đỉnh và được phân làm 8 nhóm trong đó mỗi đỉnh biểu diễn cho một người cụ thể (nghiên cứu viên). Ở ví dụ này việc bằng cách nào phân nhóm các đỉnh vào 8 nhóm như trên không phải là vấn đề chính. Trước tiên, chúng tôi áp dụng mô hình đề xuất bởi [5], kết quả thu được như hình 4. Chúng ta có thể dễ dàng nhận thấy kết quả khi áp dụng mô hình [5] là tốt tuy nhiên lại không thể hiện được cấu trúc phân nhóm của đồ thị.

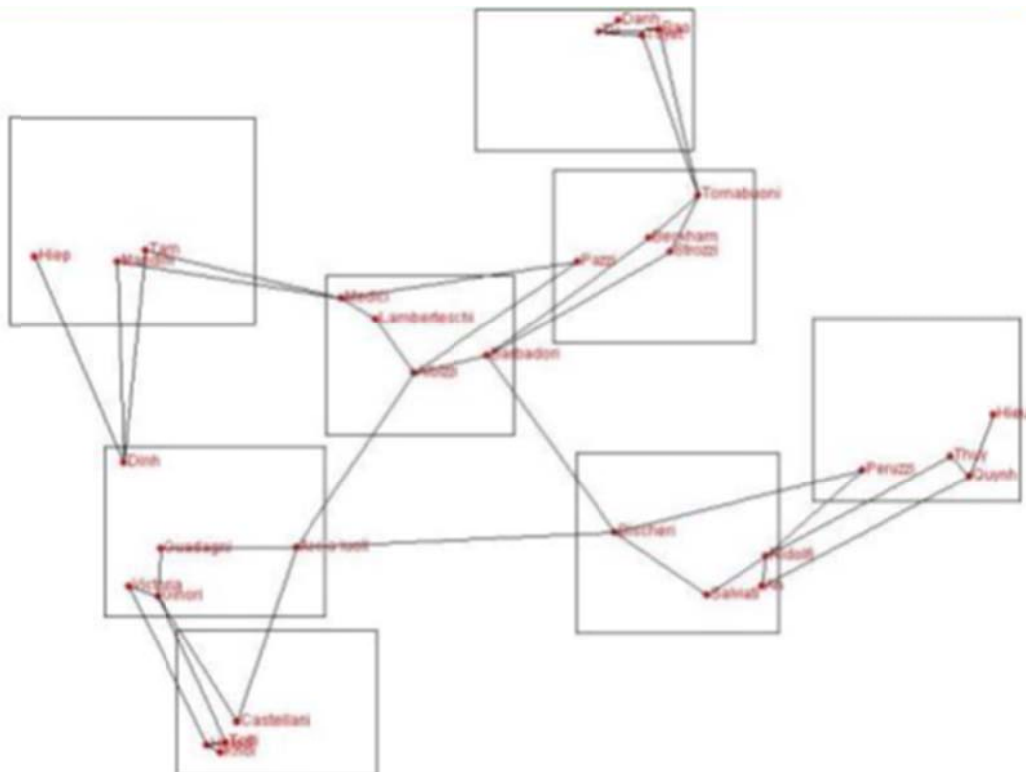


Hình 4. Hình vẽ sử dụng mô hình đề xuất bởi [5]

Chúng tôi áp dụng mô hình đề xuất cho mạng cộng tác trên trong cả 2 trường hợp: vùng hiển thị không được định nghĩa (hình 5) và vùng hiển thị được định nghĩa trước (hình 6). Trong cả 2 trường hợp có thể nhận thấy rằng cấu trúc phân nhóm của mạng cũng như cấu trúc nội tại của mỗi nhóm đều được thể hiện rõ.



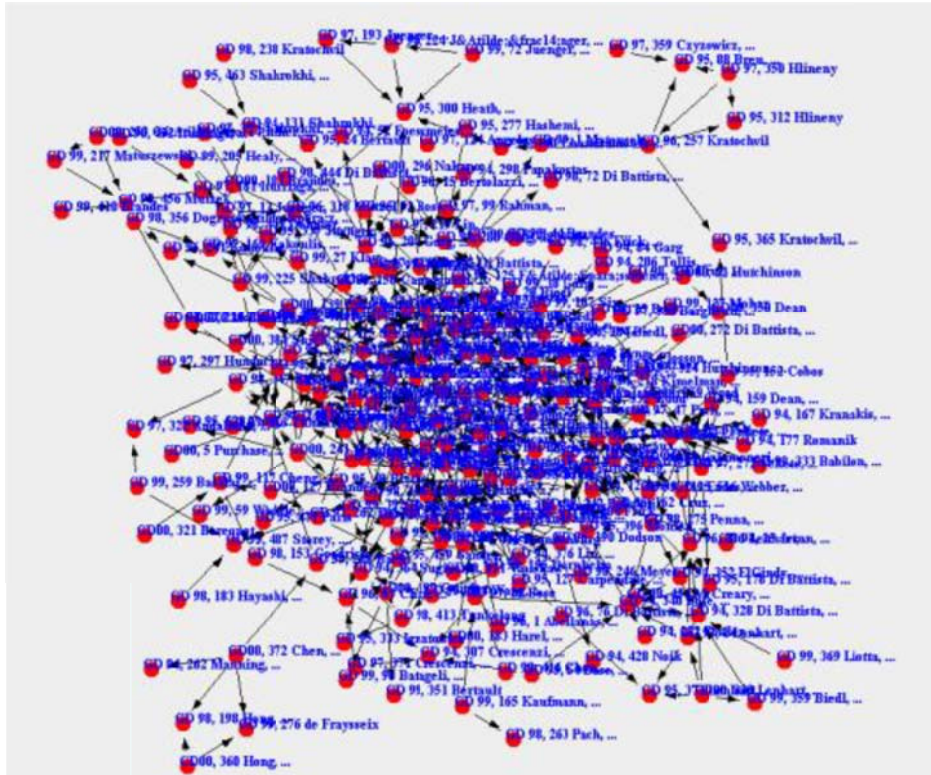
Hình 5. Hình vẽ sử dụng mô hình đề xuất – Trường hợp vùng hiển thị mỗi nhóm không được định nghĩa



Hình 6. Hình vẽ sử dụng mô hình đề xuất – Trường hợp vùng hiển thị mỗi nhóm được định nghĩa

B. Ví dụ 2 – Mạng trích dẫn

Trong ví dụ này, chúng tôi sử dụng mạng trích dẫn trong lĩnh vực Graph Drawing [10]. Mạng là một đồ thị gồm có 311 đỉnh và 647 cung. Nếu như áp dụng mô hình [5] cho mạng trích dẫn này thì chúng ta có thể dễ dàng nhận xét rằng không thể rút trích tri thức gì từ hình vẽ của mạng (Hình 7).



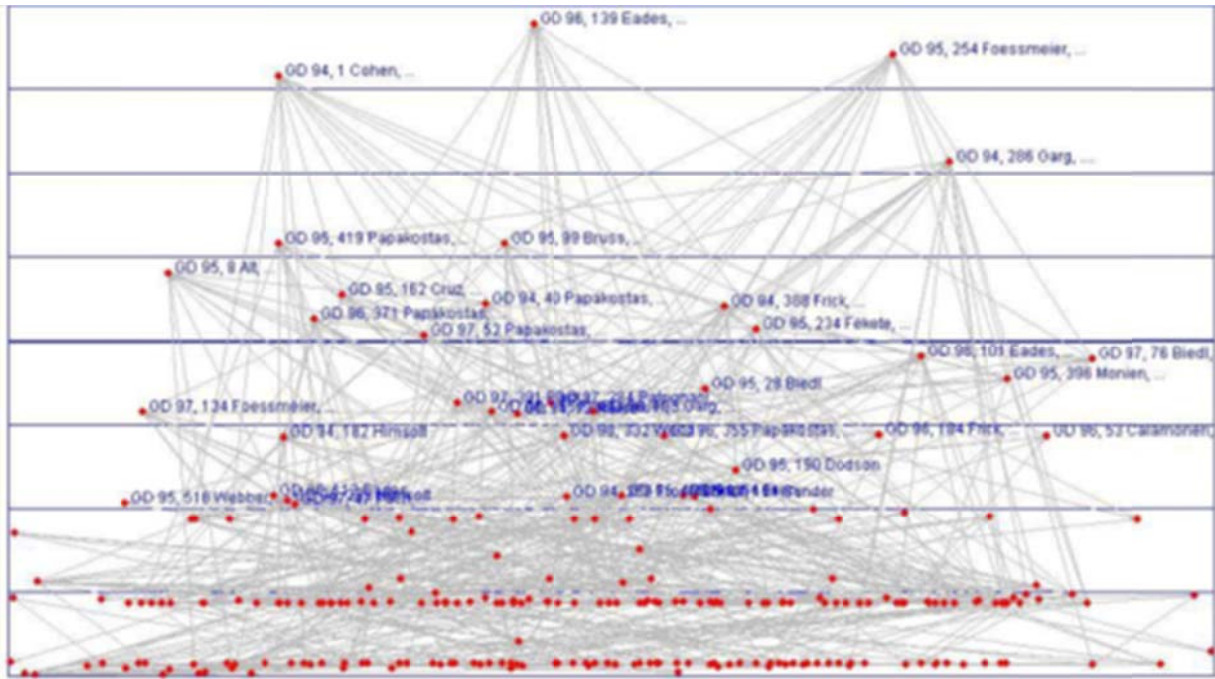
Hình 7. Hình vẽ mạng trích dẫn các bài báo khoa học thuộc lĩnh vực Graph Drawing trong giai đoạn 1994 - 2000

Tuy nhiên trong một mạng trích dẫn thì giá trị authority của mỗi nút là một giá trị quan trọng mà các nhà phân tích mạng xã hội quan tâm. Chúng tôi áp dụng phương pháp phân tích Hub – Authority dựa trên việc so sánh các đỉnh của hai đồ thị mà chúng tôi đã đề xuất trong [20]. Các bài báo có giá trị authority cao nhất được trình bày trong bảng 1.

Bảng 1. Top 10 bài báo theo giá trị Authority trong lĩnh vực Graph Drawing từ 1994-2000

Authority	
GD 96, 139 Eades, ...	Two Algorithms for Three Dimensional Orthogonal Graph Drawing.
GD 94, 1 Cohen, ...	Three-Dimensional Graph Drawing
GD 95, 254 Foessmeier, ...	Drawing High Degree Graphs with Low Bend Numbers
GD 94, 286 Garg,	On the Computational Complexity of Upward and Rectilinear Planarity Testing
GD 95, 419 Papakostas, ...	Issues in Interactive Orthogonal Graph Drawing
GD 95, 99 Bruss, ...	Fast Interactive 3-D Graph Visualization
GD 94, 388 Frick, ...	A Fast Adaptive Layout Algorithm for Undirected Graphs
GD 95, 8 Alt, ...	Universal 3-Dimensional Visibility Representations for Graphs
GD 97, 52 Papakostas, ...	Incremental Orthogonal Graph Drawing in Three Dimensions
GD 95, 234 Fekete, ...	New Results on a Visibility Representation of Graphs in 3D

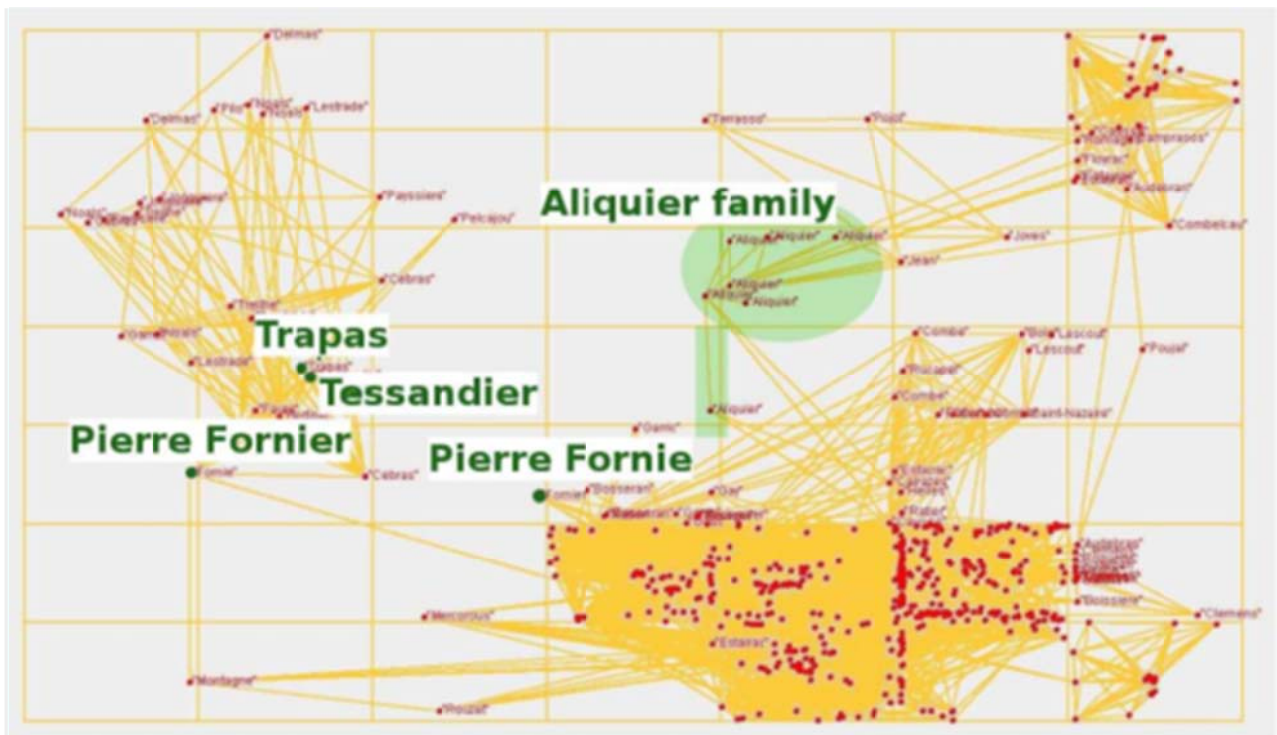
Chúng tôi áp dụng mô hình đề xuất trong ngữ cảnh vùng hiển thị của mỗi nhóm được định nghĩa sẵn. Chúng tôi phân các bài báo vào 8 nhóm khác nhau dựa trên giá trị authority của bài báo đó. Các nhóm sẽ được hiển thị trong các vùng hình chữ nhật được định sẵn từ trên xuống dưới theo giá trị authority. Nhóm trên cùng là nhóm các bài báo có giá trị authority cao nhất (hình 8). Cách biểu diễn này giúp thấy rõ một điều là các bài báo có giá trị authority cao nhận được rất nhiều trích dẫn từ các bài báo có giá trị authority thấp. Một điểm thú vị nữa là các bài báo có giá trị authority thuộc cùng một nhóm hầu như không trích dẫn lẫn nhau.



Hình 8. Hình vẽ mạng trích dẫn trong đó các đỉnh được nhóm theo giá trị authority và được hiển thị trong tầng tương ứng

C. Ví dụ 3 – Mạng hợp đồng mua bán ruộng đất

Dữ liệu trong mạng là các hợp đồng giao dịch ruộng đất trong đó có đề cập đến thông tin của người nông dân, tên lãnh chúa của vùng cũng như người làm chứng cùng các thông tin khác như ngày tháng giao dịch, vị trí mảnh ruộng ... Các nông dân được biểu diễn bởi các nút (615) của mạng, các cung (4195) có trọng số thể hiện số lần giao dịch mua bán giữa các nông dân. Chúng tôi ánh xạ mạng vừa nêu vào cấu trúc mạng Kohonen [21] và sử dụng mô hình đề xuất để biểu diễn kết quả (hình 9). Kết quả thể hiện được nhiều điều lý thú. Trước tiên, người kết nối phân trên bên trái của mạng với phần dưới bên phải là “Pierre Fornie”, một người được giới nghiên cứu lịch sử Pháp biết tên. Tiếp theo, các nhân tố quan trọng ở nửa trên bên trái mạng là “Trapas” và “Tessandier” lại cùng một gia đình với “Pierre Fornie”. Cuối cùng, phần trên bên phải mạng liên kết với phần dưới bên phải mạng chỉ bởi các thành viên trong gia đình Aliquier.



Hình 9. Mạng giao dịch ruộng đất [10] ánh xạ lên bản đồ Kohonen sử dụng mô hình vẽ đồ thị phân nhóm

V. KẾT LUẬN

Trong bài báo này, chúng tôi đã giới thiệu một mô hình lực mới dựa trên mô hình đề xuất bởi [5] để có thể vẽ được các đồ thị phân nhóm theo các yêu cầu như đã đề cập ở phần II. Hai hướng tiếp cận khi vẽ đồ thị phân nhóm đã được đề xuất bao gồm có hoặc không có điều kiện ràng buộc về vùng hiển thị của các nhóm.

Mô hình đề xuất đã được thử nghiệm trên nhiều đồ thị phân nhóm khác nhau với kích thước từ vài chục đỉnh đến vài trăm đỉnh. Ba ví dụ tiêu biểu cho các lĩnh vực áp dụng khác nhau của mô hình đề xuất đã được giới thiệu và thảo luận. Các ví dụ này đã phần nào khẳng định tính đúng đắn của mô hình. Mô hình này thiết nghĩ có thể mở rộng để tích hợp yếu tố địa lý khi thực hiện phân tích các mạng cộng tác, ví dụ như các chủ thể cùng một vùng miền, một quốc gia sẽ được phân thành một nhóm và được biểu diễn trong phạm vi đường biên địa lý tương ứng trên bản đồ.

VI. TÀI LIỆU THAM KHẢO

- [1] Oliveira Márcia, Gama João. “An overview of social network analysis”, WIREs Data Mining Knowl Discov, 2: 99-115. doi: 10.1002/widm.1048, 2012.
- [2] E. R. Gansner, and S. C. North, “An open graph visualization system and its applications to software engineering,” Software Practice and Experience, 30(11), pp. 1203-1233, 2000.
- [3] Eschbach, T.; Gunther, W.; Becker, B., “Orthogonal circuit visualization improved by merging the placement and routing phases”, VLSI Design, 2005. 18th International Conference on , vol., no., pp.433,438, 3-7 Jan. 2005 doi: 10.1109/ICVD.2005.134
- [4] Daniel A. Keim, databases and visualization, Proceedings of the 1996 ACM SIGMOD international conference on Management of data, p.543, June 04-06, 1996, Montreal, Quebec, Canada [doi>10.1145/233269.280349].
- [5] T. M. J. Fruchterman, E. M. Reingold, “Graph drawing by force-directed placement,” Software. Pract. Exper. 21(11), pp. 1129–1164, 1991
- [6] T. Kamada, S. Kawai, “An algorithm for drawing general undirected graphs,” Information Processing Letters, 31(1), pp. 7-15, 1989
- [7] Gajer P., Kobourov S. G., “GRIP: Graph dRawing with Intelligent Placement”, In Marks, Joe, Eds. Proceedings Graph Drawing”, pages pp. 222-228, Colonial Williamsburg, 2001.
- [8] S. Hachul and M. Jünger, “An Experimental Comparison of Fast Algorithms for Drawing General Large Graphs,” Pro. of the Graph Drawing 2005, vol. 3843, Lecture Notes in Computer Science, Springer-Verlag; pp. 235-250, Limerick, Ireland, 2006.
- [9] Q. D. Truong, T. Dkaki, P-J. Charrel., “Energy model for clustered graph drawing”, Proceeding of Colloque Veille Stratégique Scientifique et Technologique (VSST 2007), Marrakech (Maroc), 2007.
- [10] T. Biedl and F. J. Brandenburg, “Graph-Drawing contest report,” Proc. of the Graph drawing 2001, Lecture Notes in Computer Science, Springer-Verlag, pp. 388-403, 2001.
- [11] R. Boulet, F. Hautefeuille, B. Jouve, P. Kuntz, B. Le Goffic, F. Picarougne, and N. Villa, “Sur l’analyse de réseaux de sociabilité dans la société paysanne médiévale”, In MASHS 2007, Brest, France, 2007.
- [12] C. Görg, P. Birke, M. Pohl, and S. Diehl, “Dynamic graph drawing of sequences of orthogonal and hierarchical graphs,” In Graph Drawing, pp. 228-238, 2005.
- [13] U. Dogrusöz, B. Madden and P. Madden, “Circular Layout in the Graph Layout Toolkit”, Proc. GD '96, LNCS 1190, Springer-Verlag, pp. 92-100, 1997.
- [14] G. Melancon and I. Herman, “Circular drawings of rooted trees”, Technical Report: INS-R9817, Centre for Mathematics and Computer Science.
- [15] P. Eades, “A heuristic for graph drawing”, Congressus Numerantium 42 (11): 149–160, 1984.
- [16] Gajer P., Kobourov S. G., “GRIP: Graph dRawing with Intelligent Placement”, In Marks, Joe, Eds. Proceedings Graph Drawing, pages pp. 222-228, Colonial Williamsburg, 2001.
- [17] Ho, J., Hong, S.H., “Drawing clustered graphs in three dimensions”, Proceedings of 13th International Symposium on Graph Drawing (GD2005) (2005).
- [18] Chuang J-H. Lin C-C. Yen H-C., “Drawing Graphs with Nonuniform Nodes Using Potential Fields”, In Liotta, Giuseppe, Eds. Proceedings Graph Drawing, pages pp. 460-465, Perugia, 2004.
- [19] Balzer M., Deussen O., “Level-of-detail visualization of clustered graph layouts”, 6th International Asia-Pacific Symposium on Visualization, APVIS07, pp. 133-140, Sydney Australia, 2007.

- [20] Q. D. Truong, T. Dkaki, J. Mothe et al, "GVC: a graph-based Information Retrieval Model", Pro. of the Conférence francophone en Recherche d'Information et Applications (CORIA 2008), CNRS, pp. 337-351, Trégastel, France, 2008.
- [21] K.W. Lau, H. Yin, and S. Hubbard. Kernel self-organising maps for classification. *Neurocomputing*, 69:2033–2040, 2006.

A NEW ENERGY-BASED MODEL FOR CLUSTERED GRAPH DRAWING

Truong Quoc Dinh, Taoufiq Dkaki

ABSTRACT - Graph structure can model many kinds of data in various domains. Moreover, graphs are especially used for information visualization. When dealing with a large amount of information, the way to identify groups and subgroups among data becomes important. In this paper, we present a new graph drawing approach that helps to better identify the cluster structure in data and also the interactions that may exist between clusters. In the context of this work, we focus on the visualization aspects and assume that the clusters are already created. We propose an energy-based model for clustered graph drawing that produces a drawing that ensures each cluster will occupy a separate zone within the layout. The drawing areas assigned to the clusters can be user-specified or automatically crafted. In both two cases, the approach we suggest successfully points out the cluster structure of the graph.