

ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

---



ĐỖ XUÂN QUYỀN

**PHƯƠNG PHÁP TỐI ƯU ĐÀN KIẾN  
DÓNG HÀNG HAI ĐỒ THỊ**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**THÁI NGUYÊN - 2016**

ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

---



**ĐỖ XUÂN QUYỀN**

**PHƯƠNG PHÁP TỐI ƯU ĐÀN KIẾN  
DÓNG HÀNG HAI ĐỒ THỊ**

**Chuyên ngành: Khoa học máy tính  
Mã số: 60.48.01.01**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**Người hướng dẫn khoa học: TS. Đỗ Đức Đông**

**THÁI NGUYÊN - 2016**

## LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn của TS. Đỗ Đức Đông. Các kết quả được viết chung với các tác giả khác đều được sự đồng ý của đồng tác giả trước khi đưa vào luận văn. Các kết quả thực nghiệm nêu trong luận văn đều là chính xác và chưa từng được công bố.

Tôi hoàn toàn chịu trách nhiệm về tính pháp lý của luận văn này.

*Thái Nguyên, tháng 7 năm 2016*

Học viên thực hiện

*Đỗ Xuân Quyền*

## LỜI CẢM ƠN

Lời đầu tiên, em xin bày tỏ sự biết ơn sâu sắc đến TS.Đỗ Đức Đông người đã tận tình hướng dẫn, chỉ bảo, giúp đỡ em trong suốt quá trình làm luận văn.

Em cũng xin gửi lời cảm ơn đến các thầy cô giáo trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên, các thầy cô Viện Công nghệ thông tin đã truyền đạt những kiến thức và giúp đỡ em trong suốt quá trình học của mình.

Tôi cũng xin gửi lời cảm ơn tới Sở Giáo dục và Đào tạo Hải Phòng, Ban giám hiệu trường THPT Quang Trung Hải Phòng đã tạo điều kiện thuận lợi cho tôi tham gia khóa học và trong suốt quá trình hoàn thành luận văn.

Cuối cùng, tôi xin gửi lời cảm ơn tới các đồng nghiệp, gia đình và bạn bè những người đã ủng hộ, động viên tạo mọi điều kiện giúp đỡ để tôi có được kết quả như ngày hôm nay.

*Thái Nguyên, tháng 7 năm 2016*

Học viên

*Đỗ Xuân Quyền*

## MỤC LỤC

|   |           |
|---|-----------|
| LỜI CAM ĐOAN .....  | i         |
| LỜI CẢM ƠN .....  | ii        |
| MỤC LỤC .....   | iii       |
| DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT .....                    | v         |
| DANH MỤC CÁC HÌNH .....   | vi        |
| DANH MỤC CÁC BẢNG .....   | vii       |
| <b>MỞ ĐẦU .....</b>   | <b>1</b>  |
| <b>CHƯƠNG I: DÓNG HÀNG HAI ĐỒ THỊ VÀ CÁC PHƯƠNG PHÁP</b>        |           |
| <b>TIẾP CẬN HIỆN NAY .....</b>                                  | <b>3</b>  |
| 1.1. Bài toán đóng hàng hai đồ thị .....                        | 3         |
| 1.2. Một số phương pháp tiếp cận hiện nay .....                 | 4         |
| 1.2.1. SPINAL .....   | 5         |
| 1.2.2. FastNA .....   | 7         |
| 1.3. Kết luận chương .....                                      | 10        |
| <b>CHƯƠNG 2. PHƯƠNG PHÁP TỐI ƯU ĐÀN KIẾN .....</b>              | <b>11</b> |
| 2.1. Từ kiến tự nhiên đến kiến nhân tạo .....                   | 13        |
| 2.1.1. Kiến tự nhiên .....                                      | 13        |
| 2.1.2. Kiến nhân tạo .....                                      | 16        |
| 2.2. Phương pháp ACO cho bài toán tối ưu tổ hợp tổng quát ..... | 17        |
| 2.2.1. Đồ thị cấu trúc .....                                    | 17        |
| 2.2.2. Mô tả thuật toán ACO tổng quát .....                     | 19        |
| 2.3. Một số vấn đề liên quan .....                              | 22        |
| 2.3.1. Đặc tính hội tụ .....                                    | 22        |
| 2.3.2. Thực hiện song song .....                                | 22        |
| 2.3.3. ACO kết hợp với tìm kiếm cục bộ .....                    | 23        |

|  |           |
|--|-----------|
| 2.3.4. Thông tin heuristic .....                                       | 23        |
| 2.3.5. Số lượng kiến .....   | 24        |
| 2.3.6. Tham số bay hơi.....  | 24        |
| 2.4. Tính biến thiên của vết mùi và các thuật toán cập nhật mùi.....   | 24        |
| 2.4.1. Thuật toán tổng quát.....                                       | 25        |
| 2.4.1.1. Quy tắc chuyển trạng thái .....                               | 25        |
| 2.4.1.2. Cập nhật mùi .....  | 26        |
| 2.4.2. Đánh giá .....  | 27        |
| 2.4.2.1. Tính khai thác và khám phá .....                              | 27        |
| 2.4.2.2. Các thuật toán cập nhật mùi theo quy tắc ACS.....             | 28        |
| 2.4.2.3. Các thuật toán cập nhật mùi theo quy tắc MMAS .....           | 29        |
| 2.4.2.4. Ưu điểm khi sử dụng SMMAS và 3-LAS.....                       | 30        |
| 2.4.3. Tính bất biến.....  | 31        |
| <b>CHƯƠNG III: PHƯƠNG PHÁP TỐI ƯU ĐÀN KIẾN GIẢI BÀI TOÁN</b>           |           |
| <b>DÓNG HÀNG HAI ĐỒ THỊ.....</b>                                       | <b>34</b> |
| 3.1. Thuật toán tối ưu đàn kiến giải bài toán đóng hàng hai đồ thị ..  | 34        |
| 3.1.1. Xây dựng đồ thị cấu trúc thích hợp .....                        | 36        |
| 3.1.2. Chọn thông tin heuristic;.....                                  | 36        |
| 3.1.3. Cập nhật mùi .....  | 37        |
| 3.2. Thực nghiệm, so sánh kết quả với phương pháp SPINAL và FastNA ... | 42        |
| 3.2.1. Thực nghiệm.....  | 42        |
| 3.2.2. So sánh .....   | 46        |
| <b>KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....</b>                               | <b>49</b> |
| <b>TÀI LIỆU THAM KHẢO .....</b>  | <b>50</b> |

## DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT

|              |  |
|--------------|--|
| $\tau_{max}$ | <i>Cận trên của vết mùi</i>  |
| $\tau_{min}$ | <i>Cận dưới của vết mùi</i>  |
| $\tau_{mid}$ | <i>Cận giữa của vết mùi</i>  |
| $\tau_0$     | <i>Vết mùi được khởi tạo ban đầu</i>                               |
| $\tau_{ij}$  | <i>Vết mùi trên cạnh (i, j)</i>                                    |
| $N_c$        | <i>Số vòng lặp trong thuật toán ACO</i>                            |
| $N_a$        | <i>Số kiến sử dụng trong thuật toán ACO</i>                        |
| $\rho$       | <i>Tham số bay hơi</i>   |
| 3-LAS        | <i>Three-Level Ant System (Hệ kiến ba mức)</i>                     |
| ACO          | <i>Ant Colony Optimization (Tối ưu đàn kiến)</i>                   |
| ACS          | <i>Ant Colony System (Hệ đàn kiến)</i>                             |
| AS           | <i>Ant System (Hệ kiến)</i>  |
| CRM          | <i>Cis-Regulatory Module (Mô-đun điều tiết)</i>                    |
| EC           | <i>Evolutionary Computing (Tính toán tiến hoá)</i>                 |
| GA           | <i>Genetic Algorithm (Thuật toán di truyền)</i>                    |
| G-best       | <i>Global-best (Lời giải tốt nhất tính đến thời điểm hiện tại)</i> |
| I-best       | <i>Iteration-best (Lời giải tốt nhất trong bước lặp hiện tại)</i>  |
| MLAS         | <i>Multi-level Ant System (Hệ kiến đa mức)</i>                     |
| MMAS         | <i>Max-Min Ant System (Hệ kiến Max Min)</i>                        |
| SA           | <i>Simulated Annealing (Thuật toán mô phỏng luyện kim)</i>         |
| SMMAS        | <i>Smoothed Max-Min Ant System (Hệ kiến Max Min trơn)</i>          |
| $n_{keep}$   | <i>Số nút cần giữ lại</i>  |
| $SeedV_{12}$ | <i>Tập nút đã đóng hàng thuộc đồ thị <math>G_1</math></i>          |

**DANH MỤC CÁC HÌNH**

|  |    |
|--|----|
| Hình 1.1: Thuật toán SPINAL .....                        | 6  |
| Hình 1.2: Thuật toán FastNA .....                        | 7  |
| Hình 1.3: Thủ tục Rebuild - FastNA .....                 | 9  |
| Hình 2.1: Thực nghiệm cây cầu đôi .....                  | 14 |
| Hình 2.2: Thí nghiệm bổ sung .....                       | 16 |
| Hình 2.3: Thuật toán ACO .....                           | 20 |
| Hình 3.1: Thuật toán ACO tạo dóng hàng ban đầu .....     | 39 |
| Hình 3.2: Thuật toán Rebuild xây dựng lại lời giải ..... | 41 |



**DANH MỤC CÁC BẢNG**

|  |    |
|--|----|
| Bảng 3.1: Thông tin về dữ liệu.....  | 42 |
| Bảng 3.2: Kết quả thực nghiệm của ACOPPI theo tiêu chí GNAS .....                                | 44 |
| Bảng 3.3: Kết quả thực nghiệm của ACOPPI theo tiêu chí EC .....                                  | 45 |
| Bảng 3.4: So sánh kết quả thực nghiệm của ACOPPI với SPINAL và FastNA<br>theo tiêu chí GNAS..... | 47 |
| Bảng 3.5: So sánh kết quả thực nghiệm của ACOPPI với SPINAL và FastNA<br>theo tiêu chí EC .....  | 48 |

## MỞ ĐẦU

Tin sinh học là một lĩnh vực khoa học mới rất được quan tâm trong hơn thập kỷ gần đây, với sự phát triển của khoa học công nghệ, mối liên kết giữa sinh học và tin học trở lên khăng khít hơn, Tin học hỗ trợ cho việc giải quyết các bài toán với dữ liệu lớn của Sinh học, đồng thời đó cũng là một hướng phát triển mới của ngành Tin học về giải thuật và ứng dụng.

Dóng hàng hai đồ thị là một bài toán quan trọng trong lý thuyết đồ thị, nó giúp chúng ta xác định tính tương đồng của hai đồ thị. Về mặt sinh học nó giúp xác định tính tương đồng giữa các mạng tương tác protein. Hiện nay có nhiều tiêu chí về cách đánh giá cho đóng hàng. Một cách đánh giá thường được sử dụng hiện nay là đánh giá dựa trên lực lượng của tập cạnh (sự tương đồng về cấu trúc) và sự tương đồng giữa các nút.

Dóng hàng hai đồ thị được Aladag và Erten chứng minh là bài toán thuộc lớp NP-khó [4] và có nhiều ứng dụng. Đặc biệt, trong những năm gần đây, với sự phát triển của các kỹ thuật sinh học công nghệ cao đã cho phép các nhà nghiên cứu xây dựng được các mạng tương tác protein (Protein-Protein Interaction Network – PPI Network) tương đối đầy đủ cho nhiều loài sinh vật. Bài toán đóng hàng mạng PPI là một bài toán quan trọng trong phân tích mạng PPI nói chung. ***Các mạng tương tác protein được mô tả bằng đồ thị, bài toán đóng hàng mạng được chuyển tải về bài toán đóng hàng đồ thị.***

Phương pháp tối ưu đàn kiến (Ant Colony Optimization - ACO) là cách tiếp cận metaheuristic, được giới thiệu bởi Dorigo năm 1991 đang được nghiên cứu và ứng dụng rộng rãi cho các bài toán tối ưu tổ hợp khó.

Chính vì vậy tác giả chọn đề tài khoa học ***phương pháp tối ưu đàn kiến đóng hàng hai đồ thị***. Thực nghiệm, tác giả sẽ sử dụng bộ dữ liệu vào