

Trong phần này, bạn sẽ lần lượt giải quyết các **Problem**. Bạn cần trình bày nó và submit lên hệ thống Moodle dưới dạng .pdf với tên MSSV.pdf, ví dụ: 12345678.pdf

## 1 Động lực

Ngôn ngữ Lý thuyết Đồ thị cho phép chúng ta thực hiện nhiều điều, từ hình thức hóa bài toán/ vấn đề cho đến cấu trúc hóa dữ liệu thành dạng đồ thị để mà có thể áp dụng thuật toán/ mô hình đồ thị để đưa ra kết quả/ dự đoán cho vấn đề đang được xem xét. Mạng xã hội (social networks) là một ví dụ điển hình của một loại dữ liệu đồ thị phổ biến mà biểu diễn tương tác giữa người với người trong một nhóm hay một cộng đồng hay một xã hội trong môi trường cụ thể (Internet).

Một trong những bài toán trong lĩnh vực khoa học về mạng lưới là *"làm thế nào để hiểu về mối quan hệ giữa hai đỉnh trong mạng?"*, bài toán tưởng chừng đơn giản nhưng lại rất phức tạp. Mạng lưới biến đổi theo thời gian, và thời gian sẽ cho chúng ta biết nhiều thứ.

Nhà Vật lý học người Đan Mạch, Niels Henrik David Bohr, có đóng góp rất lớn trong lý thuyết nền tảng của cơ học lượng tử từng nói rằng:

"Prediction is very difficult, especially if it's about the future."

Tạm dịch: Bài toán dự đoán rất khó, nhất là về tương lai.

*Nils Bohr, nhà Vật lý học, giải Nobel Vật lý học*

Đúng vậy, thật khó khi dự đoán về tương lai. Điều này, dẫn đến một số câu hỏi:

- Các mẫu liên kết (association patterns) thay đổi như thế nào theo thời gian?
- Sự liên kết giữa hai nút bị ảnh hưởng bởi các nút khác như thế nào? Bài toán chúng ta muốn giải quyết ở đây là dự đoán triển vọng của một liên kết tương lai giữa hai nút, biết rằng không có mối liên hệ nào giữa các nút đó trong trạng thái hiện tại của đồ thị. Bài toán này được gọi là bài toán dự đoán liên kết (Link prediction problem).

Trong thực tế, bài toán dự đoán liên kết yêu cầu: sự phát triển của mạng xã hội có thể được mô hình hóa ở mức độ nào bằng cách sử dụng các đặc trưng nội tại của cấu trúc tô-pô mạng? Trạng thái hiện tại của mạng có thể được sử dụng để dự đoán các liên kết trong tương lai không?

Bài toán dự đoán liên kết cũng được liên hệ với bài toán **suy luận các liên kết bị thiếu (problem of inferring missing links)** từ một mạng quan sát được: trong một số miền tri thức, người ta xây dựng một mạng lưới của sự tương tác dựa trên dữ liệu quan sát được, và sau đó cố gắng suy diễn những liên kết bổ sung mà không xuất hiện trực tiếp và có khả năng tồn tại.

**Problem 1.** Trình bày điểm giống và khác nhau giữa bài toán dự đoán liên kết và bài toán suy luận liên kết bị thiếu?

**Problem 2.** Cho ví dụ về bài toán dự đoán liên kết trong một miền tri thức cụ thể?

## 2 Phát biểu bài toán

Cho trước một đồ thị vô hướng, không trọng số (unweighted, undirected graph),  $G = (V, E)$  thể hiện cấu trúc tô-pô của một mạng xã hội mà trong đó mỗi cạnh  $e = \langle u, v \rangle \in E$  thể hiện một tương tác giữa  $u$  và  $v$  xuất hiện ở một thời gian cụ thể  $t(e)$ .

Với hai thời điểm,  $t$  và  $t' > t$ , gọi  $G[t, t']$  đại diện cho đồ thị con của  $G$  bao gồm tất cả các cạnh với một khoảng thời gian giữa  $t$  và  $t'$ . Gọi  $t_0, t'_0$ , và  $t'_1$  là bốn thời điểm, trong đó  $t_0 < t'_0 \leq t_1 < t'_1$ . Ta có bài toán dự đoán liên kết:

- Đầu vào (input):  $G[t_0, t'_0]$
- Đầu ra (output): một danh sách của các cạnh không thể hiện trong  $G[t_0, t'_0]$  mà được dự đoán là xuất hiện trong mạng  $G[t_1, t'_1]$

Ta có thể xem:  $[t_0, t'_0]$  = training interval và  $[t_1, t'_1]$  = test interval.

**Problem 3.** *Làm thế nào có thể đánh giá hiệu suất của phương pháp đề xuất cho việc giải quyết bài toán này? Nói cách khác, bạn trình những metric có thể được sử dụng cho bài toán này.*

Chúng ta sẽ có một số phương pháp tiếp cận cho bài toán này như sau:

## 3 Tiếp cận đỉnh dựa trên tương đồng tô-pô đồ thị (thiên hướng cục bộ)

Trong tiếp cận này, chúng ta sẽ cho một số độ đo:

- Common Neighbors
- Jaccard's Coefficient
- Adamic/Adar
- Preferential Attachment

**Problem 4.** *Bạn chọn một trong bốn độ đo vừa liệt kê và trình bày về nó. Cho ví dụ minh họa.*

## 4 Tiếp cận đường đi dựa trên tương đồng tô-pô đồ thị (thiên hướng toàn bộ)

Trong tiếp cận này, chúng ta sẽ cho một số độ đo:

- Hit
- Katz

**Problem 5.** *Bạn chọn một trong hai độ đo vừa liệt kê và trình bày về nó. Cho ví dụ minh họa.*

## 5 Bổ sung: Nghiên cứu về nhân quả

**Problem 6.** Trình bày và so sánh các cấp độ: *Associational* (kết hợp), *Interventional* (can thiệp), *Counterfactual* (phản thực tế) trong suy luận nhân quả.