

TRƯỜNG ĐẠI HỌC VĂN LANG
KHOA TÀI CHÍNH – NGÂN HÀNG

ĐỀ THI VÀ ĐÁP ÁN
THI KẾT THÚC HỌC PHẦN
Học kỳ 1, năm học 2024-2025

I. Thông tin chung

Tên học phần:	Phân tích dữ liệu mạng xã hội		
Mã học phần:	71FINC40233	Số tín chỉ:	3
Mã nhóm lớp học phần:	241_71FINC40233_01, 02		
Hình thức thi: Trắc nghiệm kết hợp Tự luận	Thời gian làm bài:	75	phút
<i>Thí sinh được tham khảo tài liệu:</i>	☒ Có	☐ Không	

Cách thức nộp bài phần tự luận (Giảng viên ghi rõ yêu cầu):

- SV gõ trực tiếp trên khung trả lời của hệ thống thi.

II. Các yêu cầu của đề thi nhằm đáp ứng CLO

(Phần này phải phối hợp với thông tin từ đề cương chi tiết của học phần)

Ký hiệu CLO	Nội dung CLO	Hình thức đánh giá	Trọng số CLO trong thành phần đánh giá (%)	Câu hỏi thi số	Điểm số tối đa	Lấy dữ liệu đo lường mức đạt PLO/PI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
CLO 1	Hiểu và áp dụng các khái niệm cơ bản về lý thuyết đồ thị và mạng xã hội để phân tích cấu trúc của các mạng xã hội thực tế.	Trắc nghiệm	20%	1-4	2	PI2.1
CLO 2	Phân tích, khai phá dữ liệu từ các nội dung thông tin thu thập được trên mạng xã hội.	Trắc nghiệm	40%	5-12	4	PI3.3
CLO 3	Làm việc nhóm hiệu quả; tổ chức công việc tốt;	Tự luận	40%	1, 2	4	PI5.5

	thuyết trình và nói trước đám đông một cách tự tin; vận dụng tư duy phản biện trong phân tích, xử lý, đánh giá thông tin trên mạng xã hội.					
--	--	--	--	--	--	--

III. Nội dung câu hỏi thi

PHẦN TRẮC NGHIỆM (12 câu hỏi, mỗi câu 0.5)

1. Trong phân tích dữ liệu mạng xã hội, số đo trung tâm trung gian (betweenness centrality) của một nút được sử dụng để:

- A. Đo lường khả năng của nút trong việc làm cầu nối giữa các nhóm khác nhau trong mạng
- B. Xác định số lượng kết nối trực tiếp của nút đó
- C. Xác định khoảng cách trung bình từ một nút đến tất cả các nút khác trong mạng
- D. Xác định mức độ tương tác giữa các nút trong cùng một cụm

ANSWER: A

2. Phương pháp Adamic/Adar Index trong dự đoán liên kết sử dụng yếu tố nào để đánh giá khả năng liên kết giữa hai nút?

- A. Mức độ quan trọng của các nút láng giềng chung, với trọng số cao hơn cho nút có ít kết nối hơn.
- B. Độ phổ biến của các nút trong mạng, thể hiện qua tích kích thước của các láng giềng chung.
- C. Xác suất hai nút cùng tham gia vào nhiều cụm khác nhau
- D. Tổng số liên kết trực tiếp của hai nút, với trọng số cao hơn cho nút có nhiều kết nối hơn.

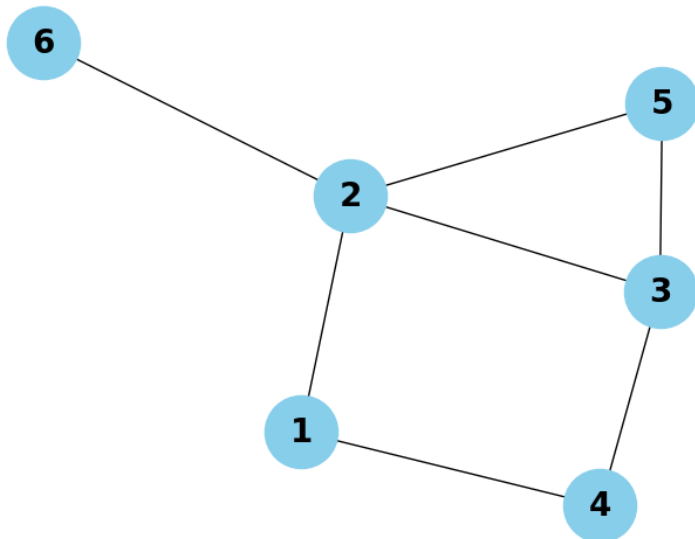
ANSWER: A

3. Trong quá trình phân tích dữ liệu mạng xã hội, khi áp dụng thuật toán phát hiện cộng đồng, mục tiêu chính là:

- A. Phân nhóm các nút sao cho các nút trong cùng một nhóm có sự liên kết chặt chẽ, trong khi liên kết giữa các nhóm là ít chặt chẽ
- B. Dự đoán các kết nối tiềm năng giữa các cộng đồng trong mạng xã hội
- C. Xác định cộng đồng có liên kết chặt chẽ nhất trong mạng xã hội
- D. Phân tích các đặc điểm của từng cộng đồng trong mạng xã hội

ANSWER: A

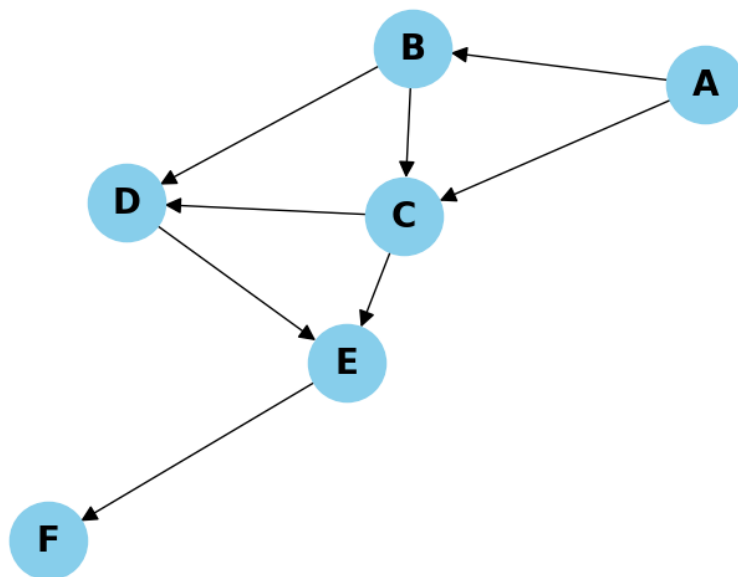
4. Cho mạng xã hội sau. Tính số đo bậc trung tâm (degree centrality) của nút 3.



- A. 0.6
- B. 0.83
- C. 0.4
- D. 0.5

ANSWER: A

5. Cho mạng xã hội là đồ thị có hướng như sau. Chọn phát biểu **sai**:

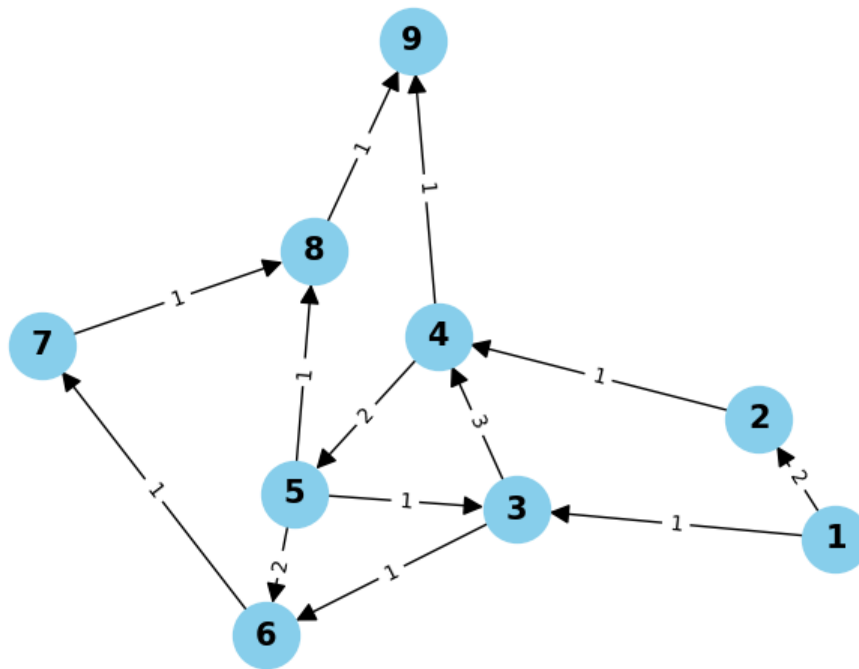


- A. A và D được kết nối bởi 4 đường đi
- B. Độ dài đường đi ngắn nhất từ A đến F là 3
- C. Lộ trình ABDCEF không phải là đường đi
- D. Đường đi ABDEF có độ dài 4

ANSWER: A

6. Cho mạng xã hội sau. In-degree & Out-degree của nút 4 có số lượng là bao nhiêu?

Weighted Directed Graph



- A. In-degree = 2, Out-degree = 2
- B. In-degree = 4, Out-degree = 3
- C. In-degree = 3, Out-degree = 4
- D. In-degree = 4, Out-degree = 4

ANSWER: A

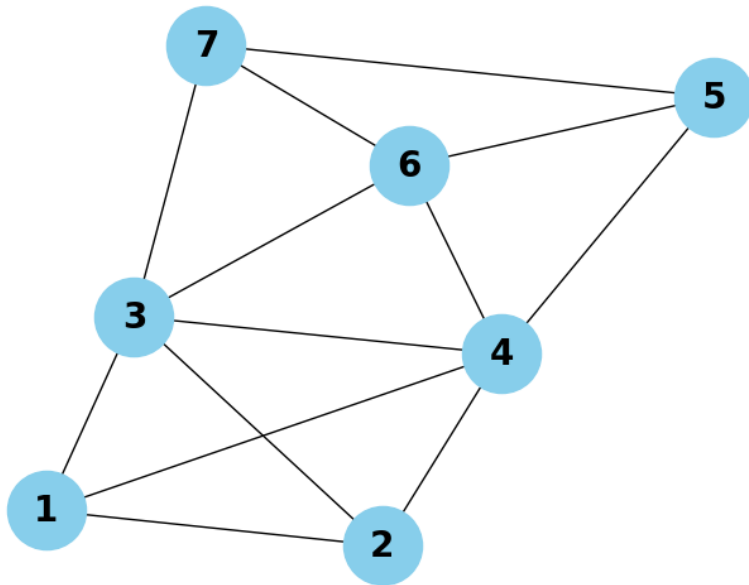
7. Áp dụng phương pháp thăm thâu Clique (CPM), hãy xác định cộng đồng từ các clique có kích thước 3 bên dưới.

{1, 2, 4}, {1, 3, 4}, {2, 4, 5}, {5, 6, 7}, {6, 7, 8}

- A. {1, 2, 3, 4, 5}, {5, 6, 7, 8}
- B. {1, 2, 3, 4}, {5, 6, 7, 8}
- C. {1, 2, 3, 4}, {2, 4, 5, 6, 7, 8}
- D. {1, 2, 3, 4}, {2, 4, 5}, {5, 6, 7, 8}

ANSWER: A

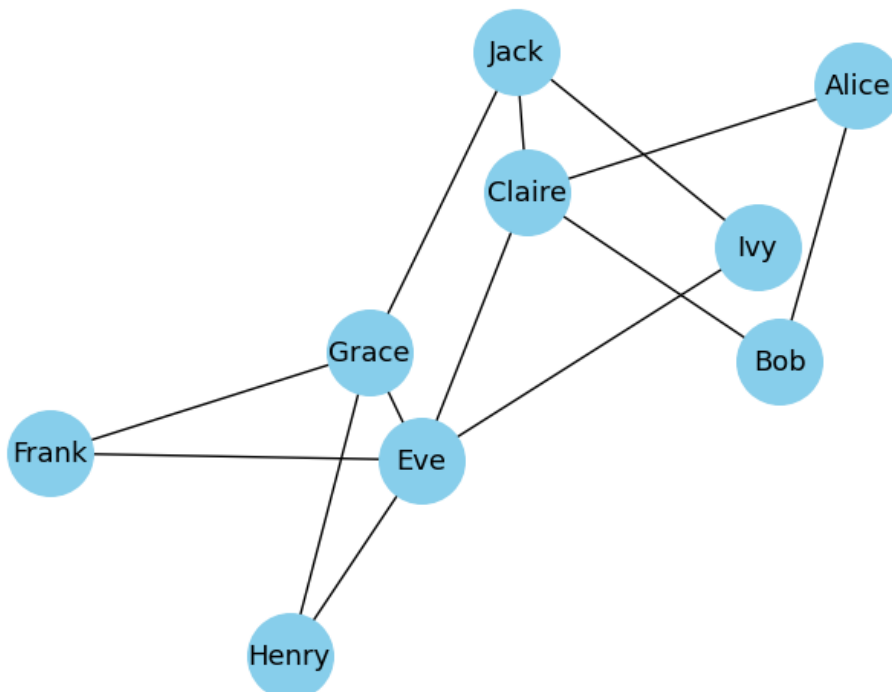
8. Cho mạng xã hội sau. Liệt kê tất cả clique cực đại (maximal clique).



- A. $\{1, 2, 3, 4\}, \{3, 4, 6\}, \{3, 6, 7\}, \{4, 5, 6\}, \{5, 6, 7\}$
 B. $\{1, 2, 3, 4\}, \{3, 4, 5, 6, 7\}$
 C. $\{1, 2, 3, 4\}, \{3, 4, 5, 7\}$
 D. $\{1, 2, 3\}, \{2, 3, 4\}, \{3, 4, 6\}, \{3, 6, 7\}, \{4, 5, 6\}, \{5, 6, 7\}$

ANSWER: A

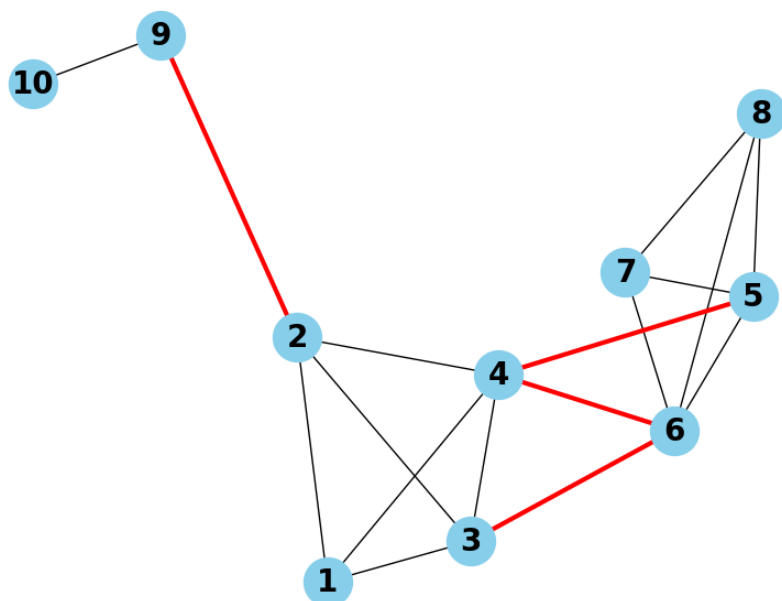
9. Một ngân hàng sử dụng phân tích mạng xã hội để phát hiện ra những khách hàng có đặc điểm chung trong việc sử dụng các dịch vụ tài chính. Tính độ tương đồng giữa Jack và Eve bằng chỉ số Cosine.



- A. 0.77
 B. 0.6
 C. 0.43
 D. 0.4

ANSWER: A

10. Cho mạng xã hội sau. Tạo một nhát cắt đi qua các cạnh màu đỏ. Tính giá trị Ratio cut.



- A. 1.33
- B. 0.62
- C. 0.56
- D. 0.45

ANSWER: A

11. Khi sử dụng mô hình hồi quy logistic để dự đoán liên kết trong mạng xã hội, phát biểu nào sau đây **sai**?

- A. Dữ liệu được chia thành tập hai tập, kiểm tra và huấn luyện, để tăng số lượng các liên kết tiềm năng trong mạng
- B. Các biến dự đoán (predictors) có thể bao gồm các đặc trưng như số bạn chung (common neighbors), chỉ số Adamic/Adar
- C. Mục tiêu chính của mô hình là dự đoán về xác suất liên kết giữa hai nút
- D. Kết quả của mô hình thường được đánh giá thông qua các chỉ số như Accuracy, AUC-ROC

ANSWER: A

12. Trong mạng xã hội, khi một người thay đổi hành vi hoặc quan điểm do ảnh hưởng từ người khác trong mạng, hiện tượng này được gọi là:

- A. Influence
- B. Social clustering
- C. Homophily
- D. Connection

ANSWER: A

PHẦN TỰ LUẬN (2 câu hỏi)

Câu hỏi 1: (1 điểm)

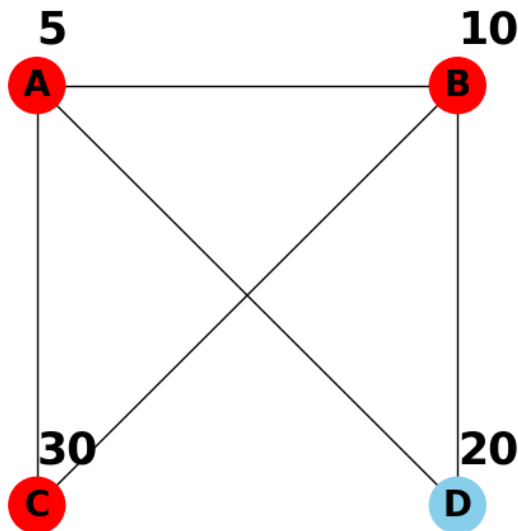
Một bệnh viện lớn phát triển mạng xã hội nội bộ nhằm kết nối các đối tượng liên quan trong lĩnh vực y tế. Mục tiêu là chia sẻ thông tin, hỗ trợ chẩn đoán, và cải thiện quy trình điều trị.

- Liệt kê 5 đối tượng có thể dùng để biểu diễn các nút trong mạng xã hội trên.
- Chọn hai nút bất kỳ trong câu a. Mô tả một mối quan hệ được thể hiện thông qua một cạnh kết nối hai nút đó.

****SV gõ trực tiếp trên khung trả lời của hệ thống thi.**

Câu hỏi 2: (3 điểm)

Ngân hàng X muốn sử dụng dữ liệu khách hàng của mình để phân tích rủi ro tín dụng. Dữ liệu mạng lưới được thể hiện dưới dạng mạng xã hội với các nút theo thứ tự A, B, C, D như hình sau:



với:

- Các nút A, B, C, D lần lượt là tên của các Khách hàng;
- Một cạnh nối 2 nút diễn tả giao dịch giữa 2 Khách hàng;
- Màu đỏ là những người có giới tính nam, màu xanh là những người có giới tính nữ;
- Các con số bên cạnh các nút thể hiện giá trị thuộc tính thu nhập, ví dụ Khách hàng A có thu nhập 5 triệu/tháng.

- Tính giá trị ma trận kề (A), ma trận chỉ báo (Δ), ma trận chỉ báo chuyển vị (Δ^T), ma trận bậc (d), ma trận bậc chuyển vị (d^T), tổng số cạnh (m). (1 điểm)
- Khi phân tích nhóm khách hàng nam ($\{A, B, C\}$), chuyên gia cho rằng những Khách hàng có thu nhập càng cao thì giao dịch càng nhiều. Tính giá trị hiệp phương sai (covariance) (0.5 điểm) và hệ số tương quan Pearson (Pearson correlation) (1 điểm), sau đó kết luận về nhận định trên (0.5 điểm).

****Làm tròn kết quả đến một chữ số thập phân.**

****SV gõ trực tiếp trên khung trả lời của hệ thống thi.**

ĐÁP ÁN PHẦN TỰ LUẬN VÀ THANG ĐIỂM

Phần câu hỏi	Nội dung đáp án	Thang điểm	Ghi chú
I. Trắc nghiệm		6.0	
Câu 1 – 12	ANSWER: A	0.5	
II. Tự luận		4.0	
Câu 1		1.0	
Nội dung a.	Liệt kê 5 đối tượng bất kỳ (cá nhân, tổ chức, thực thể) trong lĩnh vực y tế, không giới hạn trong danh sách bên dưới, mỗi đối tượng liệt kê đúng được 0.1 điểm. ● Bệnh nhân: bệnh nhân A , bệnh nhân B. ● Bác sĩ: bác sĩ nội khoa C, bác sĩ ngoại khoa D. ● Điều dưỡng: Điều dưỡng phòng Hồi sức, Điều dưỡng Khoa Nội. ● Khoa, phòng ban trong bệnh viện: Khoa Nội tiết, Khoa Tim mạch. ● Hệ thống thiết bị y tế: Máy đo đường huyết, Máy đo huyết áp. ● Nhà cung cấp: Công ty dược phẩm ABC, Nhà phân phối thiết bị y tế XYZ. Ví dụ: bệnh nhân A , bệnh nhân B, bác sĩ nội khoa C, Điều dưỡng Khoa Nội, Khoa Nội tiết.	0.5	
Nội dung b.	Mô tả đúng được mối quan hệ trong lĩnh vực giáo dục giữa 2 đối tượng. Ví dụ: ● Bệnh nhân ↔ Bác sĩ: bệnh nhân B có lịch khám định kỳ với BS. Ngoại khoa D. ● Bác sĩ ↔ Nhà cung cấp: bác sĩ nội thận C hợp tác với Công ty ABC để cập nhật các loại thuốc mới.	0.5	
Câu 2		3.0	

Nội dung a.	• A là ma trận 4 dòng 4 cột: 0 1 1 1 1 0 1 1 1 1 0 0 1 1 0 0 (0.2 điểm) • delta là ma trận 4 dòng 2 cột: 1 0 1 0 1 0 0 1 (0.2 điểm) • delta^T là ma trận 2 dòng 4 cột: 1 1 1 0 0 0 0 1 (0.2 điểm) • d là ma trận 4 dòng 1 cột: 3 3 2 2 (0.2 điểm) • d^T là ma trận 1 dòng 4 cột: 3 3 2 2 (0.1 điểm) • $m = 5$ (0.1 điểm)	1.0	
Nội dung b.	• Tính hiệp phương sai Lập danh sách các cạnh, khởi tạo XL và XR $XL = [5, 10, 10, 30, 5, 30]$ $XR = [10, 5, 30, 10, 30, 5]$ Tính kỳ vọng $E(XL)$ $E(XL) = (5+10+10+30+5+30)/6 = 15$ (0.25 điểm) Tính $E(XL \odot XR)$	2.0	

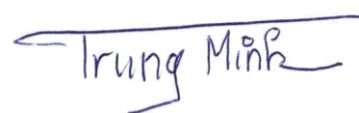
	$XL \odot XR = [5*10, 10*5, 10*30, 30*10, 5*30, 30*5]$ $= [50, 50, 300, 300, 150, 150]$ $E(XL \odot XR) = 166.7$ (0.25 điểm) $\text{cov}(XL, XR) = E(XL \odot XR) - E(XL)^2$ $= 166.7 - 15^2$ $= -58.3$ (0.25 điểm) - Tính hệ số tương quan Pearson $E(XL^2) = (25+100+100+900+25+900)/6 = 341.7$ (0.25 điểm) Phương sai $\sigma(XL)^2 = E(XL^2) - E(XL)^2$ $= 341.7 - 225 = 116.7$ (0.25 điểm) $\text{corr}(XL, XR) = \text{cov}(XL, XR) / (\sigma(XL)^2)$ $= -58.3 / 116.7$ $= -0.5$ (0.25 điểm) - Kết luận: Do $\text{corr}(XL, XR) = -0.5 < 0$, ta bác bỏ nhận định rằng những Khách hàng có thu nhập càng cao thì giao dịch càng nhiều. (0.5 điểm)		
	Điểm tổng	10.0	

TP. Hồ Chí Minh, ngày 22 tháng 11. năm 2024

Người duyệt đề

Giảng viên ra đề


Nguyễn Ngọc Chánh



Trần Trung Minh