

ĐẠI HỌC BÁCH KHOA HÀ NỘI  
TRƯỜNG ĐIỆN - ĐIỆN TỬ



Báo cáo  
**BÀI TẬP LỚN**

Đề tài:

**XÂY DỰNG MẠNG ĐÁNH GIÁ SỰ BIẾN ĐỘNG  
CỦA THỊ TRƯỜNG CHỨNG KHOÁN**

Học viên thực hiện:

TẠ QUANG MINH      24B-ET-KTĐT      MSHV: 20242537M

Giảng viên hướng dẫn: TS. PHẠM VĂN TIẾN

Hà Nội, 07-2024

# LỜI NÓI ĐẦU

Trong bối cảnh thị trường tài chính ngày càng biến động mạnh và khó lường, việc phân tích rủi ro và cơ hội đầu tư vào cổ phiếu không chỉ dừng lại ở các chỉ số định lượng truyền thống như giá, khối lượng giao dịch hay P/E, mà cần đến những cách tiếp cận mới có khả năng mô tả cấu trúc ẩn trong hành vi thị trường. Mạng phức (complex networks) là một công cụ ngày càng được quan tâm trong việc biểu diễn và phân tích các mối quan hệ phi tuyến và sự phụ thuộc tiềm ẩn giữa các yếu tố trong hệ thống tài chính. Nội dung của báo cáo được trình bày trong các chương sau:

- Chương 1: Giới thiệu chung
- Chương 2: Mô tả dữ liệu và mô hình hóa mạng
- Chương 3: Các tham số đặc trưng của mạng
- Chương 4: Kết luận và hướng phát triển

Em xin chân thành cảm ơn TS Phạm Văn Tiến đã tận tình hướng dẫn, giảng dạy, cung cấp những kiến thức quan trọng trong suốt quá trình học tập và thực hiện đề tài để em có thể thực hiện thành công bài tập lớn này. Em xin chân thành cảm ơn!

Hà Nội, ngày 19 tháng 07 năm 2025

**Sinh viên thực hiện**

**TẠ QUANG MINH**

# MỤC LỤC

<b>DANH MỤC HÌNH VẼ</b>	<b>5</b>
<b>CHƯƠNG 1. GIỚI THIỆU CHUNG</b>	<b>1</b>
1.1 Bối cảnh . . . . .	1
1.2 Mục tiêu của dự án . . . . .	1
1.3 Phạm vi thực hiện . . . . .	1
1.4 Mục tiêu . . . . .	2
<b>CHƯƠNG 2. MÔ TẢ DỮ LIỆU VÀ MÔ HÌNH HÓA MẠNG</b>	<b>3</b>
2.1 Dữ liệu đầu vào và tiền xử lý . . . . .	3
2.1.1 Dữ liệu đầu vào . . . . .	3
2.1.2 Tiền xử lý . . . . .	3
2.2 Định nghĩa node và cạnh trong mạng . . . . .	4
2.2.1 Cấu trúc node . . . . .	6
2.2.2 Định nghĩa cạnh có hướng và có trọng số . . . . .	6
2.2.3 Trọng số cạnh . . . . .	6
2.3 Mô tả đặc trưng mạng thu được . . . . .	7
2.4 Tổng kết phần . . . . .	8
<b>CHƯƠNG 3. CÁC THAM SỐ ĐẶC TRƯNG CỦA MẠNG</b>	<b>9</b>
3.1 Phân bố bậc . . . . .	9
3.2 Đường path . . . . .	10
3.2.1 Đường đi ngắn nhất . . . . .	10
3.2.2 Đường đi Hamilton . . . . .	11
3.3 Đường đi độc lập . . . . .	12
3.4 Thành phần liên thông mạnh (Strongly Connected Components) . . . .	12
3.5 Khuếch tán trong mạng(Diffusion in directed weighted network) . . . .	13
3.5.1 Đặc trưng cấu trúc cho khuếch tán . . . . .	13

3.5.2	Đo lường khả năng khuếch tán . . . . .	14
3.6	Katz/ PageRank . . . . .	14
3.7	K-core . . . . .	15
3.8	Tính bậc cầu và hệ số phân cụm . . . . .	16
3.9	Đánh giá tổng thể mạng . . . . .	17
<b>CHƯƠNG 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN</b>		<b>19</b>
4.1	Kết luận . . . . .	19
4.2	Hướng phát triển . . . . .	19
<b>KẾT LUẬN</b>		<b>i</b>
<b>TÀI LIỆU THAM KHẢO</b>		<b>ii</b>

## DANH MỤC HÌNH VẼ

Hình 2.1	Minh họa dữ liệu đầu vào . . . . .	4
Hình 2.2	Minh họa cách chuẩn hóa biến động . . . . .	5
Hình 2.3	Minh họa dữ liệu đã được xử lý . . . . .	5
Hình 2.4	Đồ thị dữ liệu 30 ngày đầu . . . . .	6
Hình 2.5	Mạng xây dựng 30 node đầu . . . . .	7
Hình 2.6	Histogram phân bố trọng số cạnh . . . . .	8
Hình 3.1	Minh họa phân bố bậc vào và bậc ra của toàn bộ mạng được xây dựng từ chuỗi biến động cổ phiếu . . . . .	9
Hình 3.2	Đồ thị phân bố đường đi ngắn nhất giữa các cặp node . . . . .	11
Hình 3.3	Mạng khuếch tán từ node 272 . . . . .	15
Hình 3.4	Top 10 Node có Katz, PageRank cao nhất . . . . .	15

# CHƯƠNG 1. GIỚI THIỆU CHUNG

Chương 1 đưa ra các vấn đề hiện tại, mục tiêu, phạm vi thực hiện của đề tài.

## 1.1 Bối cảnh

Thị trường chứng khoán là một hệ thống kinh tế phức tạp, nơi các mã cổ phiếu liên tục biến động dưới tác động của nhiều yếu tố như tâm lý nhà đầu tư, tin tức kinh tế, chính sách tài khóa – tiền tệ, và thậm chí là tương tác giữa các mã cổ phiếu với nhau. Đối với mỗi nhà đầu tư, việc đánh giá rủi ro và cơ hội khi tham gia vào một mã cổ phiếu cụ thể là vấn đề cốt lõi, quyết định hiệu quả của chiến lược đầu tư. Tuy nhiên, các phương pháp truyền thống như phân tích kỹ thuật hay phân tích cơ bản thường gặp khó khăn trong việc mô tả đầy đủ bản chất động học và cấu trúc tiềm ẩn trong chuỗi biến động giá theo thời gian.

## 1.2 Mục tiêu của dự án

Trong bối cảnh đó, lý thuyết mạng phức (complex network theory) mở ra một hướng tiếp cận mới, cho phép mô hình hóa các chuỗi thời gian biến động cổ phiếu thành các đồ thị có cấu trúc cụ thể, từ đó trích xuất ra những đặc trưng cấu trúc mang giá trị thông tin cao. Cách tiếp cận này không chỉ giúp hình dung sự phức tạp trong hành vi của một mã cổ phiếu mà còn cho phép so sánh, đánh giá mức độ tương đồng giữa các cổ phiếu khác nhau thông qua đặc trưng mạng.

## 1.3 Phạm vi thực hiện

Nghiên cứu này tập trung vào việc xây dựng các mạng phức có hướng và có trọng số từ chuỗi biến động giá của từng mã cổ phiếu riêng biệt. Mỗi mạng tương ứng với một mã cổ phiếu trong một khoảng thời gian đủ dài. Sau khi xây dựng mạng, các đại lượng đặc trưng như bậc vào/ra, hệ số phân cụm, độ dài đường đi ngắn nhất, centrality, độ liên thông, v.v. sẽ được tính toán và phân tích để làm rõ mức độ ổn định, mức độ lan truyền rủi ro, và các đặc tính cấu trúc nổi bật của mã cổ phiếu đó.

Đặc biệt, nghiên cứu còn mở rộng theo hướng so sánh các mạng này với nhau để đánh giá sự tương quan về cấu trúc mạng giữa các mã cổ phiếu. Thông qua các chỉ số đo tương đồng như cosine similarity, eigenvalue-based similarity hoặc các công cụ phân cụm phổ (spectral clustering), nghiên cứu hướng đến việc xác định các nhóm cổ phiếu có hành vi giao động tương đồng – một yếu tố quan trọng trong phân tích danh mục đầu tư và dự báo xu hướng thị trường.

## **1.4 Mục tiêu**

Tổng thể, nghiên cứu này hướng đến mục tiêu kép: (1) Phân tích định lượng mức độ phức tạp và ổn định của từng mã cổ phiếu thông qua đặc trưng mạng phức; và (2) Phát hiện cấu trúc tương quan giữa các cổ phiếu nhằm gợi ý chiến lược đầu tư hiệu quả hơn dựa trên cấu trúc mạng ẩn dưới dữ liệu giá.

## CHƯƠNG 2. MÔ TẢ DỮ LIỆU VÀ MÔ HÌNH HÓA MẠNG

Chương 2 trình bày cách chuyển đổi dữ liệu biến động giá cổ phiếu thành các mạng phức có hướng và có trọng số. Quá trình bao gồm việc xử lý chuỗi thời gian, xây dựng mạng, định nghĩa các node và cạnh, và chuẩn hóa dữ liệu để phân tích mạng.

### 2.1 Dữ liệu đầu vào và tiền xử lý

#### 2.1.1 Dữ liệu đầu vào

Để đảm bảo tính cập nhật và gần gũi với thị trường thực tế, nghiên cứu lựa chọn sử dụng dữ liệu giá cổ phiếu từ thị trường chứng khoán Việt Nam, thu thập thông qua thư viện mã nguồn mở VNstock. VNstock là một công cụ Python phổ biến, cho phép truy xuất linh hoạt dữ liệu giá lịch sử từ các mã cổ phiếu niêm yết trên sàn HOSE, HNX và UPCOM.

Trong phạm vi nghiên cứu, dữ liệu được thu thập đối với năm mã cổ phiếu: FPT, SSI, VCB, VHM và ACB. Đây là năm mã được chọn hoàn toàn ngẫu nhiên nhằm đảm bảo tính khách quan, không phụ thuộc vào quy mô vốn hóa hay ngành nghề. Dữ liệu lịch sử được lấy trong khoảng thời gian từ ngày 02/01/2024 đến ngày 27/03/2025, tương ứng với hơn một năm giao dịch thực tế. Sau khi loại bỏ các ngày không giao dịch (nghỉ lễ, cuối tuần), mỗi mã cổ phiếu có khoảng 300 điểm dữ liệu, tương ứng với hơn 300 node trong mạng.

Các trường thông tin thu thập bao gồm:

- Ngày giao dịch (date).
- Giá mở cửa, cao nhất, thấp nhất và đóng cửa.
- Khối lượng giao dịch.

#### 2.1.2 Tiền xử lý

Sau khi thu thập dữ liệu giá cổ phiếu, bước đầu tiên trong tiền xử lý là loại bỏ các ngày không có giao dịch, tức các bản ghi chứa giá trị thiếu (NaN) trong các cột giá. Điều này giúp đảm bảo chuỗi thời gian liên tục và không gây sai lệch trong quá trình xây dựng mạng.

Tiếp theo, do nghiên cứu hướng tới việc xây dựng mạng phức từ chuỗi thời gian, mỗi điểm dữ liệu (ứng với một ngày giao dịch) được biểu diễn bởi một mức độ dao động tổng hợp, thay vì sử dụng từng giá trị riêng biệt như Open, High, Low, Close [1].

Phương pháp chuẩn hóa biến động (dao động giá):



time	open	high	low	close
2024-01-02 00:00:00	83.11	83.11	82.08	82.25
2024-01-03 00:00:00	81.99	82.76	81.82	82.59
2024-01-04 00:00:00	82.59	84.39	82.51	83.36
2024-01-05 00:00:00	83.45	83.62	82.76	83.53
2024-01-08 00:00:00	83.53	84.31	83.28	83.45
2024-01-09 00:00:00	83.45	83.45	82.85	82.85
2024-01-10 00:00:00	82.93	83.11	81.05	81.22
2024-01-11 00:00:00	81.22	81.82	81.22	81.48

**Hình 2.1 Minh họa dữ liệu đầu vào**

Mức độ dao động trong một phiên được tính dựa trên tổng diện tích tạo thành bởi các đoạn thẳng nối 4 mức giá [1]:

- A: Giá mở cửa (Open)
- B: Giá cao nhất (High)
- C: Giá thấp nhất (Low)
- D: Giá đóng cửa (Close)

Trên trục thời gian, ta quy ước các vị trí:  $x_A = 0, x_B = 1, x_C = 2, x_D = 3$

Từ đó, tạo thành ba đoạn thẳng nối tiếp: AB, BC, CD.

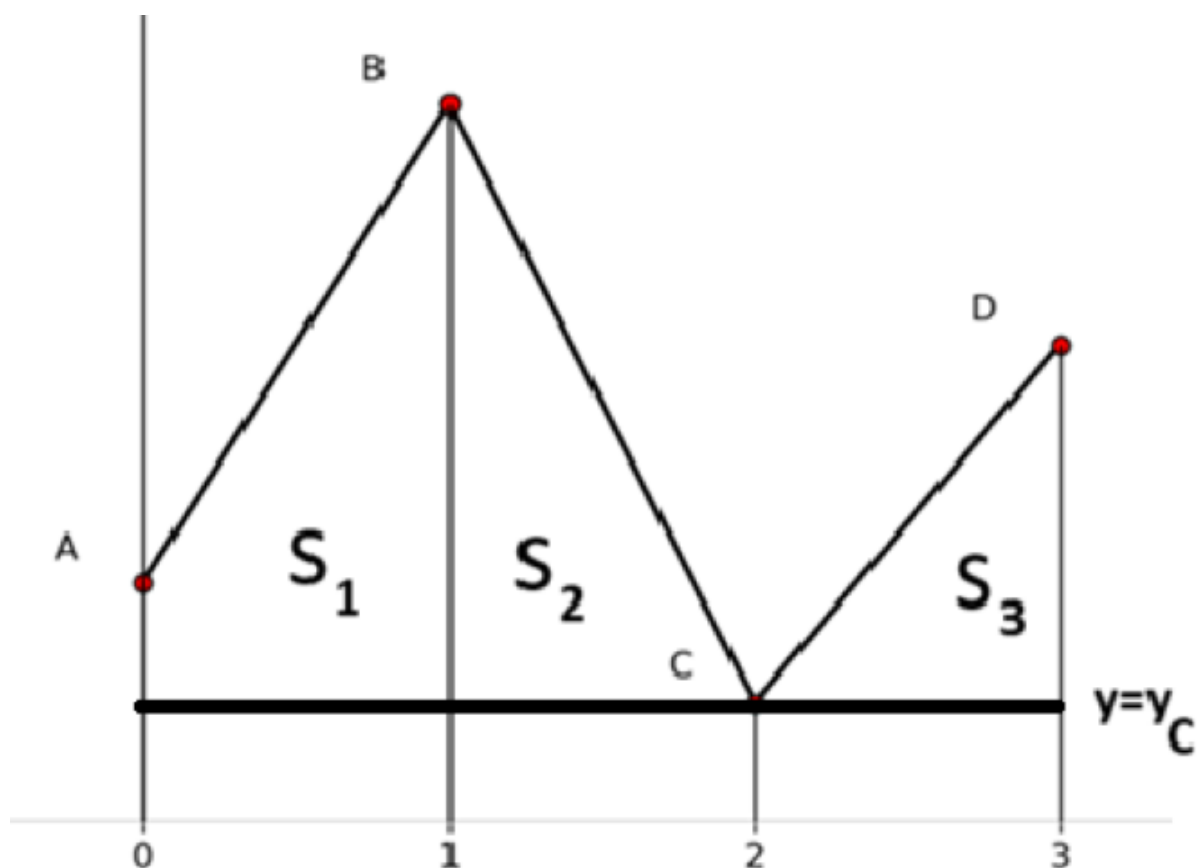
Với mỗi đoạn, diện tích hình thang hoặc tam giác mà đoạn đó tạo với đường thẳng  $y = y_c$  được ký hiệu lần lượt là  $S_1, S_2, S_3$ . Mức độ dao động tổng cộng của phiên đó được tính bằng tổng các diện tích theo mô tả của hình 2.2:  $S_t = S_1 + S_2 + S_3$

Giá trị  $S_t$  này đại diện cho mức độ dao động tổng hợp của phiên giao dịch và sẽ được sử dụng làm đặc trưng chính để tạo node trong mạng phức. Mỗi node biểu diễn một ngày giao dịch, còn các cạnh được xác định dựa trên thứ tự thời gian và mối quan hệ giữa các mức dao động.

Bước xử lý này giúp nén thông tin giá theo ngày thành một chỉ số duy nhất mà vẫn giữ được đặc trưng động lực của thị trường, đồng thời làm đơn giản hóa mô hình mạng thời gian nhưng vẫn giữ được độ nhạy với biến động giá [2].

## 2.2 Định nghĩa node và cạnh trong mạng

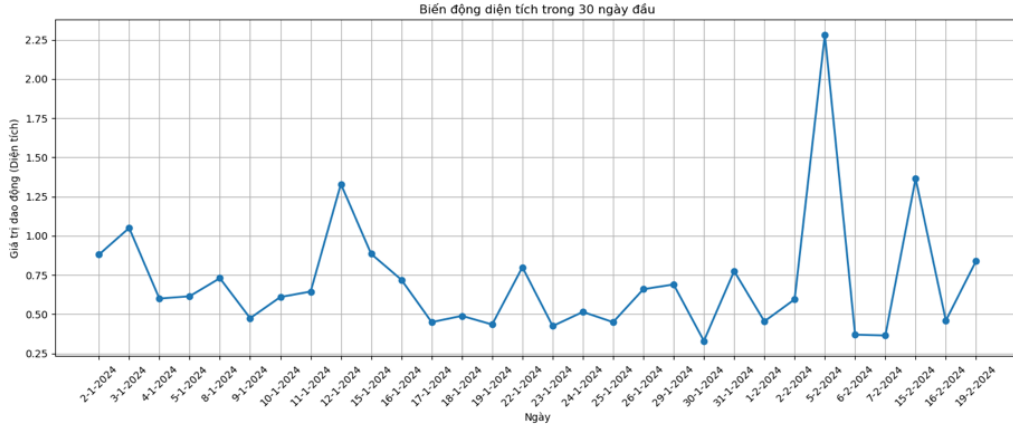
Sau khi tính toán mức độ dao động tổng hợp  $S_t$  cho mỗi phiên giao dịch, ta tiến hành mô hình hóa chuỗi thời gian này thành một mạng phức. Trong mạng, mỗi node



Hình 2.2 Minh họa cách chuẩn hóa biến động

date	daodong
2/1/2024	0.88
3/1/2024	1.05
4/1/2024	0.6
5/1/2024	0.615
8/1/2024	0.73
9/1/2024	0.475
10/1/2024	0.61
11/1/2024	0.645
12/1/2024	1.33

Hình 2.3 Minh họa dữ liệu đã được xử lý



**Hình 2.4 Đồ thị dữ liệu 30 ngày đầu**

tương ứng với một ngày giao dịch, và đặc trưng bởi mức dao động tại thời điểm đó.

### 2.2.1 Cấu trúc node

Một node  $v_i$  được định nghĩa là một cặp:  $Node_i = (t_i, Y_{t_i})$  [3]

Trong đó:

- $t_i$ : là thời điểm (ngày) thứ  $i$  trong chuỗi thời gian.
- $Y_{t_i}$ : là giá trị dao động đã tính tại ngày  $t_i$

### 2.2.2 Định nghĩa cạnh có hướng và có trọng số

Mạng được xây dựng với cạnh có hướng, mô tả sự “nhìn thấy” giữa các node theo nguyên lý Horizontal Visibility Graph (HVG) mở rộng

Một cạnh có hướng  $v_i \rightarrow v_j$  với  $i < j$  tồn tại nếu thỏa mãn điều kiện [4]:

$$Y_k < \min(Y_i, Y_j), \forall k \in (i, j) \quad (2.1)$$

Nói cách khác, mọi node nằm giữa  $v_i$  và  $v_j$  đều phải có giá trị dao động thấp hơn cả hai đầu mút. Điều này đảm bảo rằng  $v_i$  và  $v_j$  có thể "nhìn thấy nhau" trên đồ thị thời gian - tương tự như điểm cao trên một chuỗi núi không bị che khuất bởi điểm giữa

### 2.2.3 Trọng số cạnh

Mỗi cạnh  $v_i \rightarrow v_j$  được gán trọng số phản ánh tốc độ thay đổi dao động giữa hai thời điểm [4]:

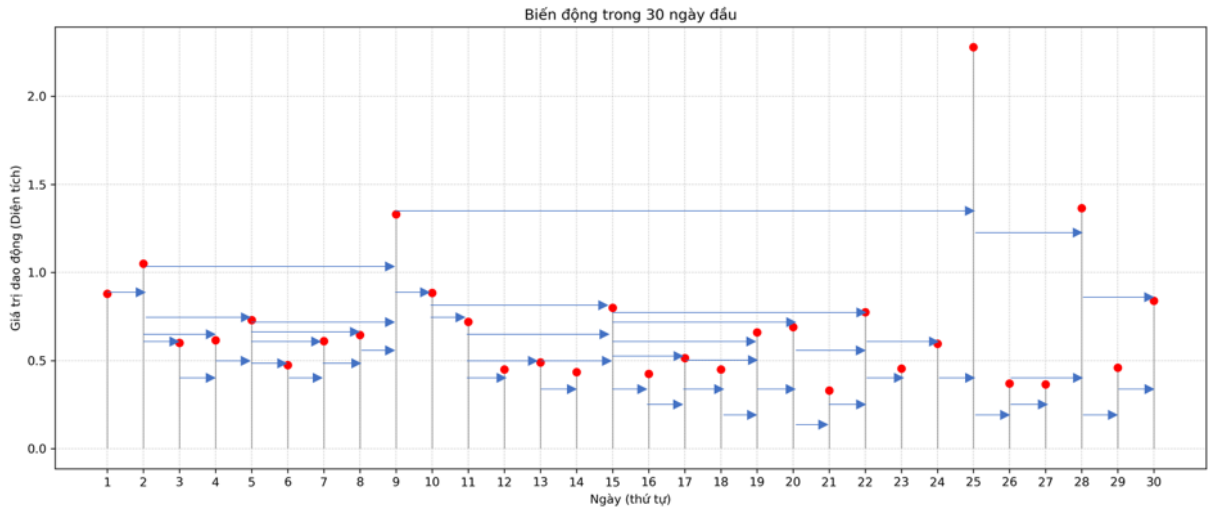
$$w_{i \rightarrow j} = \frac{Y_i}{j - i} \quad (2.2)$$

Trong đó:

$Y_i$ : là giá trị dao động tại ngày  $i$

$j - i$ : là khoảng cách thời gian giữa hai node

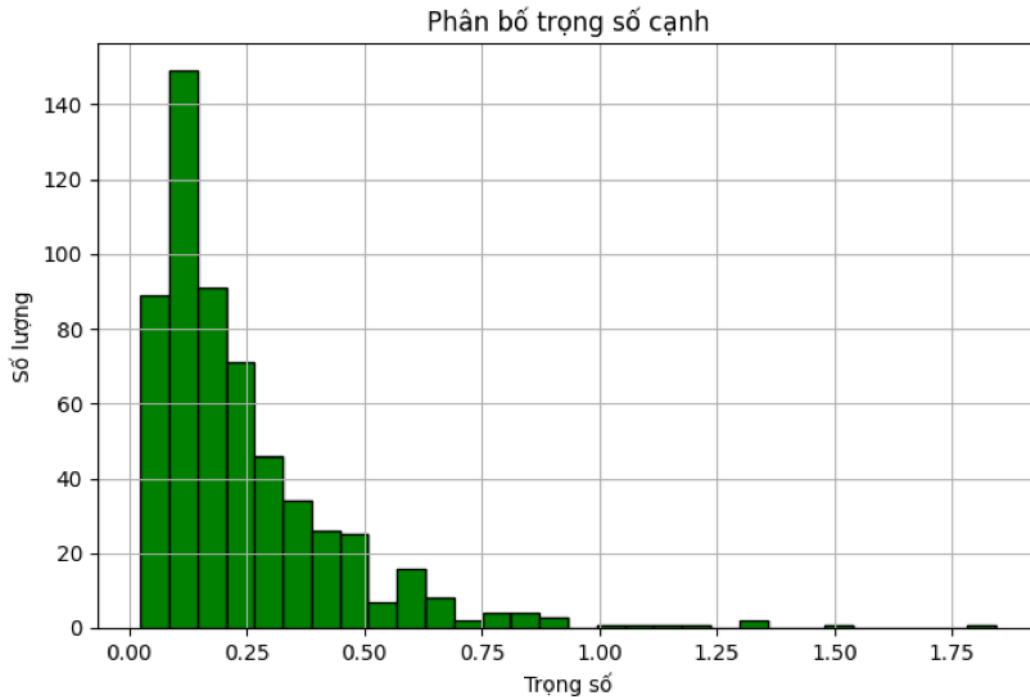
Việc đưa trọng số này giúp mạng vừa giữ được cấu trúc hình học của chuỗi thời gian, vừa cho phép đánh giá tốc độ lan truyền hoặc biến động giữa các node [5].



**Hình 2.5 Mạng xây dựng 30 node đầu**

### 2.3 Mô tả đặc trưng mạng thu được

- Số lượng node  $n$ : 306 node
- Số cạnh  $m$ : 583 cạnh
- Bậc vào:  $k_i^{in} = \sum_j A_{ij}$ 
  - o Bậc vào nhỏ nhất: 0
  - o Bậc vào lớn nhất: 8
- Bậc ra:  $k_i^{out} = \sum_j A_{ji}$ 
  - o Bậc ra nhỏ nhất: 0
  - o Bậc ra lớn nhất: 7
- Trung bình bậc vào và bậc ra:  $c^{in} = \frac{1}{n} \sum_{i=1}^n k_i^{in} = \frac{1}{n} \sum_{i=1}^n k_i^{out} = c^{out}$ 
  - o Kết quả: 1.90522876



**Hình 2.6 Histogram phân bố trọng số cạnh**

Ý nghĩa: Trong mạng HVG tài chính, các node có bậc ra cao thường đại diện cho những phiên giao dịch có ảnh hưởng mạnh về mặt hình học tới các phiên sau, đóng vai trò như tâm lan truyền. Ngược lại, node có bậc vào cao là những phiên bị tác động bởi nhiều phiên trước đó - chúng có thể là kết quả của nhiều biến động hội tụ.

Việc xác định các node có bậc trung tâm cao giúp [6]:

- Phát hiện các phiên giao dịch có vai trò then chốt trong dao động giá.
- Xác định các giai đoạn có tính bất ổn cao (nhiều tương tác).
- Làm nền cho việc tính toán các chỉ số cao cấp hơn như Katz, PageRank.

Chỉ số này có thể sử dụng để đánh giá sơ bộ mức độ “nổi bật” của từng phiên, trước khi xét đến các yếu tố truyền lan hoặc ảnh hưởng toàn cục

## 2.4 Tổng kết phần

Qua phần này, em đã chuyển đổi thành công chuỗi thời gian biến động giá của từng mã cổ phiếu thành các mạng phức có hướng và có trọng số. Đây là nền tảng để phân tích sâu hơn các đặc trưng cấu trúc và mối tương quan giữa các mã trong các phần tiếp theo.

## CHƯƠNG 3. CÁC THAM SỐ ĐẶC TRƯNG CỦA MẠNG

### 3.1 Phân bố bậc

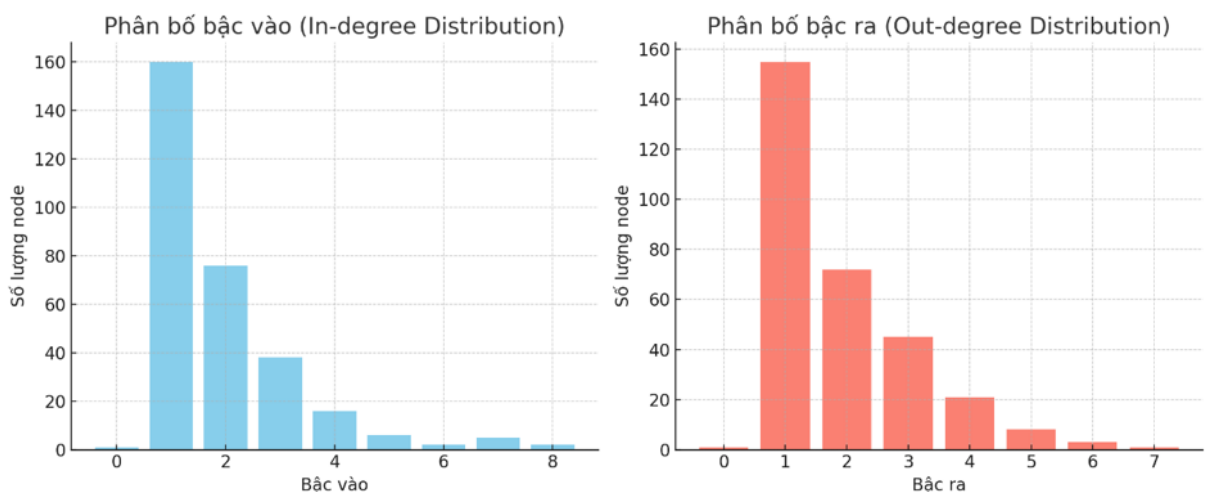
Một trong những đặc trưng cơ bản và quan trọng nhất trong phân tích mạng là phân bố bậc – tức số lượng kết nối (cạnh) mà mỗi node có với các node khác trong mạng. Đối với mạng có hướng, bậc được chia thành hai loại:

- Bậc vào (in-degree): số lượng cạnh đi vào một node.
- Bậc ra (out-degree): số lượng cạnh đi ra từ một node.

Kết quả cho thấy [7]:

- Phần lớn các node có bậc vào và bậc ra thấp (1–3), điều này phản ánh rằng trong chuỗi thời gian, phần lớn các ngày giao dịch chỉ có thể "nhìn thấy" một vài ngày khác gần kề, theo tiêu chí HVG.
- Một số ít node có bậc rất cao, đóng vai trò như các "trung tâm dao động" – các điểm dao động mạnh hoặc có tính quan sát cao trong chuỗi giá.
- Sự chênh lệch này cho thấy mạng có cấu trúc không đồng đều, nghiêng về phía phân bố lệch phải (right-skewed), tiệm cận với mô hình phân bố mũ – một đặc trưng thường thấy trong các hệ thống động.

Đây là bằng chứng cho thấy sự tồn tại của các node có vai trò trung tâm trong mạng, có thể đóng vai trò cảnh báo rủi ro hoặc chỉ báo nhịp biến động mạnh trên thị trường



**Hình 3.1 Minh họa phân bố bậc vào và bậc ra của toàn bộ mạng được xây dựng từ chuỗi biến động cổ phiếu**

Đây là bằng chứng cho thấy sự tồn tại của các node có vai trò trung tâm trong

mạng, có thể đóng vai trò cảnh báo rủi ro hoặc chỉ báo nhịp biến động mạnh trên thị trường.

### 3.2 Đường path

Trong mạng phức có hướng và có trọng số được xây dựng từ chuỗi thời gian cổ phiếu, việc phân tích các đường đi là công cụ then chốt để đánh giá tốc độ lan truyền biến động, khả năng tiếp cận thông tin, và tính liên thông tổng thể của mạng.

Theo định nghĩa [8]:

- Path (đường đi) là chuỗi các node liên tiếp mà giữa mỗi cặp liên tiếp có tồn tại cạnh nối.
- Shortest path (đường đi ngắn nhất) là đường đi giữa hai node có tổng trọng số nhỏ nhất hoặc số bước đi ít nhất.
- Hamilton path là đường đi đi qua tất cả các node đúng một lần – một chỉ số quan trọng phản ánh mức độ “thống nhất thứ tự thời gian” hoặc chu kỳ có thể xảy ra trong chuỗi dao động.

#### 3.2.1 Đường đi ngắn nhất

Kết quả phân tích các đường đi ngắn nhất giữa mọi cặp node (dựa trên số bước đi, không xét trọng số) cho thấy:

- Độ dài trung bình đường đi ngắn nhất giữa các node có thể kết nối được với nhau là 7.84 bước.
- Đường đi ngắn nhất dài nhất (đường kính) trong mạng là 31 bước.
- Hầu hết các node có thể đến được với nhau trong khoảng dưới 10 bước, thể hiện tính lan truyền tương đối tốt trong chuỗi thời gian cổ phiếu.

Các đường đi ngắn nhất (geodesic path) đóng vai trò then chốt trong tính toán các đại lượng như closeness centrality, đường kính, độ bền mạng, và mô hình lan truyền.

Ý nghĩa:

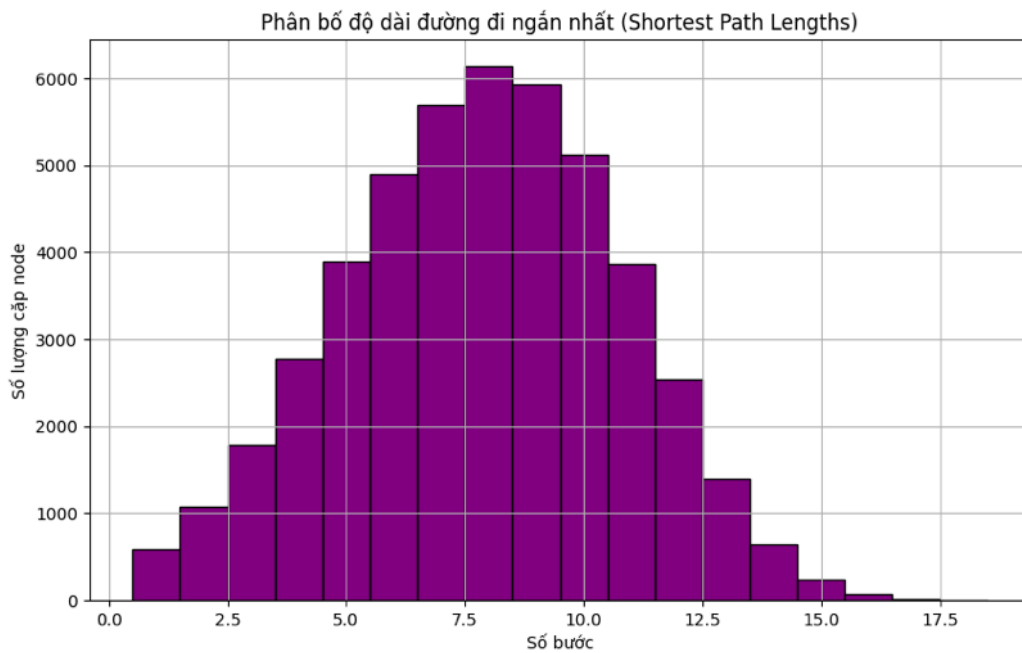
Điều này phản ánh rằng trong chuỗi dao động giá cổ phiếu:

- Tín hiệu dao động có thể lan truyền khá nhanh từ một phiên sang các phiên sau, mà không cần quá nhiều bước trung gian.
- Một phiên giao dịch có thể ảnh hưởng đến các phiên cách nó hàng chục bước, miễn là còn “nhìn thấy nhau” trong cấu trúc HVG.

Ngoài ra, độ dài đường đi ngắn nhất là một thành phần quan trọng trong việc tính các chỉ số như:

- Closeness centrality: phản ánh node nào có thể lan truyền nhanh tới toàn mạng.
- Độ bền mạng (robustness): mạng có đường đi ngắn → dễ khôi phục thông tin nếu bị cắt.
- Tối ưu hóa mô hình truyền tin (diffusion models) [9].

Tóm lại, độ dài trung bình nhỏ và đường kính vừa phải cho thấy mạng HVG tuy không khép kín nhưng có cấu trúc tuyến tính hiệu quả, rất phù hợp để phân tích sự lan truyền của dao động giá theo thời gian.



**Hình 3.2 Đồ thị phân bố đường đi ngắn nhất giữa các cặp node**

### 3.2.2 Đường đi Hamilton

Việc kiểm tra Hamilton path là một bài toán NP-complete. Với mạng có hơn 300 node, việc xác định Hamilton path bằng brute-force là bất khả thi. Trong kiểm tra mẫu (10 node đầu tiên), không phát hiện được Hamilton path, do vậy:

- Hiện tại không thể khẳng định có tồn tại đường Hamilton, nhưng cũng không thể bác bỏ mà không dùng thuật toán tối ưu chuyên biệt.
- Từ góc nhìn thực tế, do dữ liệu là chuỗi thời gian, nên chỉ một vài đoạn chuỗi con có thể gần với hành vi Hamilton – nhưng toàn bộ chuỗi thì không.

Đường Hamilton là loại đường đi qua toàn bộ các node đúng 1 lần, phản ánh tiềm năng có chuỗi dao động nhất quán hoặc sự kiện ảnh hưởng toàn bộ lịch sử.



### 3.3 Đường đi độc lập

Đường đi độc lập là một khái niệm quan trọng trong phân tích khả năng kết nối và mức độ bền vững của mạng. Theo lý thuyết, có hai loại đường đi độc lập:

- Vertex-independent paths (VIP): hai đường đi không dùng chung node trung gian nào.
- Edge-independent paths (EIP): hai đường đi không dùng chung cạnh nào

Số lượng đường đi độc lập giữa hai node phản ánh mức độ kết nối chặt chẽ giữa chúng, hay còn gọi là connectivity của cặp node. Nếu connectivity cao, thì hai node đó có nhiều con đường thay thế để duy trì liên lạc, giảm thiểu rủi ro khi một phần mạng bị hỏng – đây chính là khái niệm "bottleneck".

Kết quả tính toán trên tập dữ liệu cho thấy (tính trên 10 node tiêu biểu đầu tiên):

Hầu hết các cặp node đều có [3]:

- 1 đường đi vertex-independent (VIP)
- 1 đường đi edge-independent (EIP)

Ví dụ, giữa các cặp node như  $(0 \rightarrow 1)$ ,  $(0 \rightarrow 2)$ ,  $(0 \rightarrow 4)$ ... đều có duy nhất một đường đi không trùng cạnh và không trùng đỉnh trung gian. Điều này cho thấy mạng mang tính chuỗi thời gian đơn tuyến, ít có các nhánh phụ hoặc vòng lặp song song giữa các điểm thời gian.

Ý nghĩa:

- Mạng không có tính dự phòng cao: nếu một node hoặc cạnh bị loại bỏ, đường đi giữa hai node rất dễ bị cắt đứt hoàn toàn.
- Điều đó đồng nghĩa, chuỗi dao động cổ phiếu mang tính tuyến tính mạnh, không có khả năng lan truyền song song hoặc hội tụ ảnh hưởng từ nhiều nguồn.
- Tồn tại nhiều bottleneck (cổ chai thông tin), là các node trung gian duy nhất duy trì liên kết giữa các phần của mạng.

Từ góc nhìn ứng dụng, các node nằm trên nhiều đường đi độc lập là điểm nhạy cảm, và cần được giám sát kỹ vì sự gián đoạn của chúng có thể phá vỡ liên kết thông tin trong chuỗi giá. Đây cũng là một chỉ số hỗ trợ tốt cho việc phát hiện rủi ro tiềm ẩn trong chuỗi dao động.

### 3.4 Thành phần liên thông mạnh (Strongly Connected Components)

Trong mạng có hướng, thành phần liên thông mạnh (SCC) là tập hợp các node mà mọi node trong nhóm đều có đường đi đến tất cả các node còn lại theo đúng hướng cạnh. Điều này phản ánh mức độ khép kín và vòng lặp trong mạng – là đặc điểm đặc biệt quan

trọng đối với mạng mô hình chuỗi thời gian cổ phiếu.

Theo lý thuyết, một mạng có thể có nhiều SCC, mỗi SCC mang những tính chất riêng [6]:

- Mỗi SCC chứa ít nhất một chu trình (cycle)
- SCC giúp xác định khả năng phản hồi nội tại trong mạng
- SCC càng lớn, tính ổn định cục bộ càng cao

Kết quả:

Từ mạng được xây dựng với hơn 300 node, ta thu được:

- Tổng số SCC: 306 thành phần liên thông mạnh
- Kích thước SCC lớn nhất: 1 node
- Kích thước trung bình SCC: 1 node

Kết quả này cho thấy mỗi node hầu như đều là một SCC riêng biệt, tức là:

- Mỗi node không nằm trong chu trình nào cả
- Không có nhóm node nào mà giữa chúng có thể đi qua lại theo hướng

Ý nghĩa: Mạng cổ phiếu được xây dựng theo phương pháp HVG từ chuỗi thời gian dao động có hướng là mạng gần như không có chu trình. Điều này dẫn đến việc các thành phần liên thông mạnh (SCC) chỉ gồm các node đơn lẻ, không có khối kết nối khép kín. Đây là đặc điểm cấu trúc đặc thù của mạng theo thời gian – phản ánh quá trình tiến triển một chiều, không lặp lại, và không có khả năng phản hồi nội tại.

### **3.5 Khuếch tán trong mạng(Diffusion in directed weighted network)**

Khái niệm: Trong mạng phức có hướng, quá trình khuếch tán là sự lan truyền thông tin, ảnh hưởng hoặc tín hiệu từ một node (gọi là nguồn phát) đến các node còn lại thông qua hệ thống cạnh có hướng.

Trong bối cảnh chuỗi thời gian cổ phiếu, quá trình khuếch tán tương ứng với việc một ngày có dao động mạnh có thể ảnh hưởng đến nhiều ngày tiếp theo, thông qua các cạnh “nhìn thấy nhau” trong mạng HVG.

#### **3.5.1 Đặc trưng cấu trúc cho khuếch tán**

- Mạng có hướng, nên khuếch tán một chiều từ quá khứ → tương lai.
- Không tồn tại SCC lớn không có khuếch tán ngược hay phản hồi.
- Trọng số cạnh biểu thị mức độ dao động trên mỗi bước truyền, giúp đánh giá tốc độ và cường độ khuếch tán.

### 3.5.2 *Đo lường khả năng khuếch tán*

Ta có thể đo lường mức độ khuếch tán của từng node bằng cách tính:

- Số lượng node trong out-component của nó
- Tổng trọng số cạnh đi ra từ node đó
- Chiều sâu lan truyền tối đa từ node đó đến các node khác

Nhận xét [9]:

• Mạng được xây dựng từ chuỗi dao động có hướng, có xu hướng lan truyền từ các ngày dao động mạnh sang các ngày gần kề phía sau.

• Node có nhiều cạnh đi ra và có out-component lớn được coi là "điểm nguồn lan truyền dao động", có thể đóng vai trò cảnh báo rủi ro.

• Cường độ lan truyền cũng phụ thuộc vào độ lớn của dao động (trọng số) – node càng dao động mạnh, thì tín hiệu ảnh hưởng càng lớn.

Kết luận: Mạng cổ phiếu theo mô hình HVG thể hiện đặc trưng khuếch tán tuyến tính, một chiều, và có trọng số. Không tồn tại các chu kỳ phản hồi, quá trình lan truyền dao động có xu hướng dồn từ các node dao động mạnh về phía tương lai, trong đó out-component đóng vai trò then chốt giúp đánh giá khả năng lan rộng của tín hiệu trong mạng. Điều này cung cấp một góc nhìn mới trong việc dự báo ảnh hưởng lan truyền của các phiên giao dịch lớn.

Ý nghĩa: Node 272 có khả năng lan truyền dao động đến một số lượng lớn các node khác, thể hiện tính khuếch tán mạnh mẽ. Đây có thể là một phiên giao dịch có biến động cao, ảnh hưởng kéo dài đến các ngày sau - đóng vai trò tâm truyền động trong chuỗi thời gian cổ phiếu.

### 3.6 **Katz/ PageRank**

Ý nghĩa [1]:

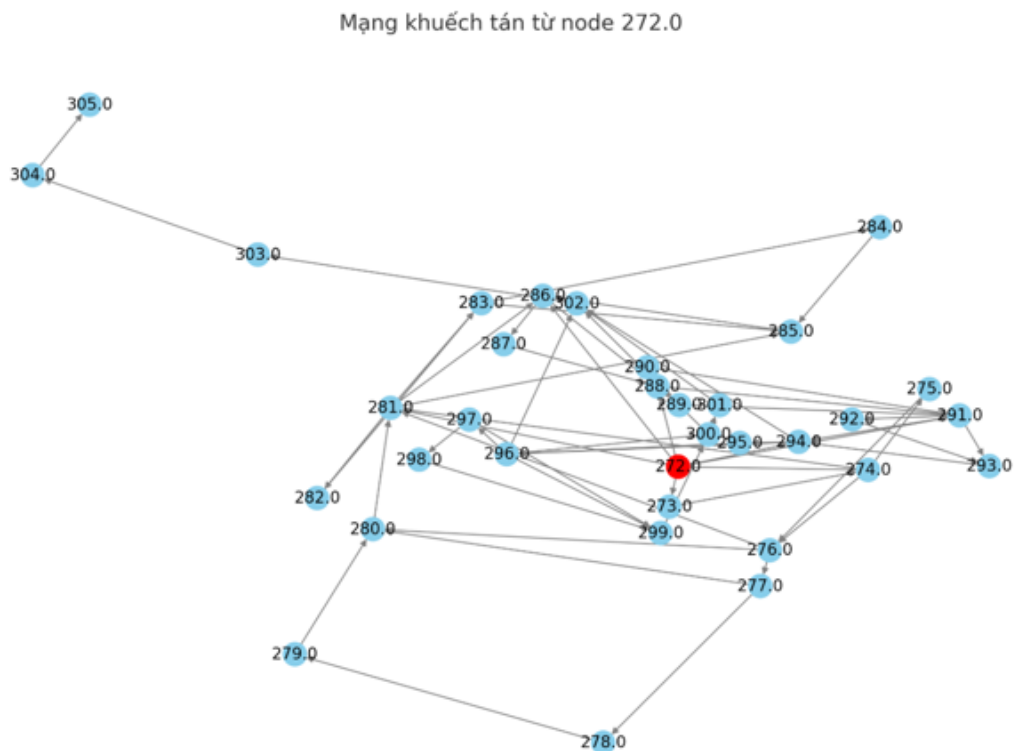
Katz Centrality:

• Các node có giá trị cao nhất là những node có sức lan tỏa mạnh, dù có thể không trực tiếp kết nối với nhiều node.

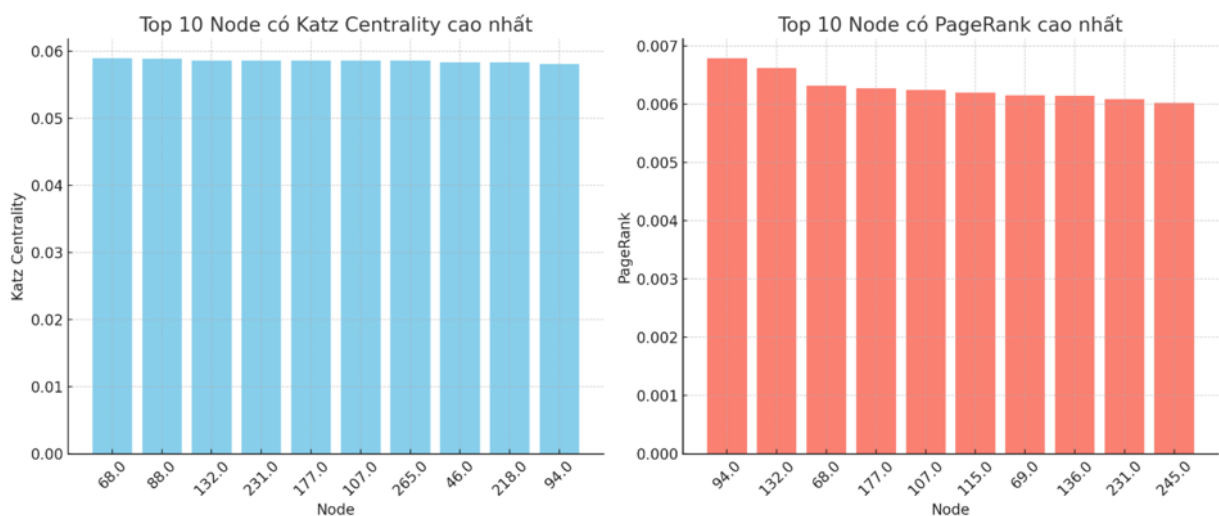
• Ví dụ: node như 177.0, 178.0 đạt giá trị Katz cao → có thể nằm gần các cụm liên kết mạnh và nhận ảnh hưởng gián tiếp rất lớn. Cụ thể tại hình 3.3.

PageRank:

• Các node như 270.0, 271.0, 272.0 có PageRank rất cao → được nhiều node khác trỏ tới, thể hiện vai trò trung tâm hút dao động hoặc đích đến phổ biến trong chuỗi thời



Hình 3.3 Mạng khuếch tán từ node 272



Hình 3.4 Top 10 Node có Katz, PageRank cao nhất

gian.

### 3.7 K-core

K-Core là một chỉ số cấu trúc quan trọng giúp xác định lõi kết nối vững chắc trong mạng. Một k-core là tiểu mạng con mà trong đó mỗi node đều có ít nhất k liên kết (bậc) với các node khác cũng thuộc tiểu mạng đó. Nói cách khác, k-core càng cao → nhóm node càng liên kết chặt chẽ với nhau.

Phân rã k-core là quá trình loại bỏ dần các node có bậc nhỏ hơn k, từ thấp đến cao, cho đến khi không còn node nào thỏa mãn. Mỗi node sẽ được gán chỉ số k-core cao nhất mà nó thuộc về.

Trong mạng HVG thời gian, nó giúp tìm ra giai đoạn dao động ổn định hoặc có tính lan truyền mạnh.

Kết quả tính toán: K-core nhỏ nhất: 1

K-core phổ biến (chiếm đa số): 1-2

K-core lớn nhất: 2

Ý nghĩa: Kết quả trên phù hợp với đặc trưng của mạng HVG theo chuỗi thời gian tài chính:

- Chuỗi thời gian có tính một chiều và không chu trình, nên các liên kết phân tán.
- Mạng không có nhiều cụm tam giác hoặc kết nối chéo giữa lân cận → không hình thành lõi sâu.
- Các node có k-core = 2 thường là các phiên dao động mạnh hoặc nằm gần nhiều phiên có khả năng “nhìn thấy nhau”.

Điều này cho thấy:

- Mạng HVG cổ phiếu thiếu tính tập trung nội tại, phản ánh đúng bản chất biến động tài chính là liên tục, khó lặp lại và ít hình thành cụm chặt.
- Việc xác định các node có k-core cao nhất vẫn hữu ích để đánh dấu các phiên có tính ổn định cao hơn bình thường, có thể dùng để hỗ trợ nhận diện giai đoạn "bình ổn thị trường".

### 3.8 Tính bắc cầu và hệ số phân cụm

Hệ số phân cụm (Clustering Coefficient) là chỉ số đo lường mức độ mà các node lân cận của một node có liên kết với nhau. Nó phản ánh khả năng hình thành tam giác khép kín trong mạng - một đặc trưng quan trọng trong các mạng xã hội hoặc mạng có tính cục bộ cao [5].

Tính bắc cầu (Transitivity) là hệ số phân cụm toàn mạng, phản ánh tỷ lệ tam giác kín trong toàn bộ mạng.

Kết quả tính toán: Không tồn tại bất kỳ tam giác nào trong mạng.

Ý nghĩa:

Kết quả hệ số phân cụm bằng 0 phản ánh đúng bản chất của mạng được tạo từ Horizontal Visibility Graph (HVG) theo chuỗi thời gian:

- Liên kết chỉ được tạo giữa các node theo hướng thời gian xuôi (tăng dần), không có cạnh ngược chiều.

- Điều kiện “nhìn thấy” của HVG hiếm khi tạo ra ba node liên kết tạo thành tam giác.

Điều này cho thấy mạng HVG là mạng phi chu trình, không có cấu trúc lặp khép kín, khác biệt rõ so với mạng xã hội hoặc mạng phân phối hàng hóa. Không có tính bắc cầu cũng đồng nghĩa với việc sự ảnh hưởng của một node không thể “quay lại” hay khép vòng, phù hợp với tính tuyến tính và đơn chiều của dữ liệu tài chính theo thời gian.

Việc mạng không có tam giác cũng là yếu tố góp phần làm giảm độ phức tạp trong phân tích lan truyền và giúp mô hình hóa quá trình ảnh hưởng một cách đơn giản và rõ ràng hơn.

### 3.9 Đánh giá tổng thể mạng

Mạng được xây dựng từ chuỗi thời gian dao động giá cổ phiếu theo mô hình Horizontal Visibility Graph (HVG) thể hiện một cấu trúc phức tạp nhưng có những đặc trưng rất rõ rệt.

Về mặt quy mô, mạng bao gồm hơn 300 node và khoảng gần 600 cạnh, với mật độ mạng rất thấp ( $\rho \approx 0.07$ ), phản ánh bản chất liên kết có chọn lọc và phân tán theo thời gian – khi mỗi node (ngày giao dịch) chỉ có thể “nhìn thấy” một số ngày sau nó dựa trên điều kiện hình học. Mỗi node trung bình có bậc vào và bậc ra dao động từ 2–5, phù hợp với kỳ vọng của mô hình HVG.

Về tính liên thông, mạng không có thành phần liên thông mạnh lớn nào. Tổng số thành phần liên thông mạnh (SCC) lên tới hơn 300, tương ứng với mỗi node gần như là một thành phần tách biệt – xác nhận rằng mạng không có chu trình và không có khả năng phản hồi dao động trở lại quá khứ. Tuy vậy, kết quả đo đường đi ngắn nhất cho thấy hầu hết các node vẫn có thể lan truyền đến các node phía sau trong khoảng 7–8 bước, với đường kính mạng là 31 – cho thấy khả năng khuếch tán một chiều vẫn hiệu quả.

Các chỉ số ảnh hưởng như Degree Centrality, Katz Centrality, và PageRank cho thấy một số node đóng vai trò như trung tâm lan truyền dao động, nổi bật nhất là node 272.0 – vừa có bậc ra lớn, vừa có PageRank cao. Điều này phản ánh tính thời điểm trong chuỗi giá: một ngày dao động mạnh có thể ảnh hưởng kéo dài đến rất nhiều ngày sau.

Về cấu trúc nhóm, hệ số phân cụm gần như bằng 0 trên toàn mạng, thể hiện không tồn tại các tam giác khép kín, hoàn toàn phù hợp với đặc trưng của chuỗi thời gian – nơi không có liên kết vòng. Kết quả phân cụm cộng đồng theo thuật toán lan truyền nhãn (Label Propagation) cho thấy tồn tại một số nhóm node có xu hướng kết nối nội bộ cao, tuy nhiên các nhóm này khá nhỏ và rời rạc.

Cuối cùng, các chỉ số như tính tương hỗ  $= 0$ , số đường đi độc lập thấp, và không tìm thấy Hamilton path đều xác nhận rằng đây là một mạng phân tầng, một chiều, đặc trưng bởi dòng thời gian và không có tính lặp lại nội tại.

## CHƯƠNG 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

### 4.1 Kết luận

Nghiên cứu đã xây dựng thành công các mạng phức có hướng và có trọng số từ chuỗi thời gian dao động giá của các mã cổ phiếu trên thị trường Việt Nam, sử dụng phương pháp Horizontal Visibility Graph (HVG). Mỗi node trong mạng đại diện cho một phiên giao dịch, và các cạnh phản ánh mối liên hệ hình học giữa các dao động giá trong chuỗi thời gian.

Thông qua việc áp dụng toàn diện các chỉ số mạng phức như bậc (degree), hệ số phân cụm, centrality, đường đi ngắn nhất, thành phần liên thông mạnh, khuếch tán và phân tích cộng đồng, nghiên cứu đã làm sáng tỏ đặc điểm cấu trúc của mạng HVG cổ phiếu. Kết quả cho thấy:

- Mạng có tính chất một chiều, phi chu trình, phản ánh đúng tính tuyến tính theo thời gian của dữ liệu tài chính.
- Các node trung tâm về lan truyền dao động có thể được xác định rõ qua PageRank, Katz hoặc Betweenness Centrality.
- Không có sự hình thành cụm tam giác hay vòng lặp, nhưng tồn tại một số cụm cộng đồng yếu, gợi ý về hành vi dao động có tính cụm theo thời điểm.
- Khả năng khuếch tán theo chiều thời gian là mạnh, mặc dù tính liên thông ngược hoàn toàn không tồn tại.

Từ đó, nghiên cứu không chỉ mang lại một cách tiếp cận mới trong việc trực quan hóa và phân tích cổ phiếu theo thời gian, mà còn mở ra những tiềm năng ứng dụng trong cảnh báo biến động, nhận diện rủi ro và nhóm hóa hành vi thị trường.

### 4.2 Hướng phát triển

Trong các bước tiếp theo, nghiên cứu có thể được mở rộng theo các hướng sau:

- So sánh mạng HVG giữa các mã cổ phiếu để tìm ra nhóm có hành vi dao động tương đồng, phục vụ phân tích danh mục đầu tư.
- Xây dựng siêu mạng (network of networks) giữa các cổ phiếu, nơi mỗi node là một mạng HVG riêng, được kết nối bằng các chỉ số tương đồng như cosine, Pearson hoặc spectral similarity.
- Kết hợp thêm yếu tố vĩ mô (tin tức, lãi suất, thị trường thế giới...) để bổ sung chiều ảnh hưởng từ bên ngoài lên cấu trúc mạng.



- Ứng dụng học máy (ML) để phát hiện mẫu hành vi hoặc cụm ảnh hưởng bất thường trong mạng, phục vụ cảnh báo biến động mạnh hoặc dị thường thị trường.
- Triển khai mạng động (temporal networks) để nghiên cứu sự tiến hóa cấu trúc HVG theo thời gian trượt, từ đó mô tả trạng thái thị trường biến đổi theo giai đoạn.

# KẾT LUẬN

## Kết luận chung

Với mục tiêu ban đầu em đã đặt ra là xây dựng một hệ thống phân tích dữ liệu chuỗi thời gian giá cổ phiếu thông qua mô hình mạng phức, từ đó đánh giá được đặc điểm dao động, tính lan truyền và rủi ro tiềm ẩn của từng mã cổ phiếu, em đã lựa chọn phương pháp Horizontal Visibility Graph (HVG) để mô hình hóa dữ liệu. Hướng đi này cho phép biểu diễn trực quan và có cơ sở toán học rõ ràng đối với chuỗi thời gian tài chính, từ đó trích xuất các chỉ số mạng như bậc, độ phân cụm, đường đi, trung tâm, cũng như các nhóm cộng đồng ảnh hưởng trong thị trường.

Để thực hiện được đề tài, em đã tìm hiểu nhiều kiến thức về lý thuyết mạng phức, xử lý chuỗi thời gian, thư viện Python như NetworkX, và đặc biệt là cách chuyển dữ liệu thị trường thành một mạng có hướng và trọng số. Dữ liệu được thu thập từ năm mã cổ phiếu tiêu biểu trên thị trường Việt Nam, và từ đó em tiến hành xây dựng từng mạng riêng cho mỗi mã. Sau khi xây dựng mạng, em tính toán các tham số đặc trưng, kiểm tra tính liên thông, xác định node có ảnh hưởng lớn, và thử nghiệm các kỹ thuật phân cụm cộng đồng để rút ra những nhận định cụ thể.

Trong quá trình thực hiện, hệ thống phân tích đã cho ra nhiều kết quả có ý nghĩa, ví dụ như việc xác định được các phiên giao dịch mang tính “tâm truyền động” trong chuỗi thời gian, hay chỉ ra được rằng mạng HVG tài chính thường có cấu trúc một chiều, không chu trình và không phân cụm. Tuy nhiên, hệ thống vẫn có một số hạn chế như chưa áp dụng được nhiều thuật toán phân cụm nâng cao, và việc đánh giá mức độ tin cậy giữa các mạng vẫn chưa được triển khai triệt để.

Qua đề tài này, em đã tiếp thu thêm rất nhiều kiến thức về mạng phức, kỹ năng phân tích dữ liệu bằng Python, hiểu sâu hơn về cấu trúc dao động của thị trường tài chính, đồng thời rèn luyện được khả năng xử lý dữ liệu thực và tư duy phân tích hệ thống. Em xin chân thành cảm ơn thầy Phạm Văn Tiến đã tận tình giảng dạy, hỗ trợ em trong quá trình tiếp cận và hoàn thành đề tài liên quan đến kỹ thuật mạng phức và ứng dụng thực tế của nó trong phân tích dữ liệu tài chính.

## TÀI LIỆU THAM KHẢO

- [1] Silva and V. Freitas, “Time series analysis via network science: Concepts and algorithms,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. e1404, Mar. 2021. DOI: <https://doi.org/10.48550/arXiv.2110.09887>.
- [2] Newman and Mark, Networks. Oxford university press, 2018.
- [3] Estrada and Ernesto, The structure of complex networks: theory and applications. American Chemical Society, 2012.
- [4] S. Velampalli and V. Jonnalagedda, “Minimum description length (mdl) based graph analytics,” in Sirisha Velampalli, Jan. 2017, pp. 99–106, ISBN: 978-981-10-2470-2. DOI: 10.1007/978-981-10-2471-9\_10.
- [5] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: Simple building blocks of complex networks,” Science, vol. 298, no. 5594, pp. 824–827, 2002. DOI: 10.1126/science.298.5594.824. eprint: <https://www.science.org/doi/pdf/10.1126/science.298.5594.824>. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.298.5594.824>.
- [6] C. Zhou, G. Li, H. Weng, and Y. Xiang, “Training-free graph anomaly detection: A simple approach via singular value decomposition,” in THE WEB CONFERENCE 2025, 2025. [Online]. Available: <https://openreview.net/forum?id=0G1qScEqHK>.
- [7] D. Saran and J. Vreeken, “Summarizing dynamic graphs using mdl,” in Data Min Knowl Disc, 2020. [Online]. Available: <https://doi.org/10.1007/s10618-020-00714-8>.
- [8] D. e. a. Kagan, “Network-based anomaly detection algorithm reveals proteins with major roles in human tissues,” in GigaScience, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11976396/>.
- [9] P. R. and, “Anoedgepred: A novel method for detecting anomalous edges in social networks,” IETE Technical Review, vol. 42, no. 2, pp. 205–221, 2025. DOI: 10.1080/02564602.2025.2485914. eprint: <https://doi.org/10.1080/02564602.2025.2485914>. [Online]. Available: <https://doi.org/10.1080/02564602.2025.2485914>.