

MẠNG XÃ HỘI

Bài 5. PHÂN TÍCH MẠNG XÃ HỘI

ThS. Lê Nhật Tùng

- 1 5.1. Tổng quan
- 2 5.2 / 5.3: Các phương pháp phân tích và công cụ
- 3 5.4. Xây dựng mạng xã hội
- 4 5.5 Chiến lược Khai thác Dữ liệu trong Mạng xã hội

- 1 5.1. Tổng quan
- 2 5.2 / 5.3: Các phương pháp phân tích và công cụ
- 3 5.4. Xây dựng mạng xã hội
- 4 5.5 Chiến lược Khai thác Dữ liệu trong Mạng xã hội

Thành phần cơ bản của Mạng xã hội

- **Nút (nodes):**

- Đại diện cho người dùng trong mạng
- Mỗi nút có các thuộc tính riêng

- **Cạnh (edges):**

- Biểu thị các tương tác giữa các người dùng
- Biểu thị mối quan hệ giữa các người dùng

Ứng dụng chính của Mạng xã hội

- **Trong lĩnh vực chăm sóc sức khỏe:**
 - Giúp kiểm soát hoạt động của nhà cung cấp
 - Đo lường hiệu quả
- **Marketing:**
 - Hỗ trợ quảng bá thương hiệu
 - Thu hút khách hàng
- **Phân tích dịch bệnh:**
 - Theo dõi sự lây lan
 - Mô hình hóa dịch bệnh

Phương pháp phân tích Mạng xã hội

- Dựa trên lý thuyết đồ thị
- Sử dụng các phương pháp thống kê
- Áp dụng các nguyên lý xã hội học

- 1 5.1. Tổng quan
- 2 5.2 / 5.3: Các phương pháp phân tích và công cụ
- 3 5.4. Xây dựng mạng xã hội
- 4 5.5 Chiến lược Khai thác Dữ liệu trong Mạng xã hội

5.2.1 Phân tích cấu trúc mạng

- **Độ tập trung (Centrality)**

- Xác định các nút quan trọng trong mạng
- Đo lường mức độ ảnh hưởng của từng nút

- **Phân tích cụm (Clustering)**

- Phát hiện các nhóm có mối liên kết chặt chẽ
- Xác định cộng đồng trong mạng

5.2.2 Các chỉ số đo lường

- **Độ trung tâm của nút**

- Degree centrality: Số lượng kết nối trực tiếp
- Betweenness centrality: Vai trò trung gian
- Closeness centrality: Khoảng cách đến các nút khác

- **Mật độ mạng**

- Tỷ lệ kết nối thực tế trên tổng số kết nối có thể
- Đo lường mức độ kết nối của mạng

5.2.3 Kỹ thuật phân tích nâng cao

- **Phân tích thời gian thực**

- Theo dõi sự thay đổi của mạng
- Phát hiện xu hướng và mẫu hình

- **Dự đoán liên kết**

- Dự báo kết nối mới có thể xuất hiện
- Phân tích mối quan hệ tiềm năng

5.2.4 Công cụ phân tích

- **Phần mềm chuyên dụng**

- Gephi
- NodeXL
- UCINET

- **Thư viện lập trình**

- NetworkX (Python)
- igraph (R)
- SNAP (C++)

- **Giới thiệu:**

- Phần mềm mã nguồn mở cho phân tích và trực quan hóa mạng
- Website: gephi.org

- **Cài đặt:**

- Tải trực tiếp từ website chính thức
- Yêu cầu Java Runtime Environment (JRE)
- Hỗ trợ Windows, Mac OS X, Linux

- **Tính năng chính:**

- Trực quan hóa mạng lưới động
- Phân tích cấu trúc cộng đồng
- Tính toán độ đo trung tâm
- Xuất báo cáo và hình ảnh chất lượng cao

- **Giới thiệu:**

- Thư viện Python cho phân tích mạng phức tạp
- Website: networkx.org

- **Cài đặt:**

- Lệnh cài đặt: `pip install networkx`
- Yêu cầu: Python 3.8+, numpy, scipy, pandas

- **Code mẫu cơ bản:**

- Import thư viện: `import networkx as nx`
- Tạo đồ thị: `G = nx.Graph()`
- Thêm cạnh: `G.add_edges_from([(1,2), (1,3)])`
- Vẽ đồ thị: `nx.draw(G)`

- **Giới thiệu:**

- Thư viện phân tích mạng lưới cho R
- Website: igraph.org/r

- **Cài đặt:**

- Lệnh cài đặt: `install.packages("igraph")`
- Load thư viện: `library(igraph)`

- **Code mẫu cơ bản:**

- Tạo đồ thị: `g <- make_graph("A-B-C-A")`
- Vẽ đồ thị: `plot(g)`
- Tính độ trung tâm: `degree(g)`

- **Giới thiệu:**

- Stanford Network Analysis Platform
- Website: snap.stanford.edu

- **Cài đặt:**

- Clone từ GitHub: `git clone snap-stanford/snap`
- Build: `make all`
- Python interface: `pip install snap-stanford`

- **Tính năng chính:**

- Xử lý mạng quy mô lớn (hàng tỷ cạnh)
- Thuật toán graph mining hiệu quả
- Hỗ trợ machine learning trên đồ thị
- Phân tích temporal networks

5.2.5 Kỹ thuật trực quan hóa

- **Biểu diễn đồ thị**
 - Force-directed layout
 - Circular layout
 - Hierarchical layout
- **Phân tích tương tác**
 - Lọc và tìm kiếm
 - Phóng to/thu nhỏ
 - Tô màu theo thuộc tính

5.2.6 Các bước phân tích thực tế

- **Thu thập dữ liệu**

- API mạng xã hội
- Crawl dữ liệu
- Khảo sát trực tiếp

- **Xử lý và phân tích**

- Làm sạch dữ liệu
- Áp dụng các phương pháp phân tích
- Trực quan hóa kết quả

- 1 5.1. Tổng quan
- 2 5.2 / 5.3: Các phương pháp phân tích và công cụ
- 3 5.4. Xây dựng mạng xã hội
- 4 5.5 Chiến lược Khai thác Dữ liệu trong Mạng xã hội

5.4.3 Quy trình xây dựng mạng xã hội

- **Xây dựng từ tập dữ liệu**

- Ví dụ: Phân tích mối quan hệ bạn bè trên Facebook
 - Dữ liệu Excel: UserID1, UserID2, RelationType
 - Chuyển sang Pandas DataFrame
 - Tạo đồ thị NetworkX với nodes=users, edges=friendships

- **Thu thập dữ liệu**

- Ví dụ: Thu thập tweet về COVID-19
 - Sử dụng Twitter API với hashtag #COVID19
 - Lưu thông tin: user, content, timestamp, interactions
 - Xây dựng mạng lưới tương tác giữa users

Các bước xử lý dữ liệu

- **Tiền xử lý dữ liệu**

- Ví dụ: Xử lý bình luận Facebook
 - Loại bỏ emoji, ký tự đặc biệt
 - Chuẩn hóa text (lowercase, remove stopwords)
 - Xử lý missing values và duplicates

- **Phân loại và phân tích**

- Ví dụ: Phân tích sentiment COVID-19
 - Train Naive Bayes với data có nhãn (positive/negative)
 - Phân loại sentiment các tweet mới
 - Visualization: biểu đồ sentiment theo thời gian

- **Khối lượng và Danh tiếng**

- Ví dụ: Chiến dịch marketing trên Facebook
 - Số lượng post: 100 posts/tháng
 - Reach: 50,000 người xem
 - Engagement rate: 5% (like, share, comment)

- **Người dùng chuyên dụng và Sự thống trị**

- Ví dụ: Influencer Marketing
 - Top influencers: >100K followers
 - Chuyên gia ngành: bác sĩ, chuyên gia y tế
 - Engagement rate: >3% mỗi post

Ví dụ thực tế: Phân tích mạng xã hội trong dịch COVID-19

- **Thu thập dữ liệu**

- Twitter API: hashtags #COVID19, #vaccine
- Facebook API: public posts từ pages y tế
- Reddit API: posts từ r/Coronavirus

- **Phân tích**

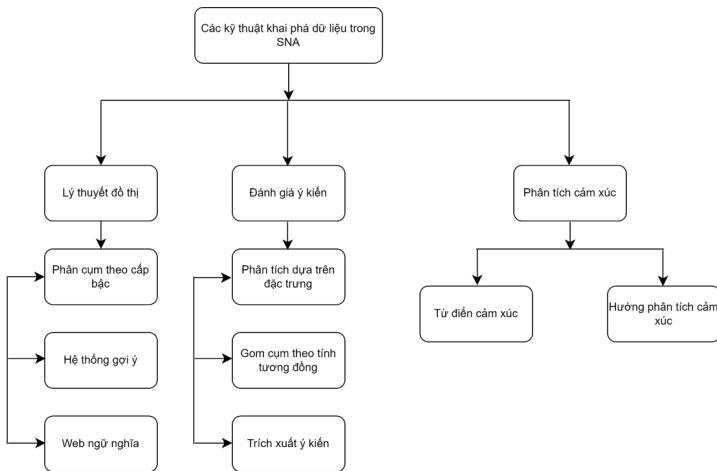
- Sentiment analysis về vaccine
- Network analysis: spread of information
- Topic modeling: main concerns

- **Kết quả**

- Trend sentiment theo thời gian
- Key influencers trong mạng lưới
- Hot topics và patterns

- 1 5.1. Tổng quan
- 2 5.2 / 5.3: Các phương pháp phân tích và công cụ
- 3 5.4. Xây dựng mạng xã hội
- 4 5.5 Chiến lược Khai thác Dữ liệu trong Mạng xã hội

Sơ đồ tổng quan các phương pháp



Hình 5.8. Sơ đồ biểu diễn các phương pháp hiện nay

Hình: Sơ đồ biểu diễn các phương pháp hiện nay

5.5.1. Lý thuyết đồ thị

- **Mục đích:** Nghiên cứu và phân tích cấu trúc mạng xã hội
- **Bao gồm:**
 - Phân cụm theo cấp bậc: Tổ chức nhóm người dùng theo thứ bậc
 - Hệ thống gợi ý: Đề xuất dựa trên mối quan hệ và tương tác
 - Web ngữ nghĩa: Kết nối và chia sẻ thông tin giữa các nút mạng
- **Ứng dụng:** Phân tích cộng đồng, xác định người ảnh hưởng

5.5.2. Đánh giá ý kiến

- **Mục đích:** Khai thác và phân tích ý kiến người dùng
- **Phương pháp:**
 - Phân tích đặc trưng: Tập trung vào các khía cạnh được đề cập nhiều
 - Gom cụm tương đồng: Nhóm các ý kiến có tính chất giống nhau
 - Trích xuất ý kiến: Lọc và phân loại các đánh giá có giá trị
- **Ứng dụng:** Đánh giá sản phẩm, phân tích phản hồi khách hàng

5.5.3. Phân tích cảm xúc

- **Mục đích:** Phân tích thái độ và cảm xúc trong nội dung
- **Thành phần:**
 - Từ điển cảm xúc: Danh sách từ và cụm từ biểu đạt cảm xúc
 - Hướng phân tích: Xác định cảm xúc tích cực/tiêu cực
- **Ứng dụng:**
 - Phân tích phản hồi khách hàng
 - Đánh giá hiệu quả marketing
 - Theo dõi danh tiếng thương hiệu

Chúc các bạn học thật tốt!