



PHÂN TÍCH VÀ KHAI PHÁ MẠNG XÃ HỘI

Trường hè Khai phá dữ liệu 2016
<http://fit.uet.vnu.edu.vn/dmss2016/>

PGS.TS. Hà Quang Thụy và ThS. TRẦN MAI VŨ
PHÒNG THÍ NGHIỆM KHOA HỌC DỮ LIỆU VÀ CÔNG NGHỆ TRI THỨC
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ
ĐẠI HỌC QUỐC GIA HÀ NỘI
16-19/08/2016

Nội dung

1. Tại sao phân tích và khai phá mạng xã hội ?
2. Phân tích và khai phá mạng xã hội là gì ?
3. Phân tích cấu trúc mạng xã hội
4. Thu thập và tiền xử lý dữ liệu mạng xã hội trực tuyến
5. Phân tích nội dung mạng xã hội trực tuyến
6. Phân tích cấu trúc mạng xã hội trực tuyến

Tại sao phân tích và khai phá mạng xã hội?

- Mạng xã hội là phổ biến
- Mạng xã hội là kho tài nguyên tiềm năng đồ sộ
- Phân tích và khai phá mạng xã hội: tính thời sự

Mạng xã hội là phổ biến

- **Mạng xã hội xuất hiện trong nhiều lĩnh vực**
 - Xã hội học, CNTT (khai phá dữ liệu), khoa học hành vi, toán học, thống kê và nhiều lĩnh vực khác
- **Mạng xã hội hiện - ẩn**
 - Mạng xã hội hiện: Quan hệ giữa các nút là rõ. Ví dụ: Facebook, Twitter, và MySpace.
 - Mạng xã hội ẩn: Quan hệ giữa các nút: kết quả qua phân tích. Ví dụ: Các nhân viên thực hiện quy trình, các khách hàng cùng sở thích, v.v.
- **Mạng xã hội tĩnh - động**
 - Tĩnh: Kết nối cố định như các bộ định tuyến trong mạng, v.v.
 - Động: Kết nối thay đổi như con người, động vật, côn trùng, gen, protein, v.v.
- **Mạng xã hội ngoại tuyến – trực tuyến**
 - Ngoại tuyến: Mạng XH thiết lập từ dữ liệu của công ty
 - Trực tuyến: Mạng XH trực tuyến trên Internet

[Alhajj14] Reda Alhajj, Jon Rokne. *Encyclopedia of Social Network Analysis and Mining*. Springer-New York, 2014

Mạng XH với Big Data và ứng dụng

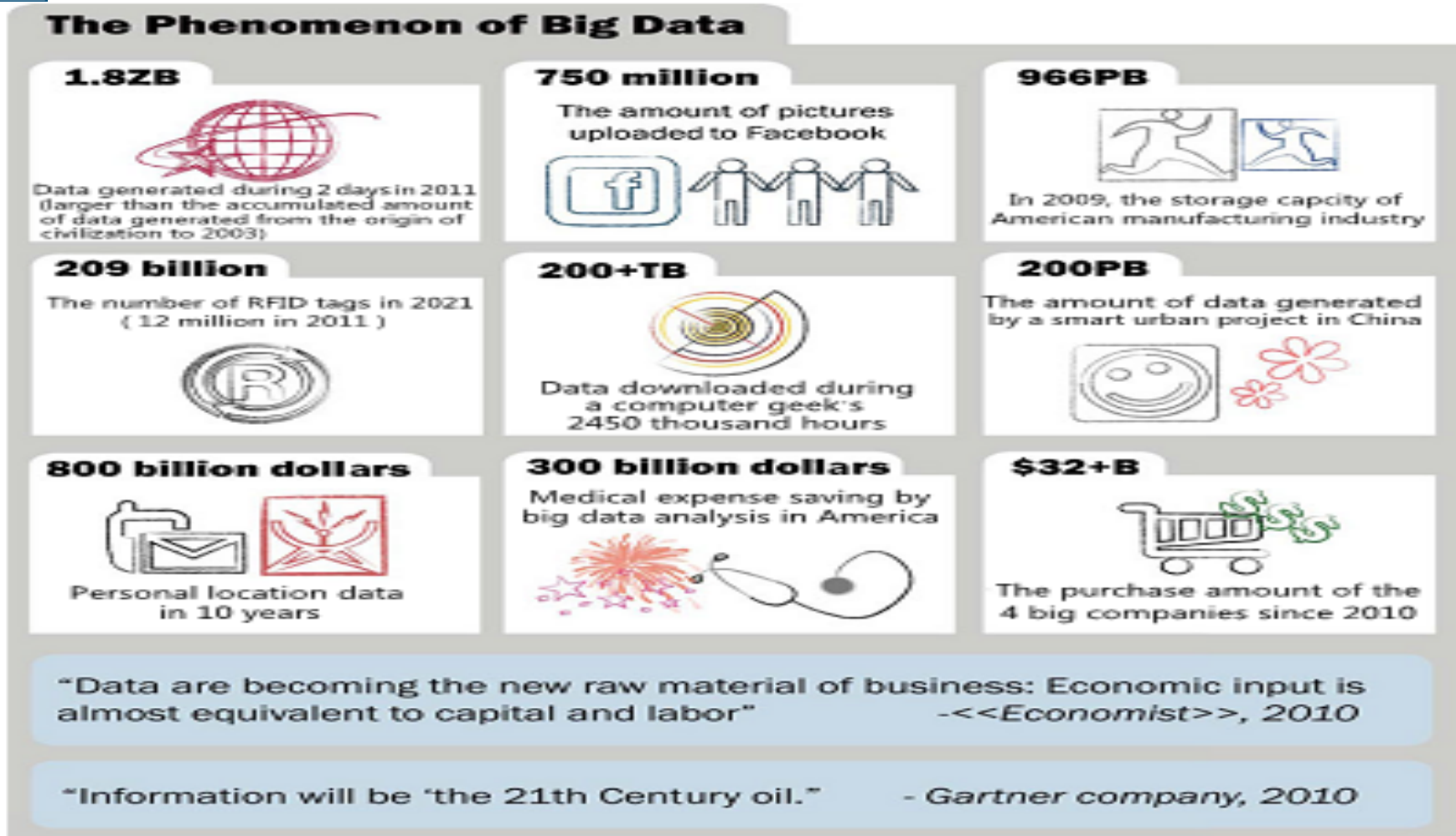
● Mạng xã hội với Big Data

- Mạng xã hội trực tuyến (Facebook, Twitter, Google+, mạng chuyên gia LinkedIn, Youtube, v.v.), mạng xã hội công ty → Big Data.
- Big Data: Volume, Variety, Velocity, Value (“*dầu mỏ của Thế kỷ 21*”)

● Một số ứng dụng phân tích mạng xã hội

- Khai phá dữ liệu, truy hồi thông tin (*information retrieval*), hệ tư vấn (*recommender systems*), khoa học web (*Web science*), nhiều ngành khoa học xã hội (đặc biệt trong **xã hội học: sociology**).
- Quản lý quan hệ KH xã hội (Social CRM, khách hàng vận động: *Advocate*), tư vấn xã hội (*social recommendation*), khai phá quan điểm (*opinion mining*), quản lý danh tiếng (*reputation management*), phóng viên công dân thời gian thực (Real-time Citizen journalism), phản ứng công dân (*Citizen response*), phân tích hành vi con người (*human behavior analysis*), v.v.
- Khoa học dữ liệu (Data Science): nghề hấp dẫn nhất thế kỷ 21

Big data không ngừng gia tăng và giá trị



- (i) Mỹ: tiết kiệm 300 tỷ US\$ ngành y tế, (ii) Châu Âu: chính phủ tiết kiệm 100 tỷ Euro (giảm gian lận, sai sót, chênh lệch thuế), v.v.

[Chen14] Min Chen, Shiwen Mao, Yunhao Liu. *Big Data: A Survey*. MONET 19(2): 171-209, 2014.

Phân tích và khai phá MXH: tính thời sự

Google Scholar

<i>Cụm từ tìm kiếm</i>	<i>2012</i>	<i>2013</i>	<i>2014</i>	<i>2015</i>	<i>2016</i>
"social network analysis"	12600	14100	15200	15900	8880
“social network mining”	290	300	370	370	200

Springer

<i>Cụm từ tìm kiếm</i>	<i>2012</i>	<i>2013</i>	<i>2014</i>	<i>2015</i>	<i>2016</i>
"social network analysis"	11100	11130	11170	11180	11120
“social network mining”	350	360	360	370	350

ACM Digital Library

<i>Cụm từ tìm kiếm</i>	<i>Tổng số</i>
"social network analysis"	1260
“social network mining”	157350

- Thống kê đơn giản từ ba nguồn lưu tài liệu (10/8/2016)

Phân tích và khai phá mạng xã hội là gì?

- Khái niệm mạng xã hội, mạng xã hội trực tuyến
- Các đặc trưng cơ bản của mạng xã hội
- Phân tích nội dung và phân tích cấu trúc mạng xã hội

Khái niệm mạng xã hội

● Khái niệm

- Định nghĩa phổ biến: Mạng tương tác/quan hệ xã hội: nút là tác nhân xã hội và cạnh là quan hệ/tương tác giữa các tác nhân đó. “*là một cấu trúc xã hội bao gồm các cá nhân hay tổ chức, thường được biểu diễn bằng các nút, cùng với các quan hệ xã hội, tương ứng với các liên kết giữa các nút*”.
- Định nghĩa tổng quát. Mạng thông tin: nút là tác nhân/thực thể có yếu tố xã hội và cạnh là quan hệ giữa các nút.
- MXH trực tuyến (online social network): MXH được thi hành bằng dịch vụ mạng xã hội trực tuyến (online social network service)

● Phương tiện xã hội

- Phương tiện xã hội (*social media*) là phương tiện được thiết kế để truyền bá ý tưởng thông qua tương tác xã hội [[Leskovec11](#)], là “phương tiện tương tác của con người để tạo, chia sẻ và trao đổi thông tin và ý tưởng trong cộng đồng ảo và mạng một cách trực tuyến” [[Aggarwal14](#)].

Tiến hóa nghiên cứu mạng xã hội

● Trước Internet

- 1930's: Xã hội học, tập trung vào cộng đồng (gắn gũi về địa lý) và tương tác con người.
- J. Barnes [Barnes54] thúc đẩy “mạng xã hội” vào NC khoa học
- 1970's: mở rộng tới khái niệm cộng đồng tổng quát (con người, động vật, các thực thể tương tác). Các đặc trưng điển hình.

● Thời đại Internet

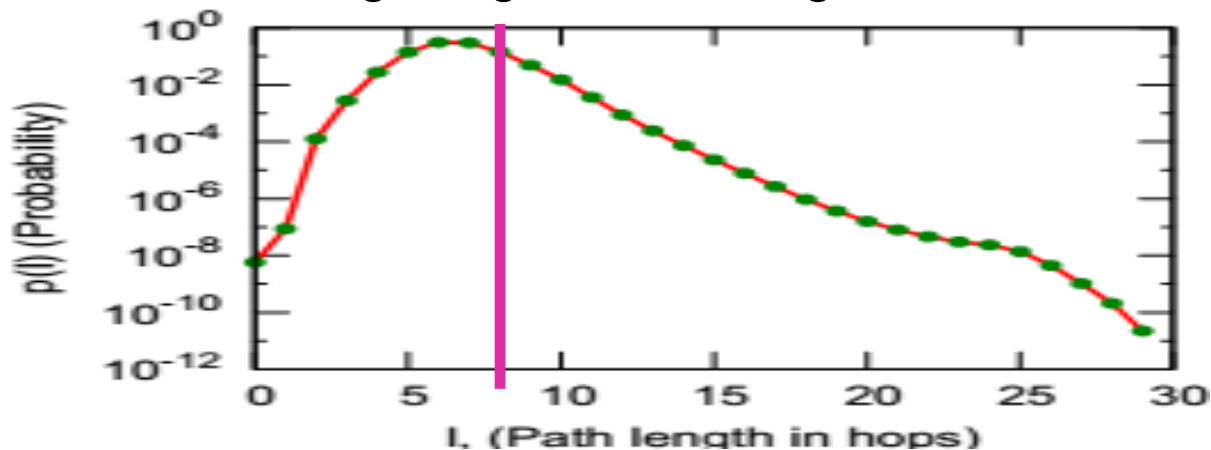
- Bùng nổ: cộng đồng dựa trên web toàn cầu (phương tiện xã hội)
- Kích thước lớn và biến đổi đòi hỏi kỹ thuật phân tích tự động hóa
- Ứng dụng rộng rãi: xã hội học (*Phân tích mô tả người sử dụng, Động lực sử dụng Facebook, Trình bày danh tính, Vai trò của Facebook trong tương tác xã hội, Tính riêng tư và việc tiết lộ thông tin*), Kinh tế (Social CRM, Tiếp thị), v.v.
- Nội dung xã hội học mới: "Netnography" = "Inter[net]" and "eth[nography]", "Child-led Research" v.v.

[Barnes54] J. Barnes. *Class and Committees in a Norwegian Island Parish*. *Human Relations*, 7, 39-58, 1954.

Đặc trưng thể giới nhỏ

● Khái niệm

- Small-world: Hầu hết cặp hai người trên thế giới kết nối nhau bằng một chuỗi ngắn thường là sáu kết nối trực tiếp.
- [Travers69] Jeffrey Travers, Stanley Milgram (1969). An Experimental Study of the Small World Problem, *Sociometry*, **32**(4): 425-443, Dec., 1969. Thủ công: cho kết quả 6.6.
- Khi mạng phát triển bổ sung nút: thêm cung kết nối
- Kiểm thử đúng hầu hết các mạng nhỏ (một vài hạn chế)
- Mạng XH lớn: Microsoft Instant Messenger 240 triệu người, 4.5TB. Phân bố đường đi ngắn nhất trung bình là 6.6, 90% không vượt quá 7.8



[Lescovec08] Jurij Lescovec. *Dynamics of large networks*. PhD Thesis, Carnegie Mellon University, 2008.

Phân bố luật lũy thừa

● Khái niệm

- Power-law degree distributions / the scale – free property
- Phân bố luật lũy thừa: số nút (cung) có k liên kết tới bằng khoảng $1/k^2$ với số $k > 2$ đa phần.

	Số mũ α (bậc vào/ra)
film actors	2.3
telephone call graph	2.1
email networks	1.5/2.0
sexual contacts	3.2
WWW	2.3/2.7
internet	2.5
peer-to-peer	2.1
metabolic network	2.2
protein interactions	2.4

● Phân bố luật lũy thừa ở hầu hết mạng XH

- Jurij Lescovec xác nhận Microsoft Instant Messenger [Lescovec08]

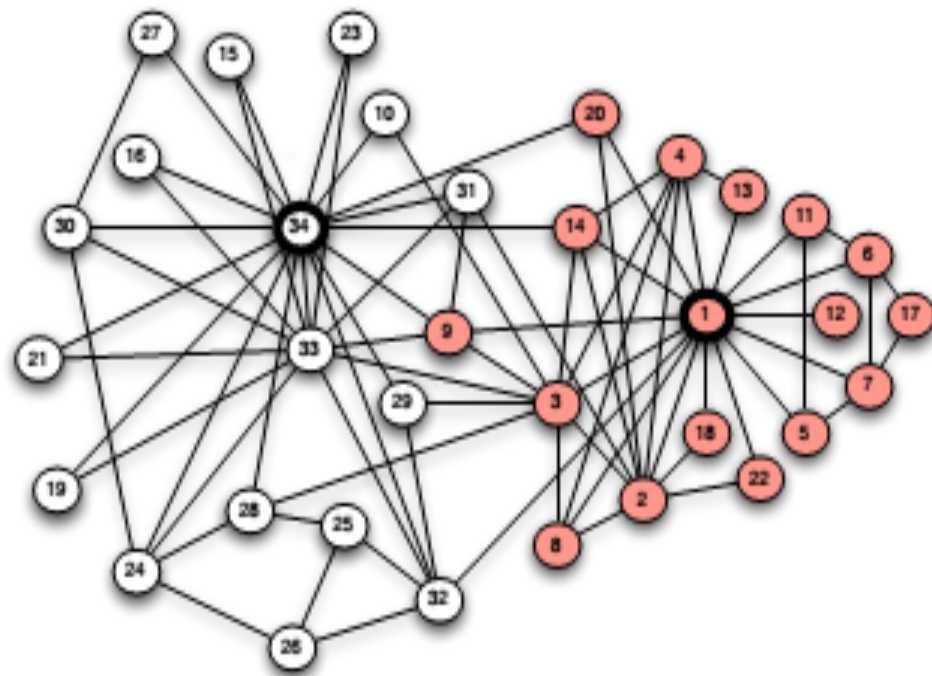
Đặc trưng tập nhân

● Khái niệm

- Network transitivity
- Cấu trúc và vận động của mạng chịu tác động bởi các nút có số lượng lớn các cung kết nối
- Vai trò của các nút trong mạng xã hội

● Ví dụ thể hiện

- Đồ thị câu lạc bộ karate của Zachary quan sát trong 3 năm.
- Đồ thị gồm 34 nút thành viên của câu lạc bộ.
- Hai tập nút: một tập quanh các nút 34 (**34 là chủ tịch**), tập còn lại quanh nút 1 (**huấn luyện viên**) thể hiện tranh chấp hai nhóm quanh **hai nút nhân**



[Easley10] David Easley and Jon Kleinberg (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, 2010.

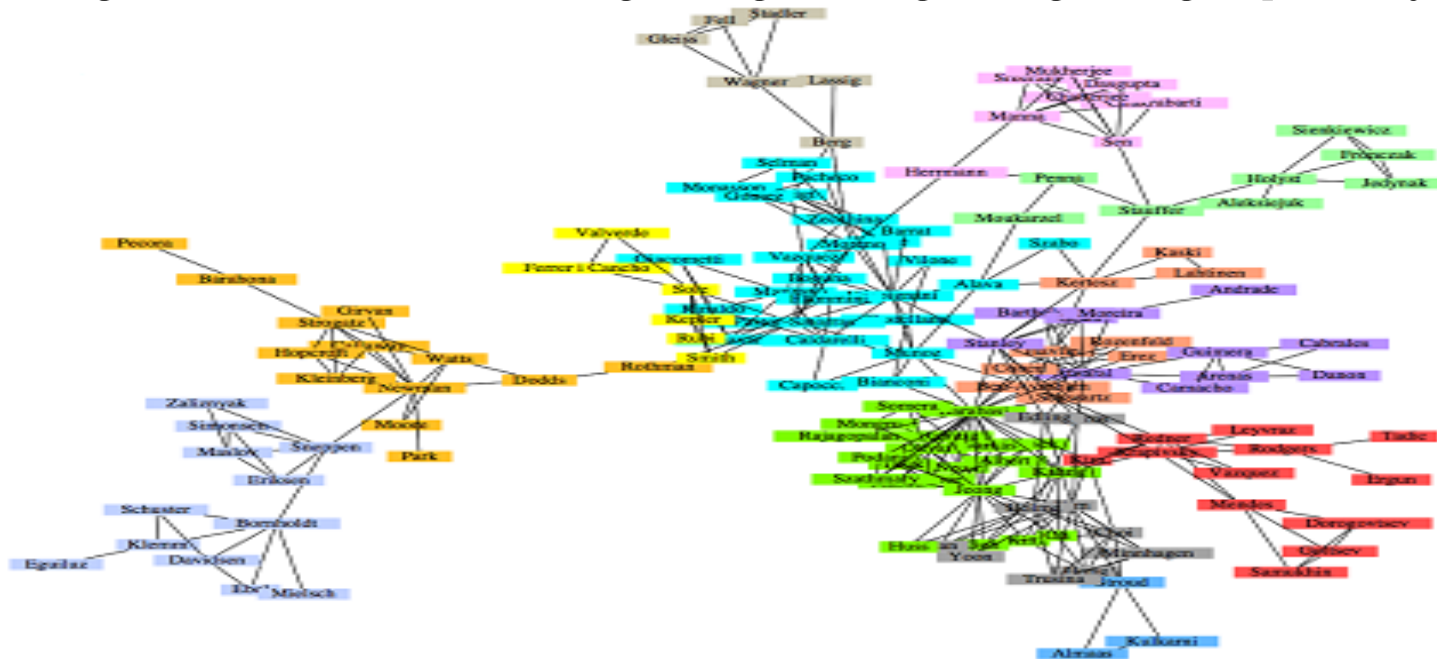
Đặc trưng cấu trúc cộng đồng

● Khái niệm

- Community structure
- Mạng được chia thành các cộng đồng, các nút trong cùng một cộng đồng liên kết chặt chẽ còn các nút khác cộng đồng liên kết yếu
- Một cộng đồng trong mạng xã hội như là “nhóm cùng sở thích” trong thế giới thực. Liên hệ “nhóm lợi ích”, “Advocacy group” ?

● Thể hiện

- Mạng CLB Karate: hai cộng đồng, Mạng đồng tác giả [Easley10]



Phân tích cấu trúc mạng xã hội

- Một số bài toán điển hình
- Giới thiệu công cụ phân tích mạng xã hội SNAP

Một số bài toán phân tích mạng xã hội

- Các kiểu phân tích
 - Phân tích dựa trên liên kết và cấu trúc
 - Phân tích dựa trên nội dung
 - Phân tích kết hợp
- Phân tích động và phân tích tĩnh
 - Phân tích tĩnh \leftrightarrow mạng xã hội tĩnh: toàn bộ mạng thay đổi chậm theo thời gian. Tập rời rạc ảnh mạng
 - Phân tích động \leftrightarrow mạng xã hội động: tương tác liên tục qua thời gian, tốc độ rất lớn. Dòng mạng.
- Một số bài toán
 - Phân tích thống kê mạng xã hội
 - Phát hiện cộng đồng trong mạng xã hội
 - Dự đoán liên kết, nút trong mạng xã hội
 - Phân tích vai trò
 - Phân loại nút trong mạng xã hội
 - Tiến hóa động mạng xã hội
 - Tính riêng tư trong mạng xã hội
 - Phân tích xung đột (*adversarial*), v.v.

đồng

- **Kiểm định thống kê**

- Kiểm định giả thuyết thống kê về mạng xã hội
- Các đặc trưng: thế giới nhỏ [Lescovec08], phân bố luật lũy thừa [Lescovec08], tác động tập nhân
- Tác động tập nhân [Easley10] : tính hạng đối tượng trong mạng xã hội sử dụng các thuật toán tính hạng trang web như PageRank, HITS...

- **Phát hiện cộng đồng**

- Cộng đồng tách rời, cộng đồng giao nhau
- Phổ phong phú các phương pháp: truyền thống, phân chia, dựa trên mô-dun hóa, dựa trên phổ, động, dựa trên suy luận thống kê, cộng đồng giao nhau, phân cấp, v.v. [Fortunato09]
- Một vài thuật toán phổ biến: họ thuật toán phân tách Girvan-Newman theo độ trung gian cạnh Girvan-Newman [Fortunato09], chia đỉnh CONGA, CONGO, gán nhãn COPRA [Gregory09], v.v.

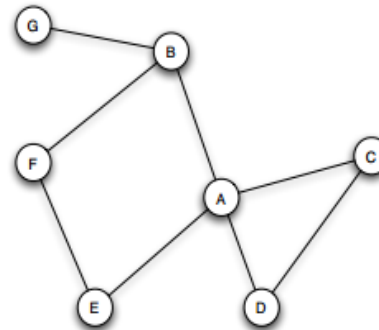
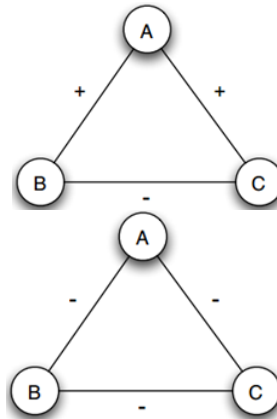
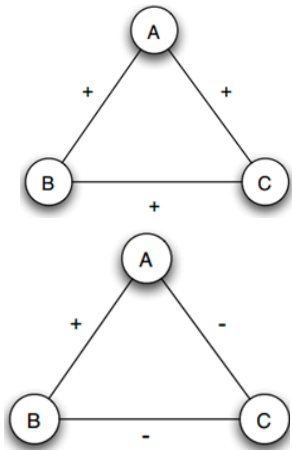
[Fortunato09] Santo Fortunato. *Community detection in graphs*. CoRRabs/0906.0612, 2009.

[Gregory09] Steve Gregory. *Finding Overlapping Communities Using Disjoint Community Detection Algorithms*. CompleNet 2009: 47-61.

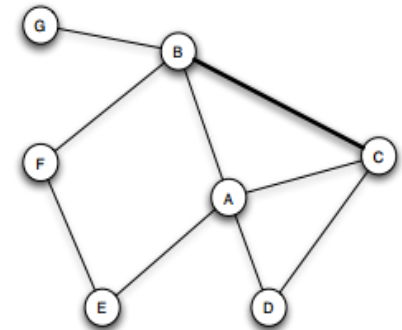
Dự đoán liên kết

• Dự đoán liên kết

- Hai nhóm phương pháp theo độ đo tương tự dựa trên cấu trúc: mô hình khả năng cực đại (Maximum Likelihood) kiểu phân cụm phân cấp và mô hình xác suất (Probabilistic) [Lu10, Wu15]
- Dự đoán liên kết âm-dương theo lý thuyết cân bằng cấu trúc: bộ ba, mạng cân bằng/không cân bằng [Easley10, Leskovec10]



(a) Trước



(b) Sau

[Lu10] Linyuan Lu, Tao Zhou. *Link Prediction in Complex Networks: A Survey*. CoRR abs/1010.0725. 2010.

[Wu15] Zhihao Wu, Youfang Lin, Jing Wang, Steve Gregory. *Efficient Link Prediction with Node Clustering Coefficient*. CoRR abs/1510.07819, 2015.

[Leskovec10] Jure Leskovec, Daniel P. Huttenlocher, Jon M. Kleinberg. *Predicting positive and negative links in online social networks*. WWW 2010: 641-650.

Phân tích vai trò

● Phân tích vai trò

- Vai trò là {quyền, nghĩa vụ, kỳ vọng, định mức và hành vi} của một người (nhóm) cần đối mặt và thi hành [Alhajj14]. mẫu hành vi đặc trưng (*characteristic behaviour pattern*).
- Hai câu hỏi: (i) Đối tượng X có vai trò gì? (ii) Ai có vai trò R?
- Tập đặc trưng của cá nhân trong quan hệ và tương tác xã hội (ví dụ, M1-M14 [Trabado12]), xác định các vai trò tương ứng (khởi tạo ý tưởng: *Information propagators/ Idea Starter*, nhận ý tưởng nhanh: *Early adopter*, quảng bá/phát tán: *Promoters/Amplifier*, quản lý: *curator*, nổi tiếng: *Celebrity*, v.v.)
- Các nhóm phương pháp: (i) Phân tích liên kết và nội dung theo mô hình xác suất, (ii) Phân tích mạng xã hội theo độ đo cấu trúc, (iii) Tối ưu hóa tổ hợp, (iv) học máy giám sát, học máy phân lớp hoặc tính hạng [Wang14, Trabado12].

[Gliwa13] E Bogdan Gliwa, Anna Zygmunt, Jaroslaw Kozlak. *Analysis of Roles and Groups in Blogosphere*. CORES 2013: 299-308.

[Trabado12] Vanesa Junquero-Trabado, David Dominguez-Sal. *Building a role search engine for social media*. WWW (Companion Volume) 2012: 1051-106

[Wang14] Chi Wang, Jiawei Han. Role Discovery. *Encyclopedia of Social Network Analysis and Mining*: 1589-1598. Springer, 2014.



Công cụ SNAP Python

- **Stanford Network Analysis Platform (SNAP)**

<http://snap.stanford.edu>

Trên 70 bộ dữ liệu mạng

- **Prebuilt packages available for Mac OS X, Windows, Linux**

[http:// snap.stanford.edu/snappy/index.html](http://snap.stanford.edu/snappy/index.html)

- **Snap.py documentation:**

<http://snap.stanford.edu/snappy/doc/index.html>

Quick Introduction, Tutorial, Reference Manual

- **SNAP user mailing list**

<http://groups.google.com/group/snap-discuss>

- **Developer resources**

Software available as open source under BSD license

GitHub repository

<https://github.com/snap-stanford/snap-python>



Công cụ SNAP C++

- Source code available for Mac OS X, Windows, Linux
<http://snap.stanford.edu/snap/download.html>
- SNAP documentation
<http://snap.stanford.edu/snap/doc.html>
Quick Introduction, User Reference Manual
Source code, see tutorials
- SNAP user mailing list
<http://groups.google.com/group/snap-discuss>
- Developer resources
Software available as open source under BSD license
GitHub repository
<https://github.com/snap-stanford/snap>
SNAP C++ Programming Guide



Dữ liệu và thu thập dữ liệu từ mạng xã hội trực tuyến

Dữ liệu cá nhân

- **Hồ sơ cá nhân**
 - Tuổi, giới tính, tình trạng hôn nhân,...
 - Nơi làm việc, nghề nghiệp, trường học, bằng cấp, học vấn,...
 - Gia đình, thành phần gia đình,...
- **Sở thích**
 - Ngắn hạn, trung hạn, dài hạn,...
- **Hành vi/Thói quen**
 - Thói quen sinh hoạt, làm việc, du lịch,...
- **Ý định/Nhu cầu**
- **Cảm xúc**
- **Tính cách**

Hồ sơ cá nhân



Sunset Trang



Former Giảng viên at Trường Cao đẳng Công nghệ Dệt may Thời trang Hà Nội
September 2009 to May 2011



Went to Trường THPT Trần Phú - Hoàn Kiếm
Past: THCS Ngo Si Lien



Lives in Hanoi, Vietnam

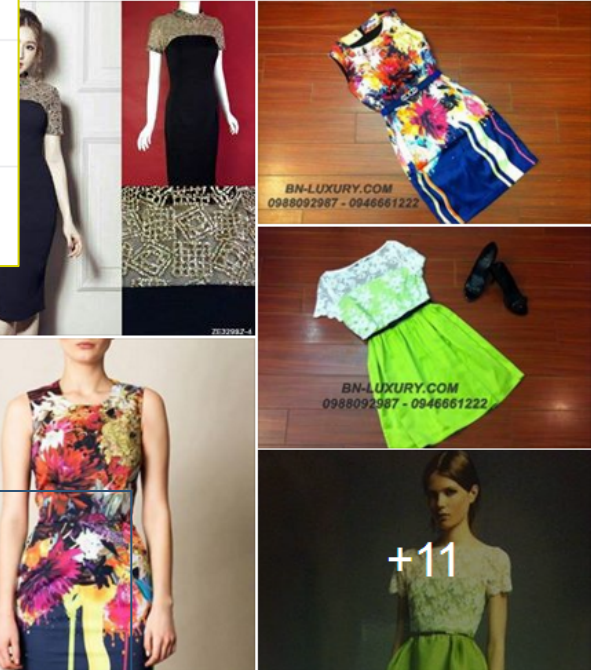


Sunset Trang added 15 new photos to the album Váy, đầm dự tiệc, party, đi biển... cho THUẾ

Monday at 20:12 · 🌐

ều váy điệu đà, party, dạ tiệc, maxi...hầu như còn mới nguyên cũng phí. Mình muốn hỏi, các chị, em có nhu cầu cần THUẾ cưới, đi ăn cưới, đi party, đi biển không ạ? vì nói thật hầu như váy của chị em chúng mình cũng thường chỉ được diện 1 lần, nếu các em có nhu cầu mình sẽ up ảnh mẫu, đảm bảo đẹp ạ.

5 việc cần đến váy áo đẹp và đặc biệt chú như: đi ăn cưới, đi sinh nhật, hoặc đơn giản như đi chơi, đi biển... See more



Like Share



Vân Anh Hoàng

20 May at 22:40 · 🌐

Hu hu. ho sù sù ko khỏi ak. Mai thuyết trình kiểu gì chứ :(((— 😞 feeling sad.

Like · Comment · Share

👍 Mai Công Đạt and 9 others like this.



Võ Hoài Linh

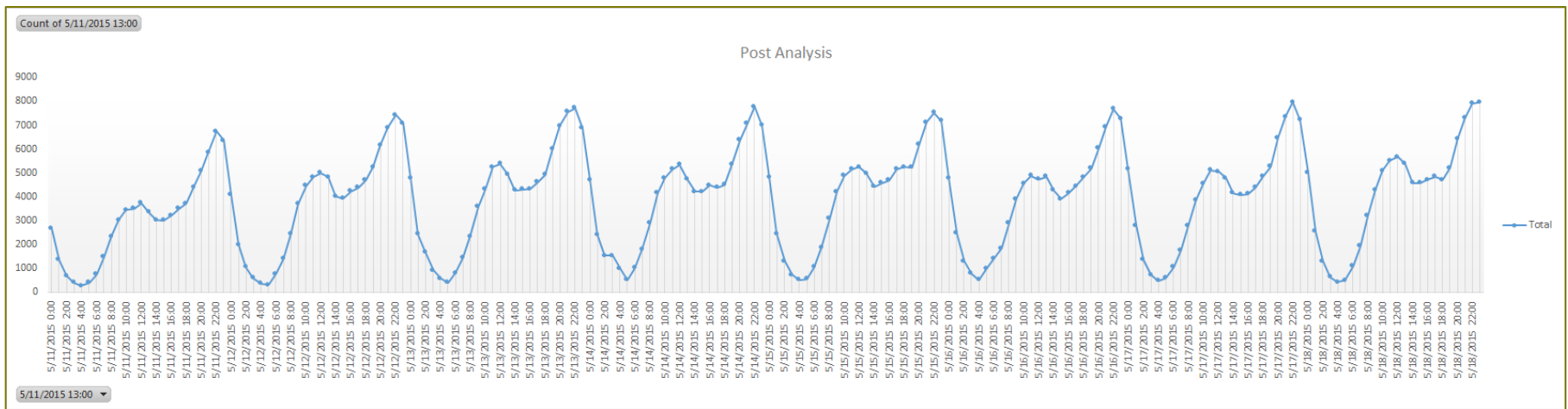
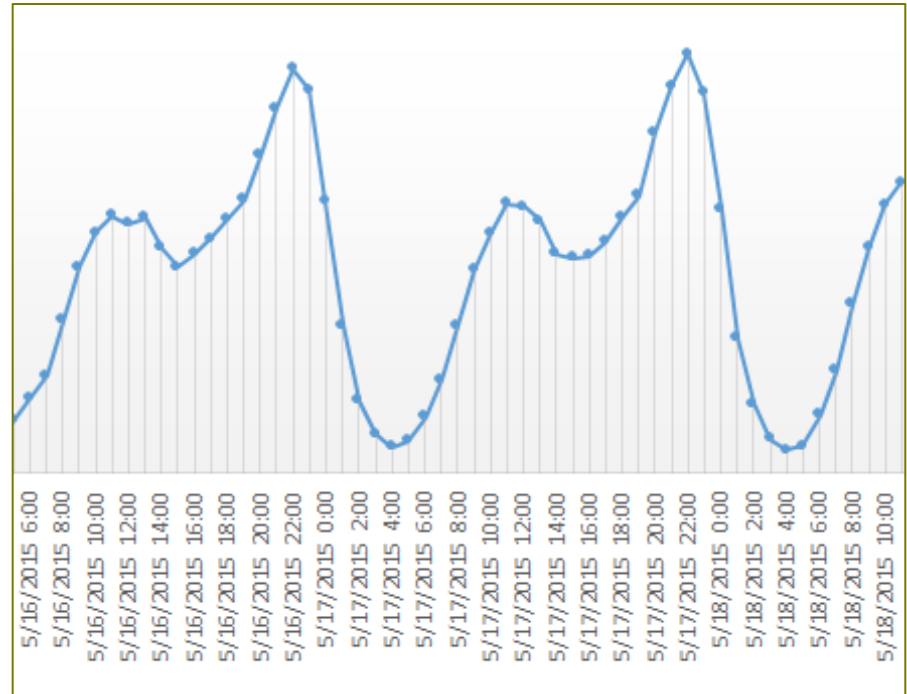
2 October near Buchs, Aargau, Switzerland 🌐

Mùa thu Thụy Sĩ, thấy cây táo cũng nhào vô chụp hình.... kkkk

3/10/2013	Này thì em cũng lặn lội đến FIFA rồi đấy kkkkkk	FIFA-Strasse 20, P.O. Box 8044 Zurich, Switzerland
2/10/2013	Mùa thu Thụy Sĩ, thấy cây táo cũng nhào vô chụp hình.... Kkkk	Buchs, Aargau, Switzerland
21/09/2013	Trời mưa quá em, ơi bầu show " lướt " mất rồi còn đâu . Đừng mưa nữa	Long Xuyên
7/09/2013	Lên đường thôi . Đài Loan thẳng tiến . Tạm biệt Sài Gòn vài hôm nhé, a	Đài Loan
5/09/2013	Giây phút nhàn hạ của nông dân miền tây sông nước .	Miền Tây
18/08/2013	Cảm ơn tình cảm của khán giả Huế (Trọng Tấn) , năm nào Linh ra diễn	Huế
7/08/2013	Uống đi đại huynh , bia quán nghèo nhưng nông nàn hương vị	Hồ Chí Minh

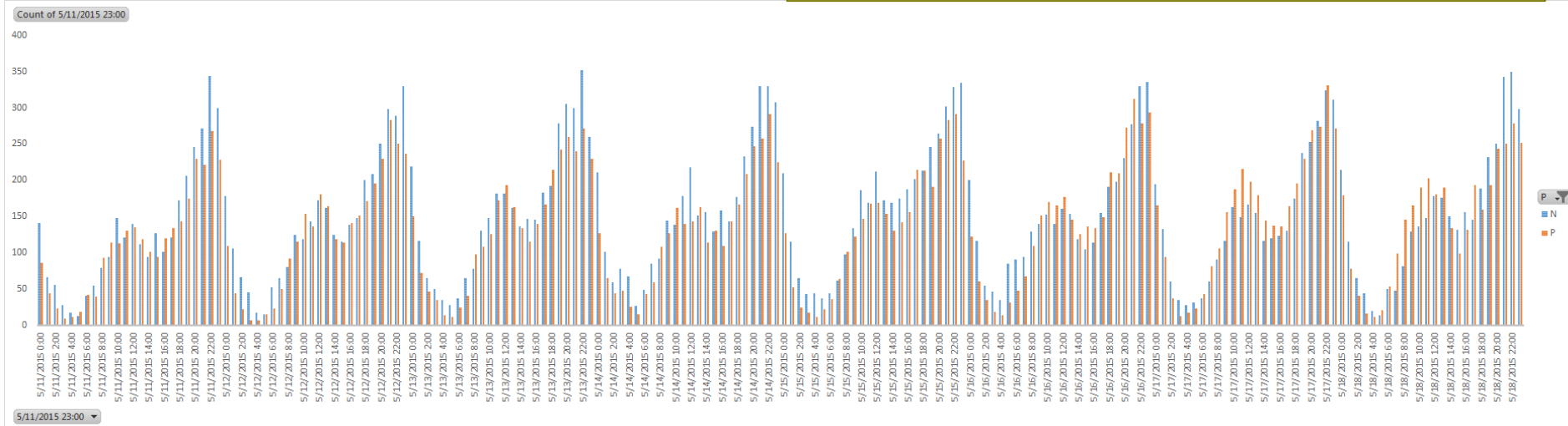
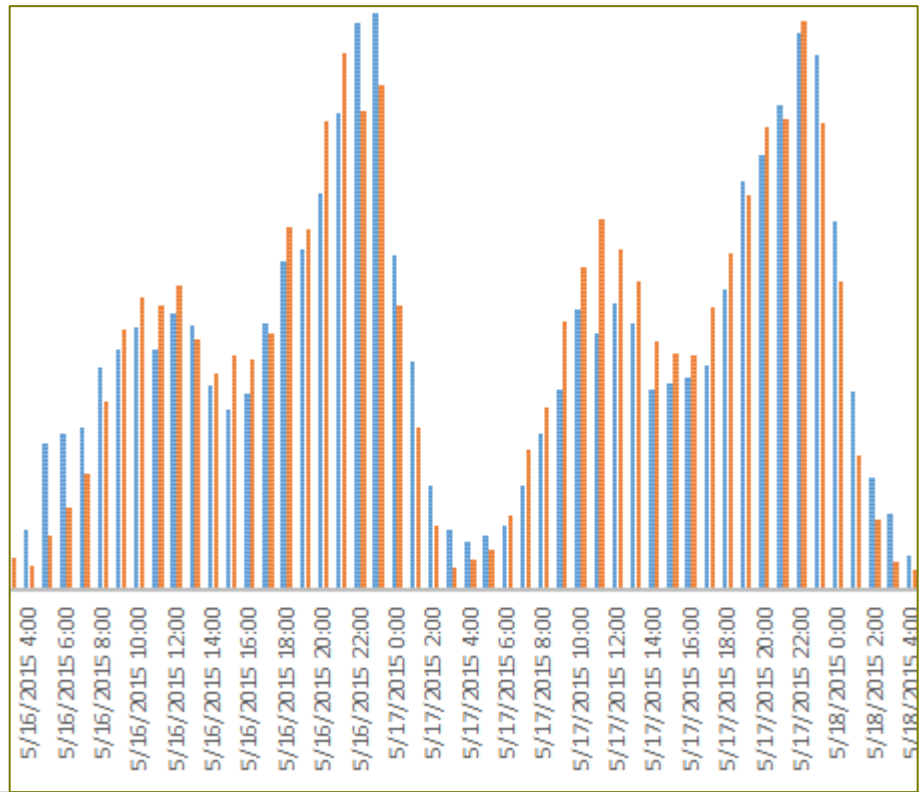
Dữ liệu cá nhân

- Phân bố thời gian viết bài trong một tuần
- Thời gian viết bài tập trung vào 18-23h hàng ngày
- Giảm trong giờ nghỉ trưa và nghỉ tối



Dữ liệu cá nhân

- Phân bố cảm xúc người dùng
- Thứ 2-6 tiêu cực nhiều hơn
- Thứ 7-CN thẳng hoa



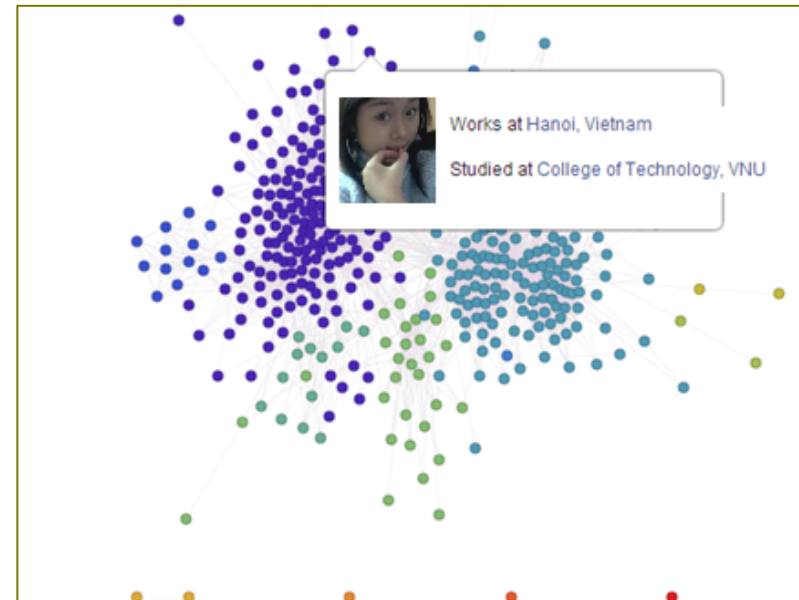
Dữ liệu quan hệ



- Các mối liên kết quan hệ bạn bè
 - Mạng bạn bè Facebook
- Quan hệ công việc, chuyên môn
 - Mạng chuyên gia LinkedIn
- Mối quan hệ theo dõi một chiều
 - Follow một chiều một cá nhân trên Twitter hay Facebook
- Mối quan hệ giữa cá nhân và doanh nghiệp/người nổi tiếng
 - Các thành viên yêu thích một thương hiệu hoặc fan của người nổi tiếng

Dữ liệu cộng đồng

- Những cá nhân có cùng sở thích
 - Tham gia cùng một group trên Facebook
 - Tham gia cùng một circle trên G+
- Những cá nhân tham cộng đồng về một lĩnh vực cụ thể
 - Tham gia vào cùng một Group hay Association trên LinkedIn
 - Tham gia vào những danh sách nhận thông tin từ tài khoản Twitter



Dữ liệu cộng đồng



TỰ LONG
(Long Chèo)

[Theo dõi](#) [Nhắn tin](#) [...](#)

[Dòng thời gian](#) [Giới thiệu](#) [Bạn bè](#) [Ảnh](#) [Xem thêm ▾](#)

Theo dõi Long để xem bài viết công khai của anh ấy trong Bảng tin của bạn.

[Theo dõi](#)



79.166 người theo dõi

Nhóm

Công khai



Hội Hot boy, Hot girl Việt Nam

[+ Tham gia](#)

67.602 thành viên

Mục đích là để giao lưu văn hoá, du lịch và lưu giữ những hình ảnh đẹp của tuổi trẻ!!! ồ...



Chợ Gỗ Hoàng Long

[+ Tham gia](#)

4.018 thành viên

Nơi gặp gỡ, giao lưu của những người đam mê Gỗ.



Câu lạc bộ 1973

[+ Tham gia](#)

7 thành viên

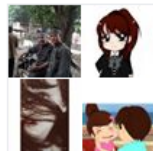


CLB TENNIS TỔ CHYM

[+ Tham gia](#)

363 thành viên

CLB TỔ CHYM NƠI HỘI TỤ ACE YÊU TENNIS, THÍCH DU LỊCH VÀ KẾT NỐI BẠN BÈ Các cá nhân hay các...



Đạo diễn Phạm Nguyễn Bắc

[+ Tham gia](#)

347 thành viên



Đền với nghệ thuật chèo

[+ Tham gia](#)

5.778 thành viên

Chèo là một loại hình nghệ thuật sân khấu dân gian Việt Nam. Chèo phát triển mạnh ở đồng...

Thu thập dữ liệu từ MXH

- Các mạng xã hội lớn đều cho phép truy xuất dữ liệu qua các API dạng Webservice
 - Lưu ý về chính sách sử dụng dữ liệu
 - Lưu ý về chính sách thu thập dữ liệu
- Facebook
 - Chỉ cho phép truy xuất dữ liệu của chính bản thân và các Page hay group mở
 - Một số Partner được phép sử dụng một số kênh mất chi phí
- Twitter
 - Cho phép thu thập và sử dụng toàn bộ dữ liệu
- LinkedIn
 - Trả phí cho việc đăng tuyển và tìm ứng cử viên công việc

Thu thập dữ liệu từ Twitter

- **Search API**

- Cho phép thu thập dữ liệu dựa trên từ khóa tìm kiếm và một số tiêu chí lọc dữ liệu.
- Dữ liệu trả về trong 7 ngày gần nhất

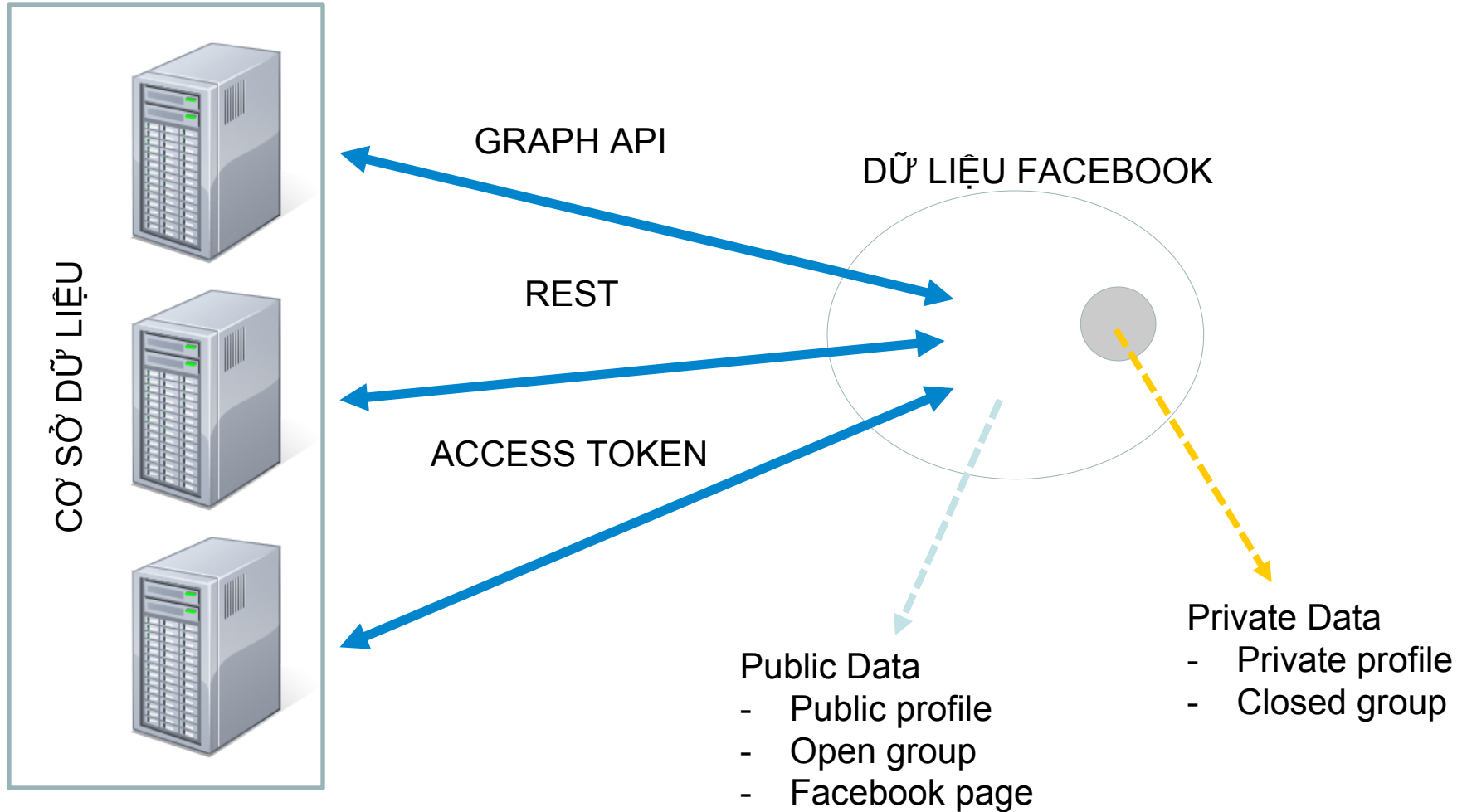
- **Streaming API**

- Được phép truy xuất để lấy dữ liệu Twitter theo thời gian thực
- Cho phép thiết lập các tham số liên quan đến địa điểm đưa các tweet

- **Thư viện**

- Java: Twitter4J, jTwitter
- Python: Tweepy

Thu thập dữ liệu từ Facebook





Thu thập dữ liệu từ Facebook

- **Graph API**
 - Tương tác với dữ liệu Open group, page, user tham gia app và tài khoản của chính bản thân
- **Access Token**
 - User access token: chỉ truy xuất đến thông tin cá nhân và một số thông tin của bạn bè trực tiếp
 - App access token: chỉ truy xuất đến thông tin của các user tham gia vào app (với điều kiện user cho phép)
 - Page access token: chỉ truy xuất vào thông tin của page
- **Thư viện**
 - Java: RestFB, Facebook4J
 - Python: Facebook SDK for Python



Phân tích nội dung mạng trực tuyến

Phân tích quan điểm

- Phân tích tình cảm (Sentiment Analysis - khai phá quan điểm: Opinion Mining)
 - Phân tích quan điểm (opinion), tình cảm/tâm lý (sentiment), đánh giá (evaluation), thẩm định (appraisal), thái độ (attitude), và cảm xúc (emotion) của con người đối với các thực thể như sản phẩm (product), dịch vụ (service), tổ chức (organization), cá nhân (individual), vấn đề (issue), sự kiện (event), chủ đề (topic) và các thuộc tính của chúng
- Thuật ngữ:
 - sentiment analysis / opinion mining /opinion extraction /sentiment mining / subjectivity analysis/ affect analysis /emotion analysis /review mining... Phổ biến nhất: opinion mining (hàn lâm - công nghiệp)/ sentiment analysis (công nghiệp)

Phân tích quan điểm

- **Mức tài liệu (document-level sentiment classification)**
 - Toàn bộ tài liệu thể hiện một quan điểm tích cực (positive) / tiêu cực (negative). Phân lớp chứa/không quan điểm.
 - Bài toán phân lớp: Tài liệu chứa quan điểm: tích cực / tiêu cực
- **Mức câu (sentence level: subjectivity classification)**
 - Cho quan điểm tích/tiêu cực hoặc trung tính (neutral). Trung tính ~ không có quan điểm.
 - Phân lớp câu: khách quan (objective sentences) và chủ quan (subjective sentence)
 - Câu chủ quan không tương đương câu có quan điểm.
 - Câu khách quan “Tôi mua chiếc xe tháng trước và chiếc gạt nước đã bị rơi” có quan điểm.

Phân tích quan điểm

- Mức thực thể và khía cạnh (Entity and Aspect level / aspect level / feature level / feature-based opinion mining and summarization)
 - Đối tượng và các khía cạnh của đối tượng
 - Mức khía cạnh phát hiện chính xác thích gì và không thích gì
 - Toàn bộ đối tượng và từng khía cạnh,
 - Chi tiết theo từng khía cạnh, hấp dẫn và phức tạp nhất.
 - Một số ví dụ
 - "mặc dù dịch vụ là không đáng kể, tôi vẫn thích nhà hàng này". Toàn bộ nhà hàng: tích cực, khía cạnh dịch vụ : Tiêu cực
 - "chất lượng cuộc gọi của iPhone là tốt, nhưng tuổi thọ pin của nó là ngắn". Khía cạnh cuộc gọi: tích cực", khía cạnh tuổi thọ pin: tiêu cực.

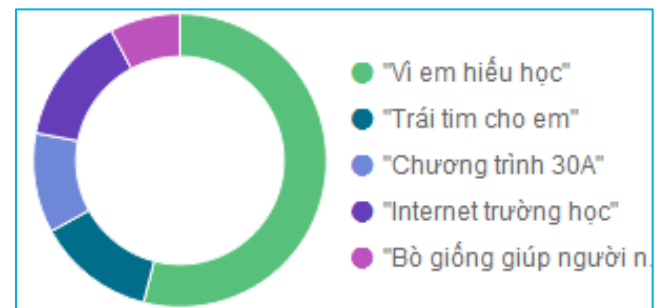
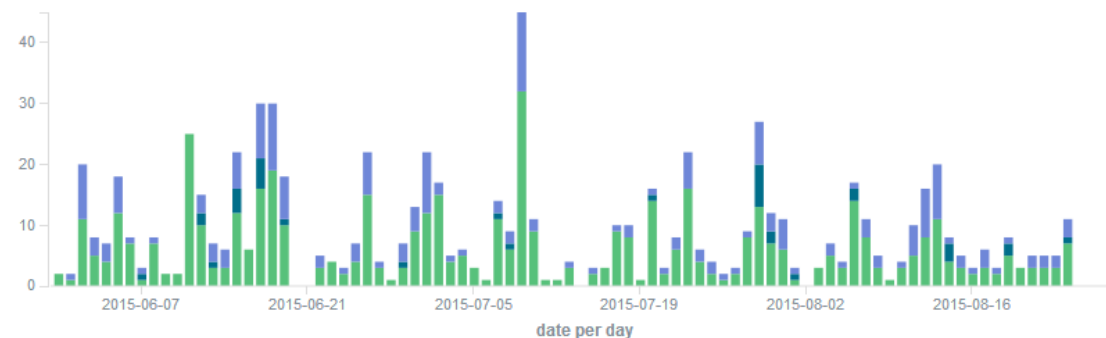
Phân tích quan điểm

- **Quan điểm so sánh**
 - So sánh hai hay nhiều thực thể với nhau
 - So sánh nhất: tốt/tồi nhất
 - So sánh hơn kém
 - So sánh không thứ bậc: hai khía cạnh khác nhau
- **Từ quan điểm**
 - Từ quan điểm: một dấu hiệu nhận diện quan điểm
 - Từ vựng quan điểm (SentiWordNet)

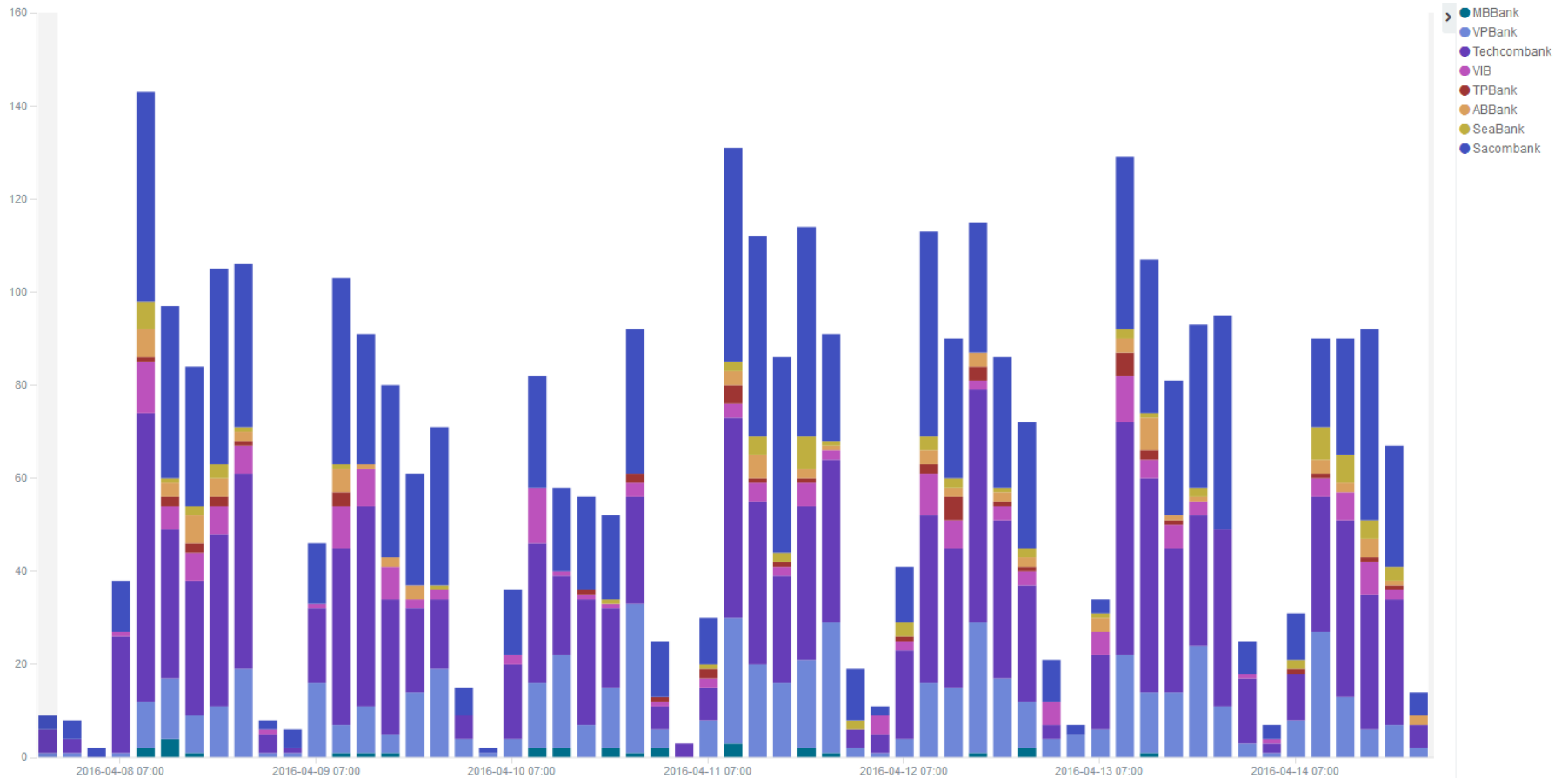
Quản lý thương hiệu

- Theo dõi, giám sát mức độ thâm nhập, sức lan tỏa, và ảnh hưởng của thương hiệu trên truyền thông trực tuyến/mạng xã hội
- Đo đếm mức độ truyền thông/dự luận trực tuyến đề cập tới các hoạt động, sự kiện, chiến dịch hoặc bất cứ một vấn đề nào đó của doanh nghiệp, tổ chức theo thời gian.
- Phát hiện nhanh chóng các luồng dự luận thiếu chính xác, sai lệch ảnh hưởng bất lợi đến vị thế và uy tín của thương hiệu nhằm đưa ra quyết định và giải pháp can thiệp kịp thời

Viettel - Thống kê quan điểm theo ngày



Quản lý thương hiệu



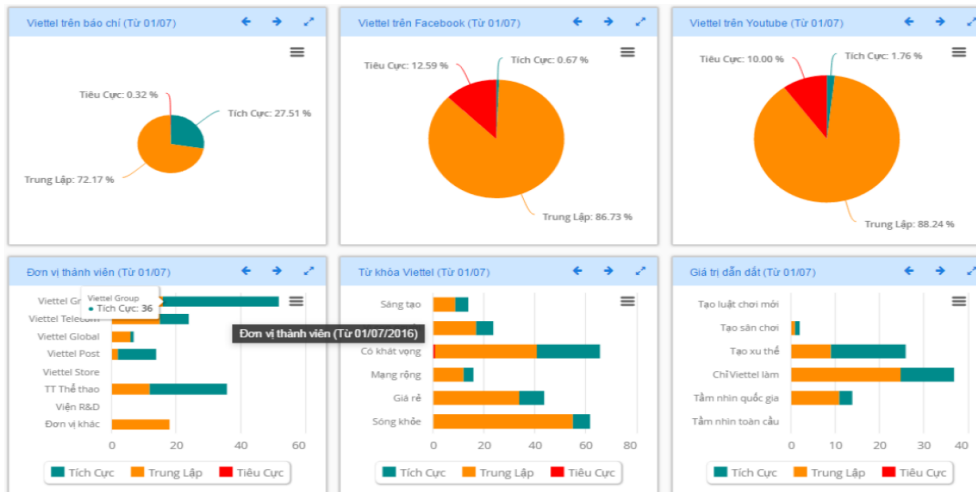
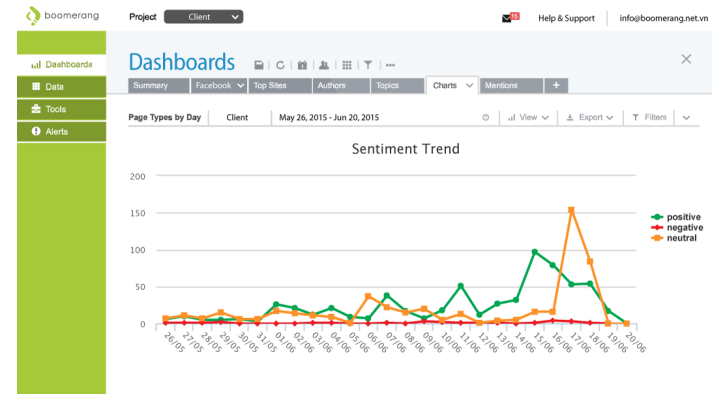
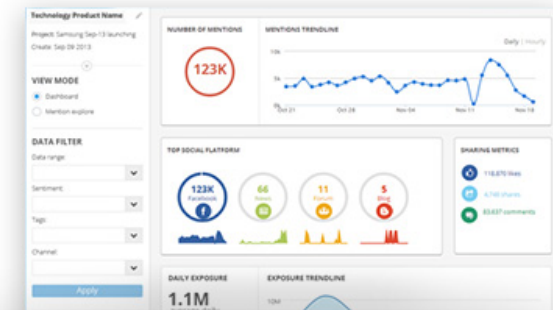
Quản lý thương hiệu

- Barclays phát hành ứng dụng Mobile Banking chỉ cho người 18 tuổi trở lên nhưng sau đó cho cả người 16-17 tuổi vì phản hồi khách hàng
- Theo Evry, các ngân hàng lớn trên thế giới đều sử dụng công cụ giám sát mạng xã hội và phân tích quan điểm nhằm lắng nghe, chăm sóc khách hàng và quản lý thương hiệu



Quản lý thương hiệu

- Doanh nghiệp ứng dụng
 - Viettel, BIDV
- Sản phẩm
 - Viettel Social Monitoring
 - Younet Media & BuzzMetric
 - Boomerang
 - SMCC.VN (InfoRe)
 - iMonitor (Datasection)



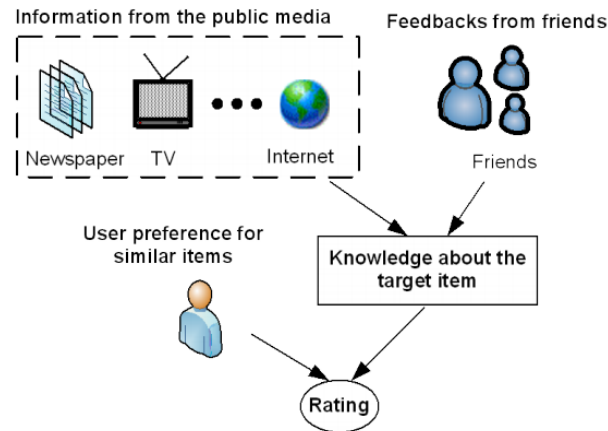
Tư vấn xã hội

- Định nghĩa hẹp:



- Định nghĩa rộng:

- Là hệ tư vấn bất kì nhắm đến lĩnh vực phương tiện xã hội
- Đối tượng: con người, cộng đồng, mục, thẻ,...
- Nguồn sử dụng: dữ liệu phương tiện xã hội (các mối quan hệ xã hội, tương tác người dùng, gắn thẻ,...)



Tự vấn xã hội

Jobs you may be interested in

- LAZADA GROUP**
Identity & Access Management Analyst
Lazada Group — Vietnam
[View Job](#)
- SDL**
Senior Researcher - Insight and Analytics
SDL plc — Vietnam
[View Job](#)
- SDL**
Researcher - Insight and Analytics
SDL plc — Vietnam
[View Job](#)

See more jobs >

People like you are watching

Foundations of Business Analytics: Prescriptive Analytics
Everyone is talking about big data these days, but that's just the starting point for drawing high-value, actionable insights from your...
17,441 views
[View more courses >](#)

Vietcombank nói gì về việc chủ thẻ mất 500 triệu sau 1 đêm?
(Dân trí) - Theo phản ánh của bà Hoàng Thị Na Hương, khách hàng của Ngân...
DANTRI.COM.VN | BỞI DÂN TRÍ

Xem thêm từ Dân trí. [Thích](#)

[Anh Le](#) và 149 người khác 23 bình luận 35 lượt chia sẻ

[Thích](#) [Chia sẻ](#)

MỌI NGƯỜI CŨNG CHIA SẺ

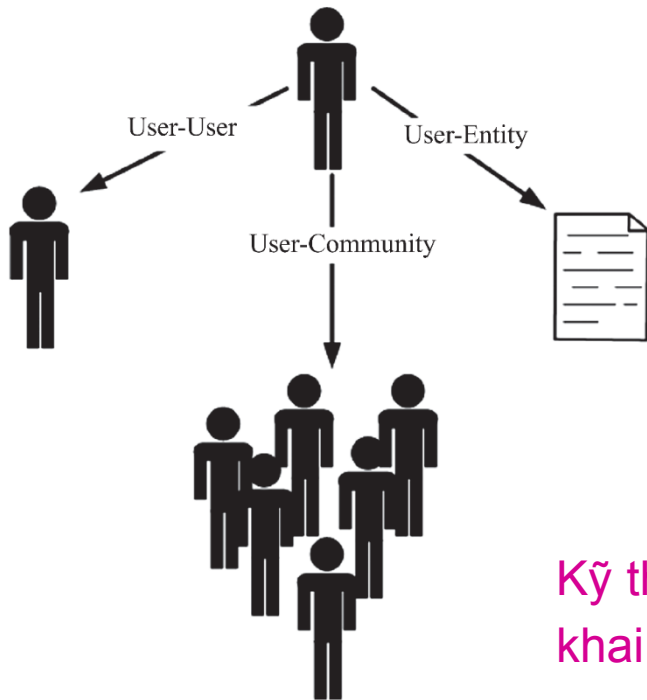
Đã tìm ra nguyên nhân chủ thẻ Vietcombank mất 500 triệu đô ng
Vietcombank khẳng định nguyên nhân khiến chủ thẻ mất nửa tỉ đồng khi đang ngủ là do trước đó, khách...
INFONET.VN · 3.789 LƯỢT CHIA SẺ [Chia sẻ](#) [Lưu](#)

Hàng loạt ngân hàng cảnh báo nguy cơ mất tiền sau vụ hacker
(Dân trí) - Nhiều ngân hàng đồng loạt gửi tin nhắn đến các khách hàng sử dụng thẻ ATM cảnh báo không...
DANTRI.COM.VN · 257 LƯỢT CHIA SẺ [Chia sẻ](#) [Lưu](#)

Chủ thẻ Vietcombank bỗng dưng mất 500 triệu đô ng chỉ qua một đêm
Chiều ngày 11.8, Dân Việt có nhận được đơn tường trình của chị H.T.N.Hương, khách hàng của ngân hà...
BAOMOI.COM · 1.057 LƯỢT CHIA SẺ [Chia sẻ](#) [Lưu](#)

Phân tích hành vi người dùng

- Mục đích
 - Phân tích, mô hình và dự đoán một hoặc nhiều hành vi của người dùng trên mạng xã hội
- 3 loại hành vi của một cá nhân trên MXH



User-User (link generation)

Trở thành bạn, gửi tin nhắn, cùng chơi trò chơi, theo dõi, hoặc mời tham gia một sự kiện

User-Community

Tham gia hay rời khỏi một cộng đồng hoặc đóng góp thảo luận vào một cộng đồng

User-Entity (content generation)

Viết bài, đăng ảnh

Kỹ thuật: Xử lý ngôn ngữ tự nhiên, học máy, thống kê, khai phá dữ liệu

Ứng dụng phân tích hành vi người dùng

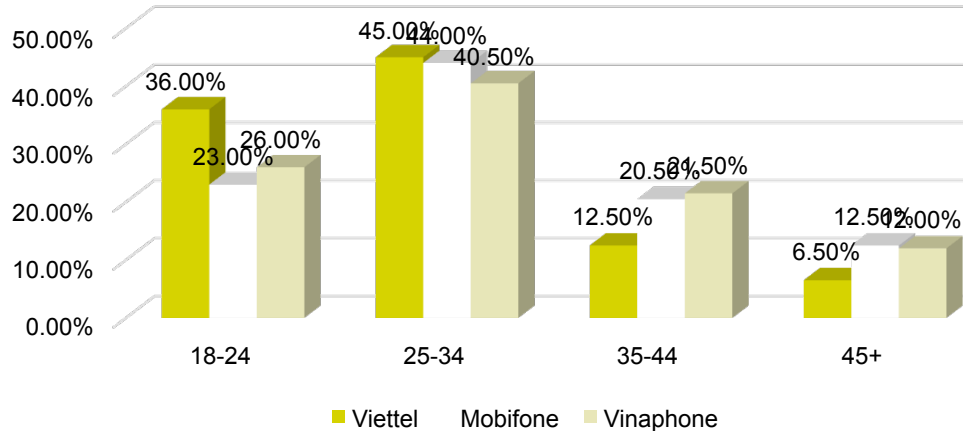
- **Phân tích khách hàng (Customer insight)**
 - Phân tích tập và hiểu tập khách hàng nội bộ nhằm cải thiện chiến dịch tiếp thị, bán có mục tiêu và đưa dịch vụ tốt hơn
 - Hiểu biết kế hoạch sản phẩm của khách hàng để tìm kiếm sản phẩm khách hàng có khả năng sắp mua nhất nhằm đưa ra thông điệp tiếp thị đúng đắn
- **Hiệu quả tại một số công ty lớn**
 - OCBC hiểu khách hàng hơn thông qua hành vi và thói quen khách hàng qua đây tăng sự gắn kết giữa khách hàng-ngân hàng
 - Westpac phát hiện xu hướng khách hàng và gom nhóm các sản phẩm để tiếp thị khách hàng, hiệu quả đạt được tăng 37% hiệu quả đăng ký sử dụng dịch vụ và 60% khách hàng liên hệ với ngân hàng
 - Manulife ước lượng thu nhập và thói quen của khách hàng nhằm tạo ra các dịch vụ tối ưu cho khách hàng sử dụng

PHÂN TÍCH KHÁCH HÀNG VIỄN THÔNG

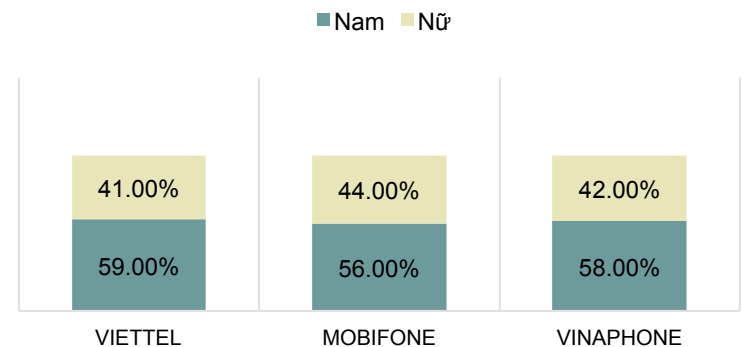
Phân tích thông tin tất cả các khách hàng có đầu số 09 của 3 nhà mạng Viettel (096, 097, 098), Mobifone (090, 093), Vinaphone (091, 094)

	Số lượng tài khoản Facebook	Số lượng tài khoản Facebook hoạt động trong tháng 6/2016
Viettel	11.2 triệu	6.5 triệu
Mobifone	8.0 triệu	4.7 triệu
Vinaphone	5.5 triệu	2.9 triệu

Phân bố độ tuổi



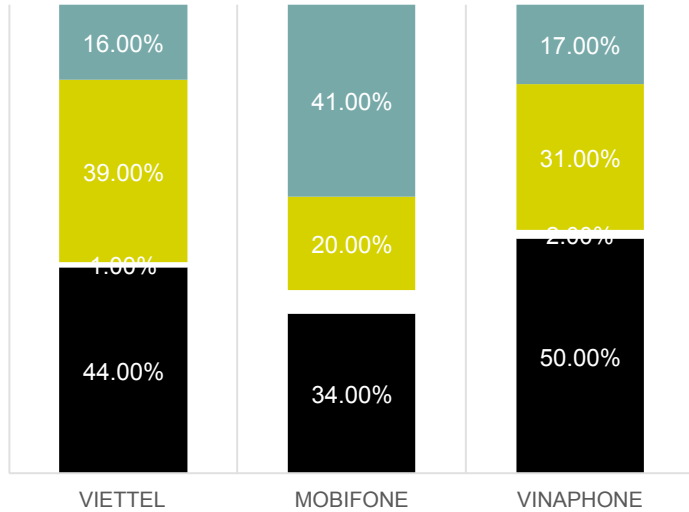
PHÂN BỐ GIỚI TÍNH





PHÂN TÍCH KHÁCH HÀNG VIỄN THÔNG

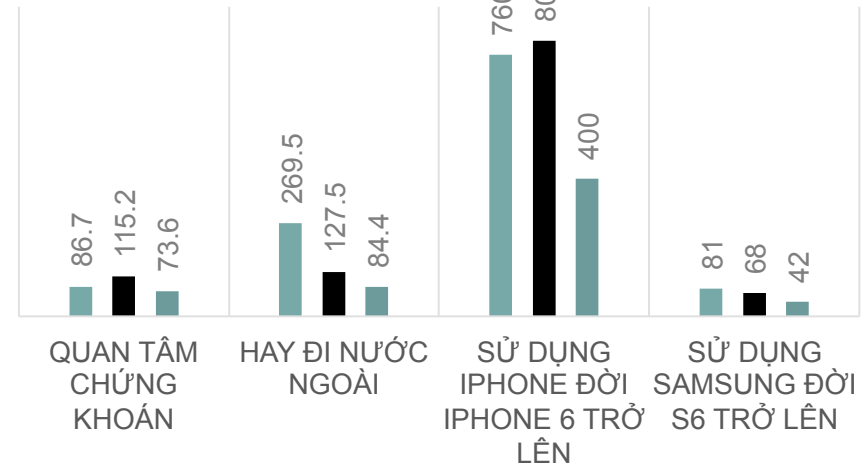
■ Khác ■ Đà Nẵng ■ Hà Nội ■ Hồ Chí Minh



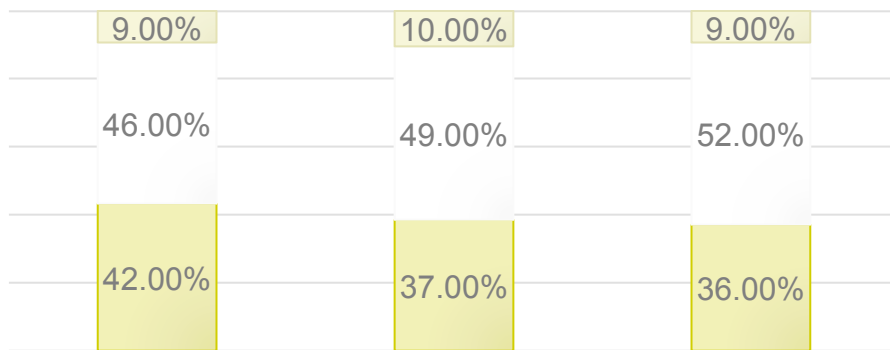
PHÂN TÍCH SỞ THÍCH, HÀNH VI, THÓI QUEN

(Đơn vị: 100)

■ Viettel ■ Mobifone ■ Vinaphone



Tình trạng hôn nhân



■ Độc thân ■ Đã kết hôn ■ Đang hẹn hò



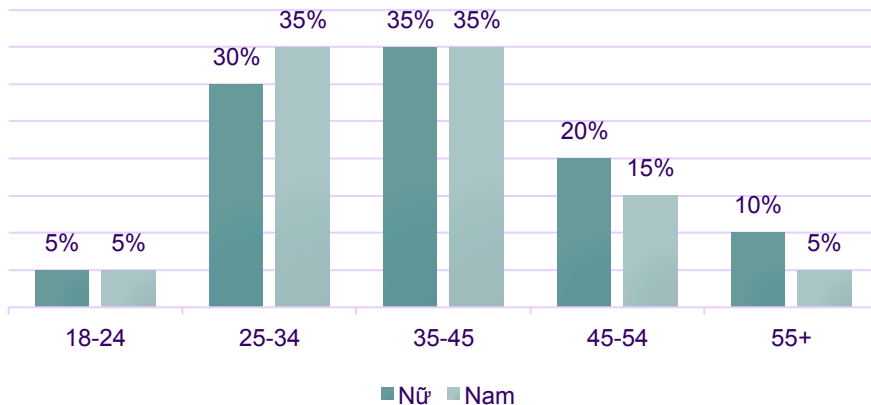
PHÂN TÍCH KHÁCH HÀNG NGÂN HÀNG

Hệ thống phân tích khách hàng của một ngân hàng tầm trung tại Việt Nam với hơn 2 triệu khách hàng, 300 triệu dữ liệu giao dịch chuyển khoản hàng năm

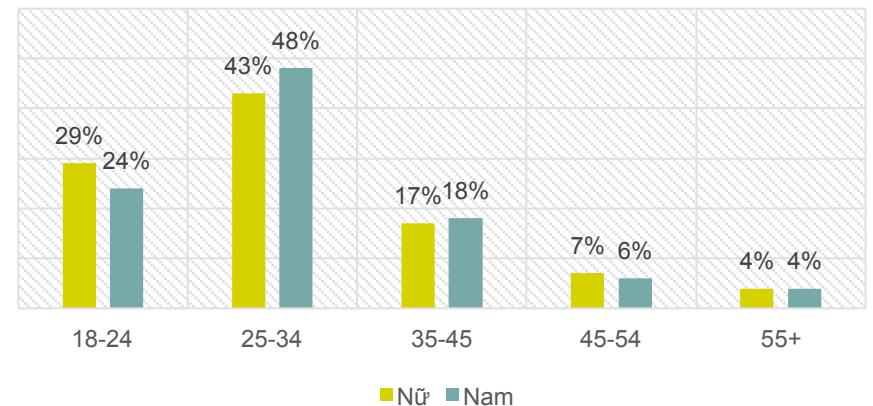
	Số lượng khách hàng 04/2016	Số lượng tài khoản Facebook	Số lượng tài khoản Facebook hoạt động trong tháng 6/2016	Phân bố giới tính (Nam/Nữ)
--	-----------------------------	-----------------------------	--	----------------------------

Tổng số khách hàng	2.048 triệu	836 nghìn	615.8 nghìn	50/50
Khách hàng Vip	19.5 nghìn	10.4 nghìn	8.6 nghìn	50/50

Phân bố giới tính và độ tuổi KHÁCH HÀNG VIP

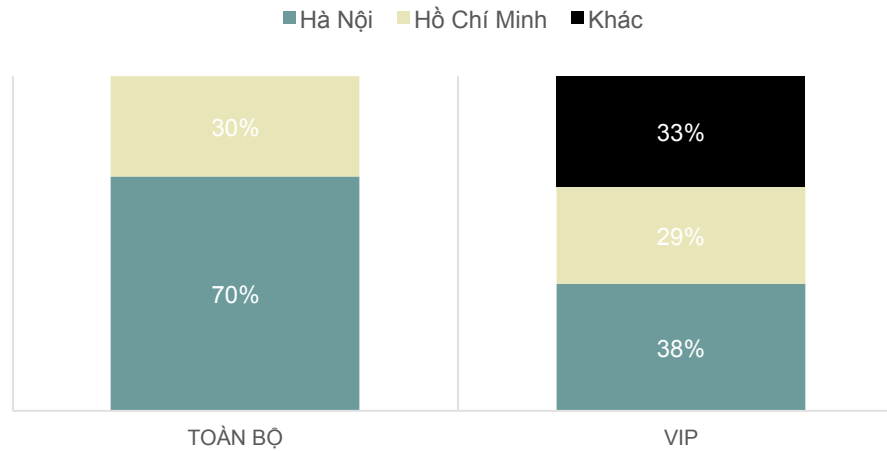


Phân bố giới tính và độ tuổi toàn bộ khách hàng



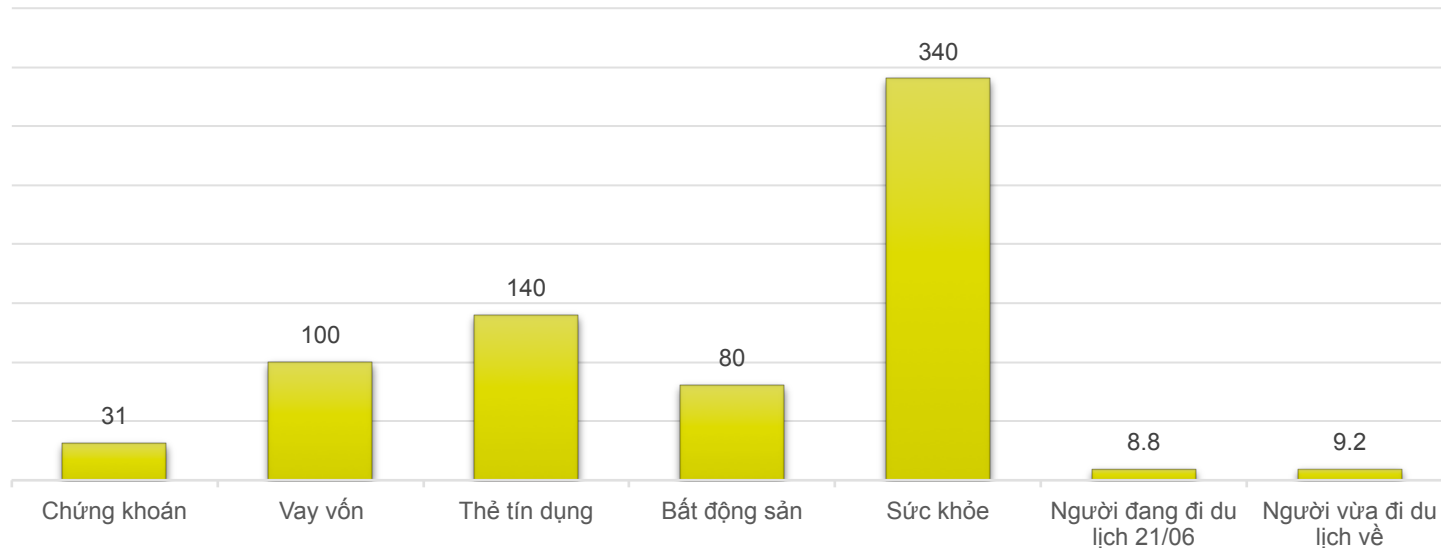
PHÂN TÍCH KHÁCH HÀNG NGÂN HÀNG

PHÂN BỐ THEO THÀNH PHỐ



PHÂN TÍCH XU HƯỚNG VÀ HÀNH VI KHÁCH HÀNG

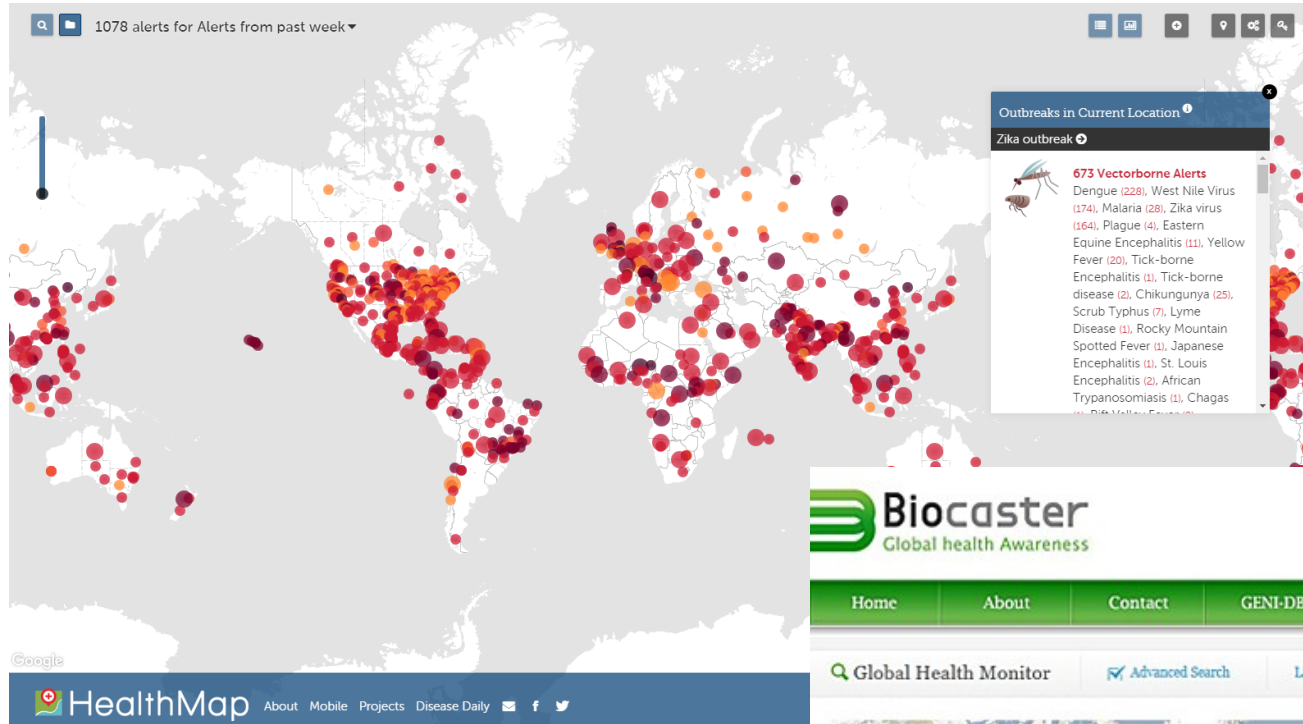
(Đơn vị: 1000 người)



Giám sát sự kiện trên mạng xã hội

- **MXH là công cụ truyền thông quan trọng:**
 - Tuyên truyền về thảm họa, thiên tai (động đất, sóng thần Nhật Bản, bão Katrina, tràn dầu Louisiana,...)
 - Vận động tranh cử, tổng tuyển cử (Mỹ, HQ, Anh,...)
 - Giúp quan chức Chính phủ quản lý, theo dõi các luồng thông tin
 - Thu thập thông tin TB phục vụ mục đích chính trị
- **Công cụ giám sát sự kiện**
 - Hiểu ngôn ngữ tốt hơn
 - Trích xuất sự kiện, nhận dạng thực thể
 - Gom nhóm sự kiện, dịch máy thống kê
 - Thường tập trung vào một số miền dữ liệu cụ thể
 - Bệnh truyền nhiễm
 - Thiên tai, thảm họa
 - Chính trị

Giám sát sự kiện trên mạng xã hội



Biocaster
Global health Awareness

- Home
- About
- Contact
- GENI-DB
- Ontology
- Trends
- Downloads
- Login

Global Health Monitor [Advanced Search](#) Language filter: en Last update 04:36 AM JST: Next update 05:06 AM

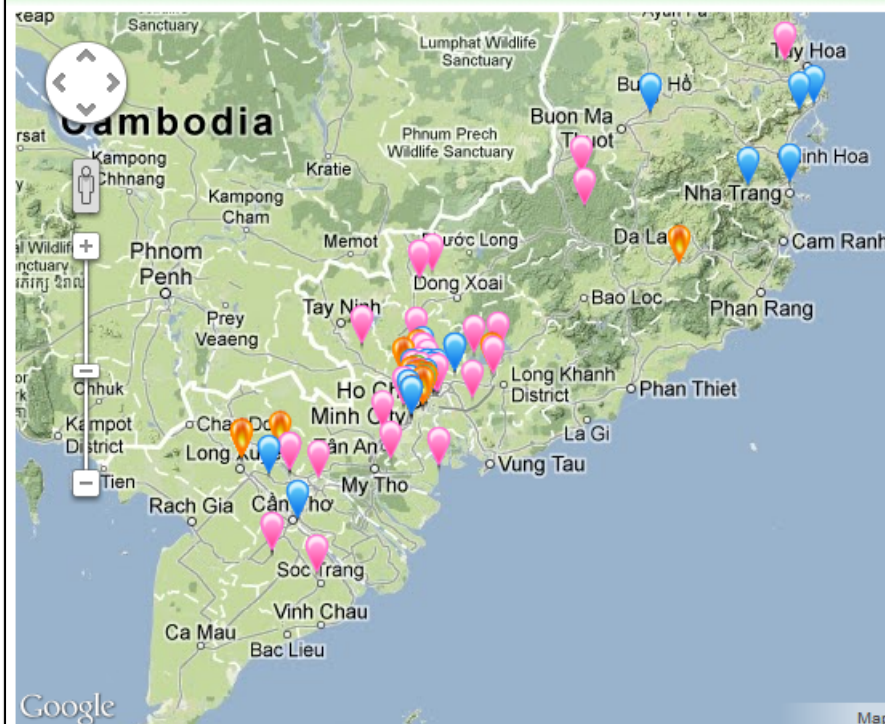


Giám sát sự kiện Việt Nam

Event Monitoring

Trang chủ | Giới thiệu

Xem:



Tóm tắt sự kiện

Map Satellite

Loại sự kiện : **Hình sự**

Thời gian : **Fri Sep 07 04:57:00 GMT+07:00 2012**

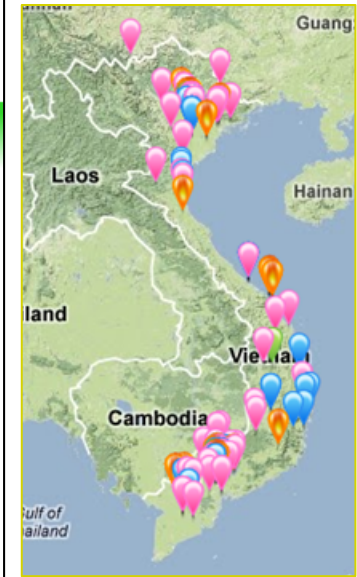
Nguồn ngữ : **VN**

Địa điểm : **Bình Dương**

Đất nước : **Việt Nam**

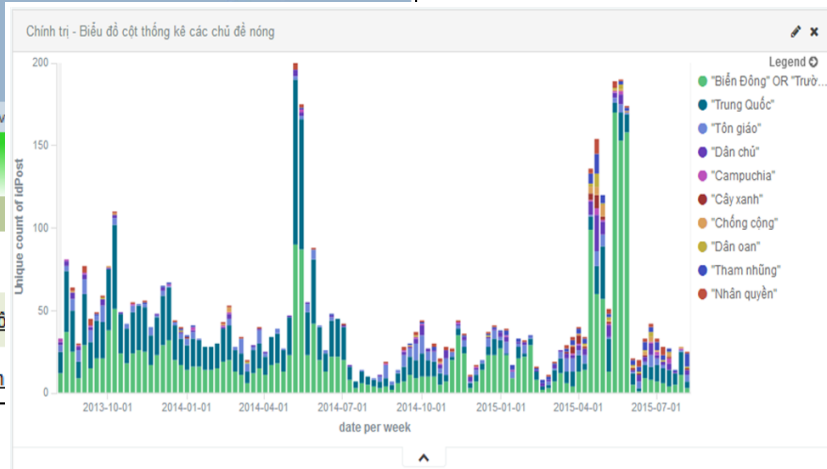
Đưa tin : **Bộ Thông Tin và Truyền Thông - Pháp luật**

Link : [Bình Dương: Dân quân xã về cơ đánh người gây thương tích!](#)

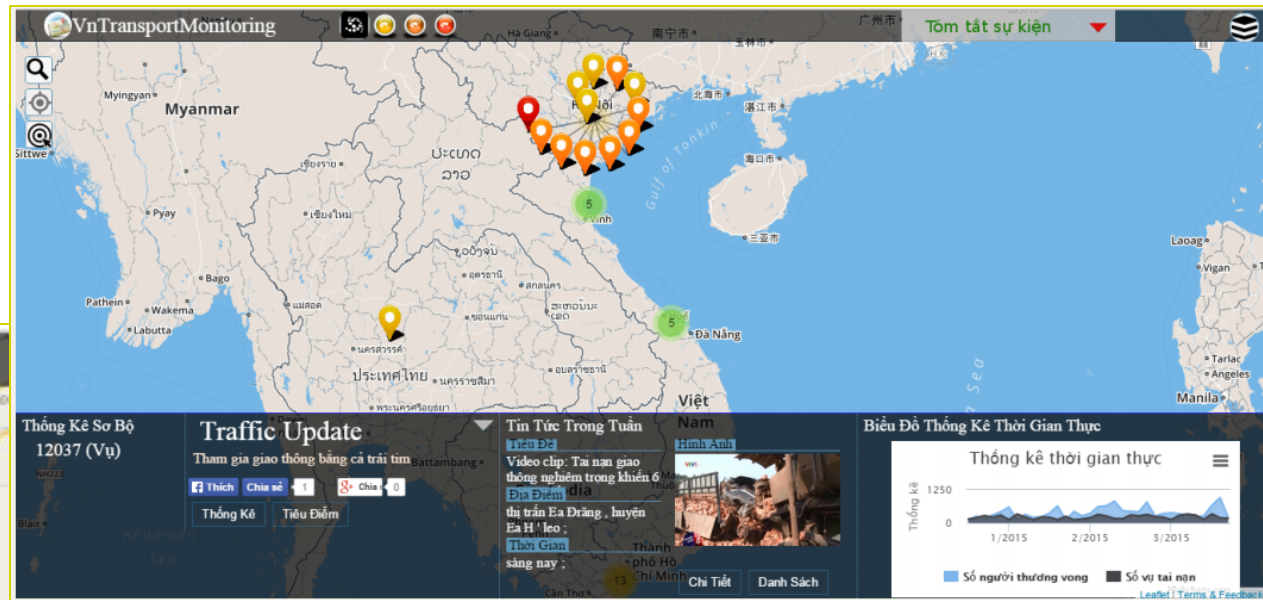


- Hỏa hoạn
- Tai nạn giao thông
- Hình sự
- Văn hóa
- Dịch bệnh
- Động đất

Thời gian	Sự kiện	Nội dung
Wed Oct 03 00:00:00 GMT+07:00 2012	Hình sự	Giết vợ chỉ vì ghen
Wed Oct 03 00:00:00 GMT+07:00 2012	Hình sự	Tìm nhân chứng vụ sinh viên bị sát hại trên xa lộ
Mon Nov 04 10:11:00 GMT+07:00 2013	Hình sự	Hôm nay xét xử kẻ đánh vợ dã man, bắt vợ xem



Hệ thống giám sát tai nạn



VnTransportMonitoring

Tom tắt sự kiện

Thống Kê Sơ Bộ: 12037 (Vụ)

Traffic Update
Tham gia giao thông bằng cả trái tim

Tin Tức Trong Tuần
Tên Đề
Video clip: Tai nạn giao thông nghiêm trọng khiến 6 người chết
Địa Điểm
thị trấn Ea Đrăng, huyện Ea H'leo;
Thời Gian
sáng nay;

Biểu Đồ Thống Kê Thời Gian Thực
Thống kê thời gian thực

Thống kê: 1250

1/2015 2/2015 3/2015

■ Số người thương vong ■ Số vụ tai nạn



VnTransportMonitoring

Thống Kê Sơ Bộ: 84.892 (Vụ)

Traffic Update
Tham gia giao thông bằng cả trái tim

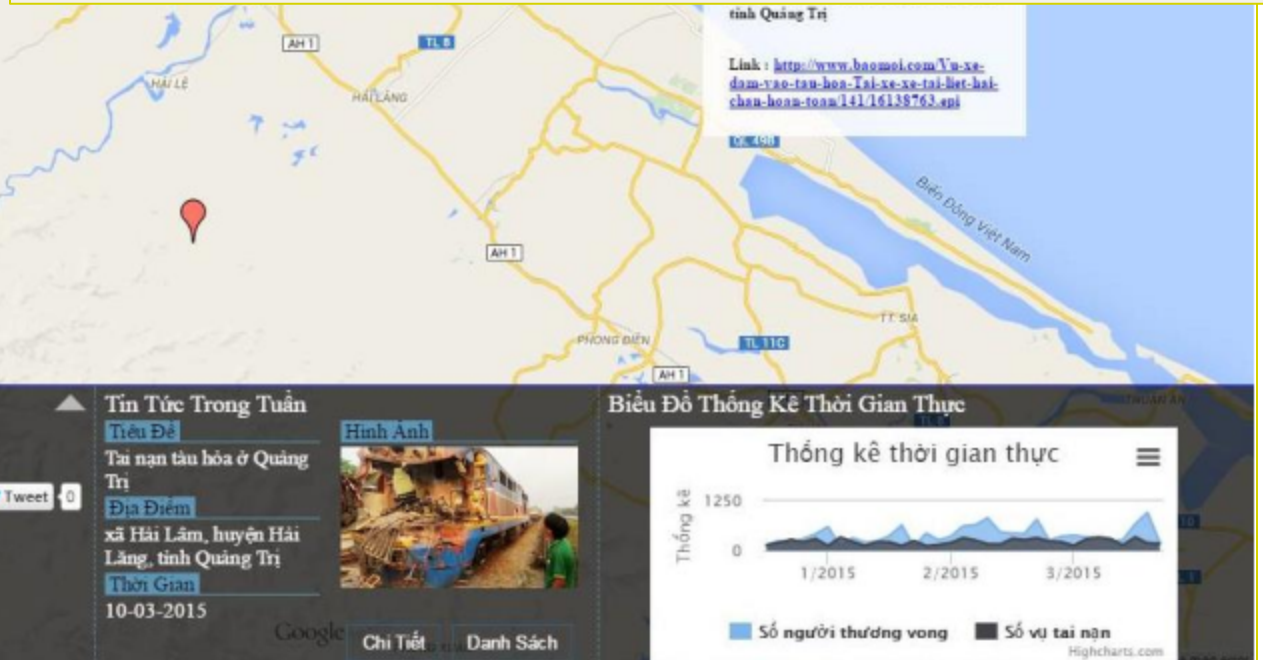
Tin Tức Trong Tuần
Tên Đề
Tai nạn tàu hỏa ở Quảng Trị
Địa Điểm
xã Hải Lâm, huyện Hải Lăng, tỉnh Quảng Trị
Thời Gian
10-03-2015

Biểu Đồ Thống Kê Thời Gian Thực
Thống kê thời gian thực

Thống kê: 1250

1/2015 2/2015 3/2015

■ Số người thương vong ■ Số vụ tai nạn



VnTransportMonitoring

Thống Kê Sơ Bộ: 84.892 (Vụ)

Traffic Update
Tham gia giao thông bằng cả trái tim

Tin Tức Trong Tuần
Tên Đề
Tai nạn tàu hỏa ở Quảng Trị
Địa Điểm
xã Hải Lâm, huyện Hải Lăng, tỉnh Quảng Trị
Thời Gian
10-03-2015

Biểu Đồ Thống Kê Thời Gian Thực
Thống kê thời gian thực

Thống kê: 1250

1/2015 2/2015 3/2015

■ Số người thương vong ■ Số vụ tai nạn



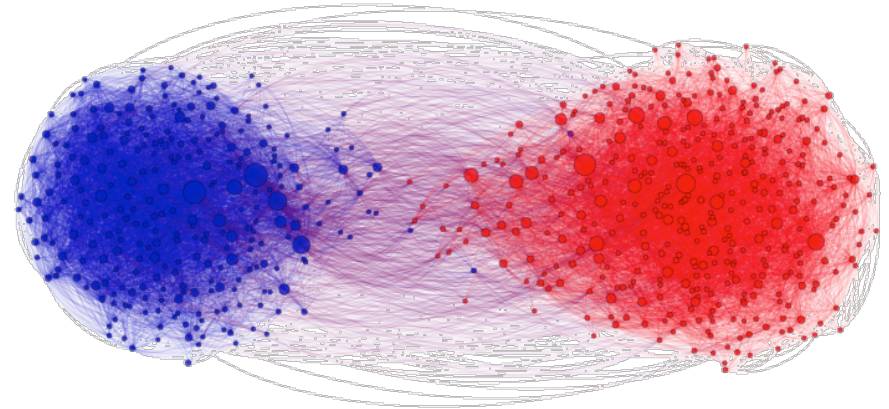
Phân tích cấu trúc mạng trực tuyến

Phát hiện cộng đồng



Phân tích cộng đồng giúp hiểu được người dùng

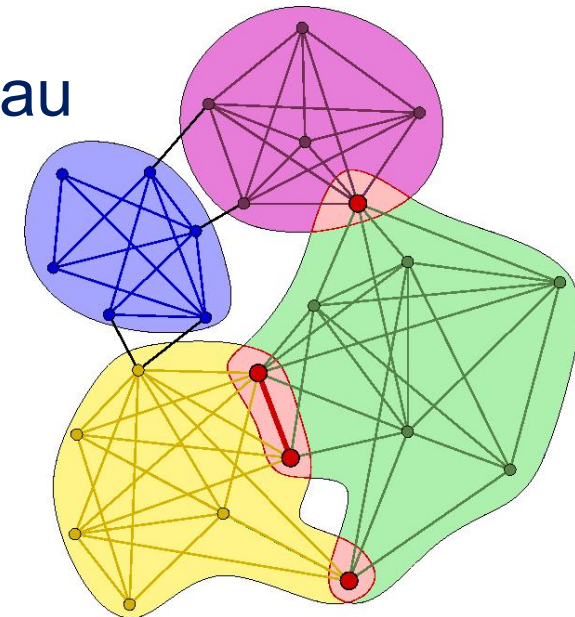
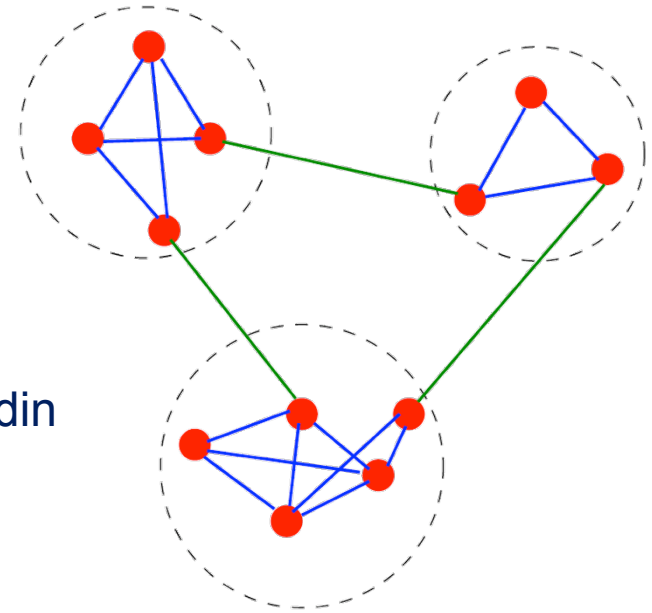
Phân tích cộng đồng đưa ra góc nhìn về sự tương tác của người dùng



Một số hành vi người dùng chỉ được phát hiện và quan sát khi họ tham gia cộng đồng

Phát hiện cộng đồng

- Cộng đồng trên MXH
 - Cộng đồng minh bạch (explicit)
 - Khai báo qua việc đăng ký rõ ràng
 - Các open/close group trên Facebook/LinkedIn
 - Cộng đồng ẩn (explicit)
 - Không được khai báo
 - Một số nhóm người thường tương tác trao đổi với nhau về chủ đề yêu thích
- Cộng đồng giao nhau và không giao nhau



Kỹ thuật phát hiện cộng đồng

- Một số kỹ thuật phổ biến
 - Phân cụm phân cấp
 - Hierarchical agglomerative clustering
 - Phân cụm theo đồ thị
 - Kernighan-Lin, Spectral bisection
 - Phân cụm phân hoạch
 - K-mean, Fuzzy K-mean
 - Phân cụm theo phổ
 - Spectral Clustering
 - Thuật toán phân chia
 - Girvan-Newman, Conga, Congo
- Công cụ
 - <http://www.cs.bris.ac.uk/~steve/networks/software/conga.html>
 - <http://perso.crans.org/aynaud/communities/index.html>

Cám ơn !
감사 !
ありがとう !
謝謝 !
Thank you!