

Báo cáo học thuật luận án lần 3
Phân tích mạng xã hội dựa theo mô hình chủ đề và ứng dụng

Người hướng dẫn khoa học: PGS.TS Đỗ Phúc

Nghiên cứu sinh: Hồ Trung Thành

Nội dung trình bày

- ▶ Tổng quan mục tiêu nghiên cứu và các bài toán của luận án
- ▶ Thử nghiệm mô hình đề xuất và đánh giá kết quả
- ▶ Những đóng góp của luận án
- ▶ Tài liệu tham khảo

Mô hình mạng xã hội



- ▶ Mạng xã hội được mô hình hóa bằng đồ thị $G(V,E)$ với V là tập các cá nhân (actor), E là tập các liên kết xã hội giữa các cá nhân
 - ▶ Từng cá nhân $v \in V$ có thể có các đặc trưng riêng
 - ▶ Từng liên kết $e \in E$ cũng có thể có các loại liên kết khác nhau như liên kết trao đổi thông tin, like, dislike, friend

Mục tiêu nghiên cứu của luận án

- ▶ Luận án tập trung vào 2 mục tiêu nghiên cứu:
 1. Phát triển mô hình khám phá vai trò của cá nhân trên MXH dựa theo mô hình chủ đề có yếu tố thời gian và khảo sát sự biến thiên chủ đề quan tâm của cá nhân.
 2. Đề xuất mô hình khám phá cộng đồng cá nhân trên MXH dựa theo mô hình chủ đề có yếu tố thời gian và khảo sát sự biến thiên những đặc trưng trong cộng đồng MXH.
- ▶ Các yếu tố trọng tâm trong luận án: thông điệp (chủ đề), vai trò (cá nhân, cộng đồng), thời gian

Đối tượng nghiên cứu của luận án

- ▶ Mô hình chủ đề
- ▶ Chủ đề và thông điệp được cá nhân quan tâm trao đổi trên MXH.
- ▶ Vai trò (cá nhân và cộng đồng): cá nhân là người gửi, người nhận thông điệp và cộng đồng là nhóm những cá nhân có cùng sự quan tâm trao đổi chủ đề.
- ▶ Thời gian cá nhân gửi và nhận thông điệp.

Phát triển mô hình khám phá vai trò của cá nhân trên MXH dựa theo mô hình chủ đề có yếu tố thời gian và khảo sát sự biến thiên chủ đề quan tâm của cá nhân

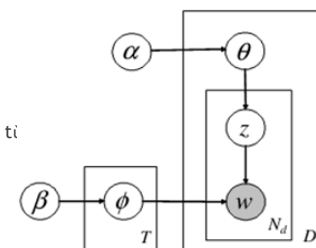
Mục tiêu nghiên cứu 1

Nhiệm vụ thực hiện trong mục tiêu nghiên cứu 1

1. Khám phá chủ đề tiềm ẩn từ thông điệp trên MXH và gán nhãn chủ đề dựa theo mô hình chủ đề
2. Phát triển mô hình TART (Temporal – Author – Recipient – Topic) khám phá actor vector chủ đề quan tâm của cá nhân trong V theo từng giai đoạn thời gian

1. Khám phá chủ đề tiềm ẩn và gán nhãn dựa theo mô hình chủ đề

- ▶ Tiếp cận mô hình chủ đề (Topic Modeling) LDA của David Blei (2003)
 - ▶ Là một dạng mô hình xác suất, mạng Bayes
 - ▶ Văn bản có nhiều chủ đề, một chủ đề được mô hình hóa thành phân bố xác suất trên tập từ
 - ▶ Dùng phương pháp Gibbs Sampling (lập cải thiện dần) để tìm ma trận $T \times W$ và $D \times T$
- ▶ Phát biểu bài toán của mô hình LDA
 - ▶ Cho
 - ▶ $D = \{d_1, d_2, \dots, d_n\}$ tập các văn bản trong kho ngữ liệu
 - ▶ $W = \{w_1, w_2, \dots, w_m\}$ tập các từ trong kho ngữ liệu D . Mỗi văn bản $d_i \in D$ chứa một tập từ của W
 - ▶ k tập các chủ đề cần khám phá
 - ▶ Tìm
 - ▶ Topic vector của k chủ đề ($T \times W$) ($\phi_{z,w}$)
 - ▶ Document vector của các văn bản ($D \times T$) ($\theta_{d,z}$)



Mô hình LDA
(Latent Dirichlet Allocation)

1. Khám phá chủ đề tiềm ẩn và gán nhãn dựa theo mô hình chủ đề

- ▶ Tiếp cận mô hình chủ đề (Topic Modeling) của David Blei (2003)
- ▶ **Topic vector** $v(T_j)$ là vector $\langle f_1(w_{i1}), f_1(w_{i2}), \dots, f_1(w_{im}) \rangle$ với thành phần thứ k ứng với từ thứ k (w_k) và $f_1(w_{ik})$ là tần số phản ánh xác suất xuất hiện của từ w_k trong chủ đề T_i .

- ▶ Topic vector $T \times W$ - $\phi_{z,w}$

Ví dụ: $V(T1) = \langle 0.012, 0.034, \dots, 0.067 \rangle$

$V(T2) = \langle 0.017, 0.041, \dots, 0.053 \rangle$

Theo Topic Modeling của David Blei, đây là ma trận $T \times W$, mỗi dòng của ma trận $T \times W$ là một Topic vector.

$$\phi_{z,w} = \frac{n_w^{(z)} + \beta_w}{\sum_{w' \in V} n_{w'}^{(z)} + \beta_{w'}}, z \in Z, w \in V.$$

1. Khám phá chủ đề tiềm ẩn và gán nhãn dựa theo mô hình chủ đề

- ▶ Tiếp cận mô hình chủ đề (Topic Modeling) của David Blei (2003)
- ▶ **Document vector** $d_i \in D$ là vector $v(d_i) = \langle f_2(T_{i1}), f_2(T_{i2}), \dots, f_2(T_{ik}) \rangle$

với thành phần thứ j là chủ đề T_j trong văn bản, $f_2(T_{ij})$ là con số phản ánh xác suất xuất hiện chủ đề T_j trong văn bản d_i .

- ▶ Documents vector $D \times T$ - $\theta_{d,z}$

Ví dụ: $V(d1) = \langle 0.72, 0.14, \dots, 0.23 \rangle$

$V(d2) = \langle 0.19, 0.32, \dots, 0.41 \rangle$

Theo Topic Modeling của David Blei đây là ma trận $D \times T$, mỗi dòng của ma trận $D \times T$ là một Document vector.

$$\theta_{d,z} = \frac{n_z^{(d)} + \alpha_z}{\sum_{z' \in Z} n_{z'}^{(d)} + \alpha_{z'}}, d \in D, z \in Z.$$

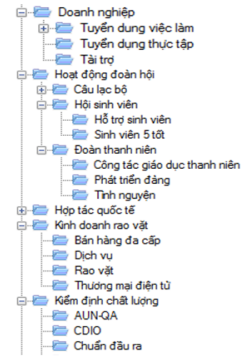
1. Khám phá chủ đề tiềm ẩn và gán nhãn chủ đề dựa theo mô hình chủ đề

► Giới hạn mô hình LDA

► Chỉ tìm ra chủ đề thông qua tập từ đặc trưng mô tả chủ đề nhưng chưa cho biết khái niệm chủ đề cụ thể được tìm ra là gì?

P (từ / chủ đề)

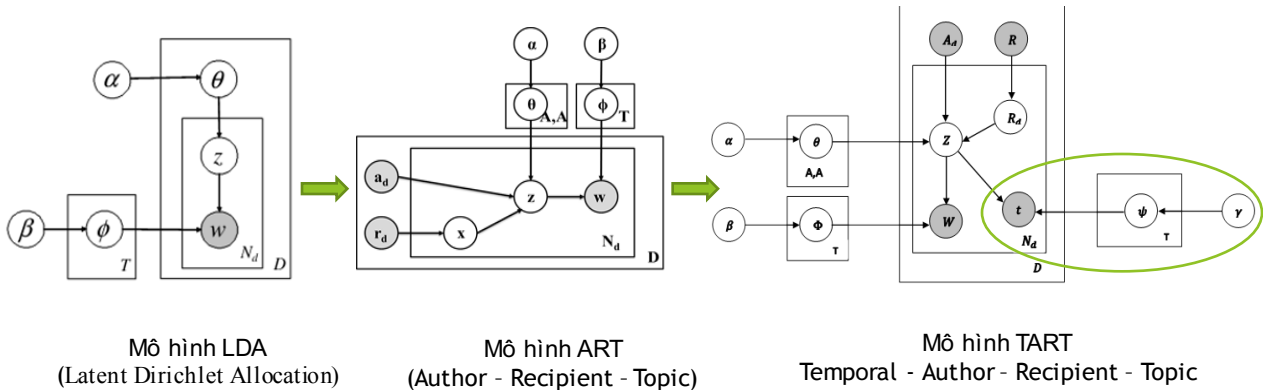
Chủ đề 1		Chủ đề 2		Chủ đề 3		Chủ đề 4	
Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất
công_tác	0.01197	đại_học	0.01306	fampage	0.01339	Đoàn	0.02464
chúng_ta	0.01051	học_thuật	0.01259	L_ê_nin	0.01292	hoạt_động	0.02261
qui_dịnh	0.01051	tham_gia	0.01053	tư_tướng	0.01194	mọi_người	0.02002
hè	0.00903	học	0.00969	phòng	0.01072	thanh_niên	0.01712
năm_học	0.0086	thê_thao	0.00928	olympic	0.01069	sinh_viên	0.01459
học	0.00827	sinh_viên	0.00905	Mác	0.01046	đại_viên	0.01346
sinh_viên	0.00631	tinh_thần	0.00818	hội_thi	0.01003	học_sinh	0.01313
Bác	0.00616	kỹ_năng	0.00774	chung_kết	0.00844	Bác	0.01274
Đoàn	0.00614	đại_học	0.00644	thời_đại	0.00773	chương_trình	0.01188
hoạt_động	0.00579	thi	0.00633	công_việc	0.00752	kỹ_năng	0.01172
vui	0.00543	giáo_tiếp	0.00581	sinh_viên	0.00725	tinh_nguyên	0.01165
tinh_nguyên	0.00524	đạt	0.00482	ảnh_sáng	0.00722	chuyển	0.01122
hội	0.0051	ngoại_ngữ	0.00419	chúng_tôi	0.00614	ĐHQG	0.01025
phong_trào	0.00417	chương_trình	0.00419	bản_lĩnh	0.00559	an_toàn	0.00772
đoàn_viên	0.00373	hoạt_động	0.00414	hành_trình	0.00515	tổ_chức	0.00767



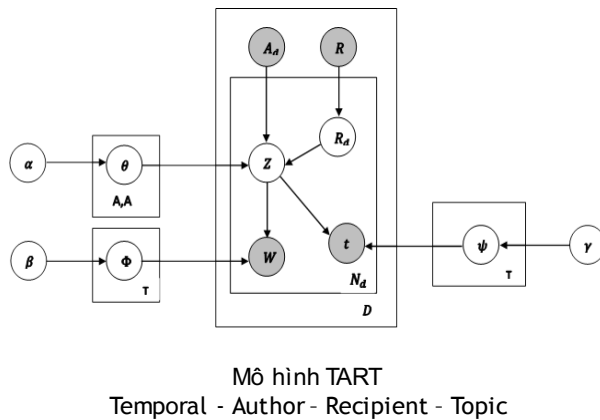
► Các công việc cần thực hiện:

- Gán nhãn chủ đề để biết các thông điệp trao đổi nội dung gì?
 - xây dựng hệ thống cây khái niệm chủ đề
 - phân lớp chủ đề dựa trên cây khái niệm chủ đề

2. Phát triển mô hình TART



2. Phát triển mô hình TART



- ▶ Nhiệm vụ của mô hình TART
 - ▶ Phân tích vai trò của cá nhân tham gia mạng xã hội dựa theo mô hình chủ đề có yếu tố thời gian
 - ▶ Dùng yếu tố thời gian để chia nhỏ các yếu tố trong mô hình ART như tập cá nhân gửi, tập cá nhân nhận, tập chủ đề
 - ▶ Tìm ra chủ đề quan tâm của cá nhân (actor vector) theo từng giai đoạn thời gian
 - ▶ Khảo sát sự biến thiên chủ đề quan tâm của từng cá nhân.

2. Phát triển mô hình TART

Phát biểu bài toán của mô hình TART

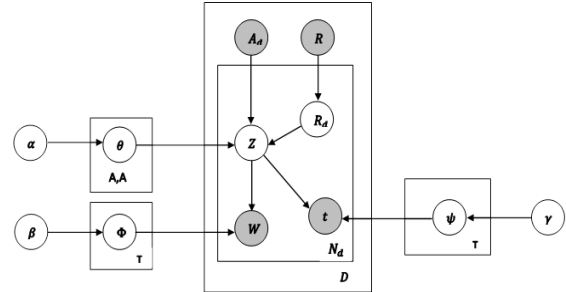
- ▶ Cho mạng xã hội $G(V,E)$, trong đó tập V là tập các cá nhân, tập E là tập các thông điệp trao đổi giữa các actor. Gọi D là tập các thông điệp trao đổi trên mạng xã hội, T là các chủ đề trao đổi trong các thông điệp.
- ▶ Mỗi actor $v_i \in V$ được đặc trưng bằng actor vector gồm $\langle v_{i1}, v_{i2}, \dots, v_{ik} \rangle$, trong đó thành phần thứ k phản ánh xác suất trao đổi chủ đề T_k của actor v_i .

→ Tìm actor vector chủ đề quan tâm của những cá nhân trong V theo từng giai đoạn thời gian

2. Phát triển mô hình TART

Phát biểu bài toán của mô hình TART

- ▶ Cho:
 - ▶ Tập thông điệp trao đổi
 - ▶ Tập cá nhân trên MXH
 - ▶ Số chủ đề K cần khám phá từ tập thông điệp
- ▶ Tìm:
 - ▶ actor vector chủ đề quan tâm của những cá nhân trong V theo từng giai đoạn thời gian
 - ▶ Ví dụ: Cá nhân có ID =025, với 3 chủ đề quan tâm T1, T2 và T3. Kết quả như sau:
 - ▶ Thời gian = 01-2009, vector chủ đề quan tâm <0.012, 0.044, 0.606>
 - ▶ Thời gian = 02-2009, vector chủ đề quan tâm <0.082, 0.004, 0.516>
 - ▶ Thời gian = 03-2009, vector chủ đề quan tâm <0.012, 0.044, 0.452>



Mô hình TART
Temporal - Author - Recipient - Topic

2. Phát triển mô hình TART

- ▶ Thực hiện mô hình TART tạo 4 ma trận kết quả :
 - ▶ T (chủ đề) x W (từ) và A (tác giả) x T (chủ đề),
 - ▶ R (người nhận) x T (chủ đề) và T (chủ đề) x T (thời gian)
- ▶ Dựa trên 4 ma trận, ta có phân bố chủ đề và từ ϕ_{zw} , phân bố chủ đề và thời gian ψ_{zt} , phân bố cá nhân gửi và chủ đề θ_{az} , phân bố cá nhân nhận và chủ đề θ_{rz} .

$$\theta_{az} = \frac{m_{az} + \alpha}{\sum_z (m_{az} + \alpha)}$$

$$\phi_{zw} = \frac{n_{zw} + \beta}{\sum_w (n_{zw} + \beta)}$$

$$\psi_{zt} = \frac{m_{zt} + \gamma}{\sum_t (m_{zt} + \gamma)}$$

$$\theta_{rz} = \frac{m_{rz} + \alpha}{\sum_z (m_{rz} + \alpha)}$$

2. Phát triển mô hình TART

Kết quả mô hình TART: các vector chủ đề quan tâm của cá nhân (actor vector) qua thời gian

Cá nhân =025, với 3 chủ đề quan tâm T1, T2 và T3:

Thời gian = 1-2009, actor vector <0.012,0.044,0.606>

Thời gian = 2-2009, actor vector <0.082,0.004,0.516>

Thời gian = 3-2009, actor vector <0.012,0.044,0.452>

Cá nhân =026, với 3 chủ đề quan tâm T1, T2 và T3:

Thời gian = 1-2009, actor vector <0.072,0.004,0.016>

Thời gian = 2-2009, actor vector <0.082,0.054,0.018>

Thời gian = 3-2009, actor vector <0.076,0.027,0.506>

Đóng góp của mục tiêu nghiên cứu 1

- ▶ Xây dựng cây khái niệm chủ đề
- ▶ Khám phá chủ đề tiềm ẩn từ thông điệp và gán nhãn chủ đề
- ▶ Phát triển mô hình TART để khám phá vai trò của cá nhân trên MXH có yếu tố thời gian
- ▶ Đề xuất phương pháp khảo sát sự biến thiên chủ đề quan tâm của cá nhân trên MXH theo thời gian

Đề xuất mô hình khám phá cộng đồng cá nhân trên MXH dựa theo mô hình chủ đề có yếu tố thời gian và khảo sát sự biến thiên những đặc trưng trong cộng đồng MXH

Mục tiêu nghiên cứu 2

Cộng đồng mạng xã hội



Cộng đồng có liên hệ trong MXH



Các cộng đồng rời rạc trong MXH

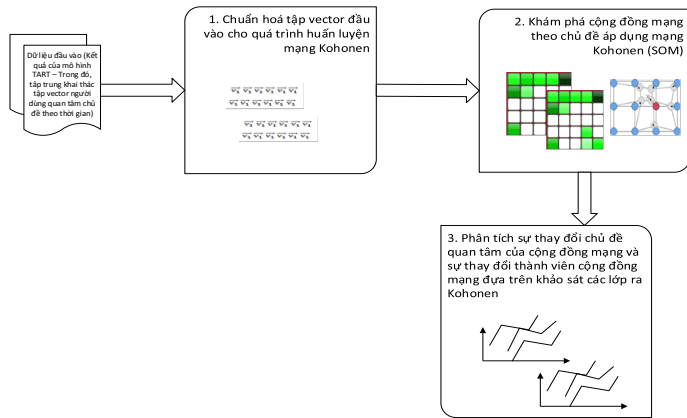
Tiếp cận dựa theo mô hình chủ đề, cộng đồng là tập hợp các cá nhân cùng quan tâm đến các chủ đề. Mỗi cá nhân được đặc trưng bằng một vector chủ đề quan tâm và mức độ cùng quan tâm đến chủ đề trong cộng đồng nhiều hơn so với những cộng đồng khác

Cho c là một cụm (cộng đồng) theo chủ đề, $c \in C$, trong đó C là tập hợp các cộng đồng. Tập hợp các cộng đồng trên mạng được ký hiệu $C = \{C_1, C_2, C_3, C_4, \dots, C_k\}$ với k là số cụm, mỗi cụm C_i có vector chủ đề.

Trong đó, các cụm có tính chất:

- (1) Rời rạc nhau: $C_i \cap C_j = \emptyset$ nếu các cụm không cùng quan tâm đến một hay nhiều chủ đề cụ thể nào đó
- (2) Và hợp các cụm $\bigcup_{i=1}^k C_i = C$.

Mô hình khám phá cộng đồng cá nhân trên MXH dựa theo mô hình chủ đề



Mô hình khám phá cộng đồng cá nhân theo chủ đề và khảo sát sự biến thiên chủ đề quan tâm của cá nhân

Nhiệm vụ thực hiện của mô hình

- ▶ **Chuẩn hoá vector nhập:** chuẩn hoá dữ liệu đầu vào phù hợp với dữ liệu huấn luyện của mạng Kohonen.
- ▶ **Khám phá cộng đồng sử dụng mạng Kohonen (SOM):** áp dụng kỹ thuật SOM để gom cụm cá nhân theo chủ đề quan tâm, mỗi cụm là một cộng đồng theo chủ đề và tương ứng với 1 neuron tại lớp ra Kohonen.
- ▶ **Khảo sát sự biến thiên thành viên và chủ đề quan tâm của cộng đồng:** dựa trên lớp ra Kohonen phân tích sự biến thiên thành viên cộng đồng và chủ đề mà cộng đồng quan tâm theo từng giai đoạn thời gian.

Mô hình khám phá cộng đồng cá nhân trên MXH dựa theo mô hình chủ đề

Phát biểu bài toán của mô hình khám phá cộng đồng

- ▶ Cho đồ thị $G=(V,E)$, trong đó tập V là tập các cá nhân, tập E là tập các thông điệp trao đổi giữa các cá nhân và cho T là tập chủ đề, v là tập vector chủ đề quan tâm của cá nhân, tìm cộng đồng cá nhân có cùng chủ đề quan tâm theo từng giai đoạn thời gian.

- ▶ Đầu vào: tập vector chủ đề quan tâm của cá nhân theo từng giai đoạn thời gian

- ▶ Ví dụ: có 4 actor vector, mỗi vector gồm 7 thành phần tương ứng xác suất quan tâm 7 chủ đề

	T-0	T-1	T-2	T-3	T-4	T-5	T-6	Cá nhân
\vec{v}_1	(0.64444, 0.34545, 0.46826, 0.0, 0.33721, 0.0, 0.0)							Mr.tajkjd
\vec{v}_2	(0.30435, 0.44565, 0.33333, 0.30435, 0.33333, 0.52941, 0.0)							dsvantan
\vec{v}_3	(0.39601, 0.48718, 0.0, 0.35484, 0.0, 0.38462, 0.0)							nguyen.nhi.334491
\vec{v}_4	(0.34694, 0.40741, 0.0, 0.39227, 0.0, 0.36, 0.0)							trang.harry.7

- ▶ Đầu ra: danh sách các cụm (cộng đồng) theo từng giai đoạn thời gian

- ▶ $C = \{C_1, C_2, C_3, C_4, \dots, C_k\}$ với k là số cụm, mỗi cụm C_i có vector chủ đề

- ▶ Trong đó, các cụm có tính chất:

- ▶ Rời rạc nhau: $C_i \cap C_j = \emptyset$ nếu các cụm không cùng quan tâm đến một hay nhiều chủ đề cụ thể nào đó.

- ▶ Và hợp các cụm $\bigcup_{i=1}^k C_i = C$.

Khảo sát sự biến thiên chủ đề quan tâm của cộng đồng theo thời gian

Phát biểu bài toán của mô hình khám phá cộng đồng

- ▶ Sự biến thiên chủ đề quan tâm của cộng đồng c được biết dựa trên tần suất thay đổi chủ đề quan tâm của tập cá nhân u trong cộng đồng. Ký hiệu là $Z(c, t, N_z)$.
 - ▶ Mỗi cộng đồng quan tâm đến nhiều chủ đề trong giai đoạn thời gian
 - ▶ Sự thay đổi này thể hiện qua sự khác nhau giữa thành phần của hai tập chủ đề tại thời điểm $t - 1$ là $Z(c, t-1, N_z)$ và tại thời điểm t là $Z(c, t, N_z)$ mà cộng đồng quan tâm chủ đề
- đề xuất độ đo $\partial_\pi(c, t - 1, t, N_z)$ để đo lường mức độ biến thiên chủ đề quan tâm z của cộng đồng c tại thời điểm t
- ▶ Độ đo này là tỉ lệ giữa: hiệu số (số chủ đề N_z và phần giao giữa số chủ đề trong cộng đồng tại thời điểm $t-1$ với tập số chủ đề trong cộng đồng tại thời điểm t) chia cho số chủ đề N_z , giá trị của $\partial_\pi(c, t - 1, t, N_z)$ nằm trong khoảng $[0, 1]$.
 - ▶ Bằng 1 thì tập N_z thường xuyên được thay đổi bởi cộng đồng c
 - ▶ Bằng 0 là không có chủ đề nào thay đổi trong cộng đồng c trong khoảng thời gian từ $t-1$ đến t

$$\partial_\pi(c, t - 1, t, N_z) = \frac{N_z - |Z(c, t - 1, N_z) \cap Z(c, t, N_z)|}{N_z} \in [0, 1]$$

Đóng góp của mục tiêu nghiên cứu 2

- ▶ Đề xuất mô hình khám phá cộng đồng cá nhân có yếu tố thời gian
- ▶ Đề xuất phương pháp khảo sát sự biến thiên đặc trưng của cộng đồng trên MXH theo thời gian

Thử nghiệm và đánh giá kết quả

Mục tiêu nghiên cứu 1

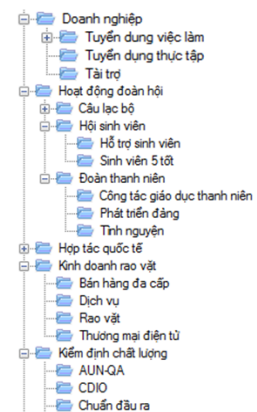
Mô tả dữ liệu thử nghiệm mô hình TART



Hình 2. Nội dung trao đổi của cá nhân trên MXH

Bảng 1. Bảng thống kê số thông điệp rút trích từ diễn đàn và MXH trong giáo dục

Nguồn	Số thông điệp
Diễn đàn ĐHQG-HCM	13,208
Fanpage của UEL	16,917
Fanpage của HCMUS	1,159
Nhóm CLB của UEL	42,702
Fanpage của UIT	10,491
Nguồn khác từ Facebook	22,483

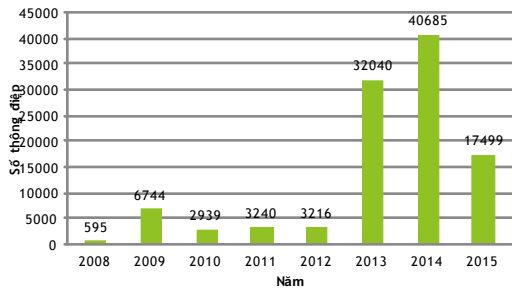


Hình 3. Một phần cây khái niệm chủ đề phân cấp trong lĩnh vực giáo dục đại học

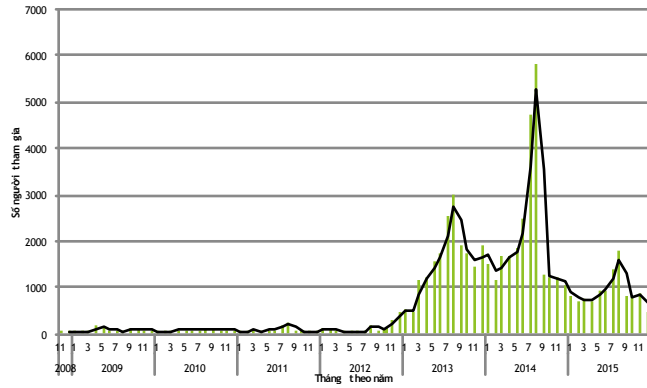
Mô tả dữ liệu thử nghiệm mô hình TART

► Dữ liệu 1. Nội dung trên diễn đàn – MXH Facebook

- 106.960 thông điệp trao đổi
- 35.056 tài khoản cá nhân



Hình 4. Trình bày lịch sử thông điệp được gửi theo từng năm trong tập dữ liệu thu thập



Hình 5. Trình bày lịch sử số cá nhân tham gia gửi thông điệp theo từng tháng và năm từ năm 2008 đến năm 2015.

Mô tả dữ liệu thử nghiệm mô hình TART

Giáo dục

Công Vinh có bất ngờ luật với 21 điểm ☹️/!

Được 21 điểm cho ba môn Văn, Sử, Địa, Công Vinh có điểm cao thứ ba trong số thí sinh dự thi vào hệ thi chức Đại học Luật Hà Nội...

60 từ viết tắt tiếng Anh giới trẻ chuộng nhất

Thành Hóa sẽ giảm hơn 100 trường học

Bộ đại học sau 3 năm sẽ theo ngành minh thích có quá muộn

Hình 6. Nội dung bài viết và trao đổi của nhân trên trang báo điện tử vnExpress.net

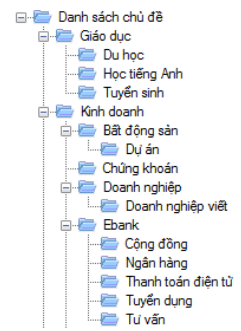
Lan Hạ

Công Vinh có bất ngờ luật với 21 điểm

Được 21 điểm cho ba môn Văn, Sử, Địa, Công Vinh có điểm cao thứ ba trong số thí sinh dự thi vào hệ thi chức Đại học Luật Hà Nội...

Trước đó ngày 29/11, nhiều thí sinh dự thi tại chức Đại học Luật Hà Nội đã ngỡ ngàng khi biết rằng Công Vinh có một tài năng thi và ngày làm bài nghiêm túc, kết thúc buổi thi, Công Vinh vui vẻ chụp ảnh lưu niệm khi một số giám thị thì thầm nói kín. Chúng tôi hỏi người Nghệ An công đồng cấp cho quý từ thên của sinh viên đồng hương Nghệ An - Hà Tĩnh tại Đại học Luật Hà Nội.

Các cựu cầu thủ như Nguyễn Hồng Sơn, Triệu Quang Hà từng thi đỗ và học tại Đại học Luật Hà Nội.

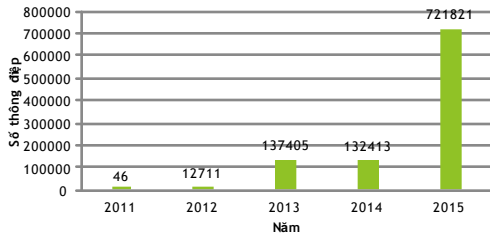


Hình 7. Một phần cây khái niệm chủ đề phân cấp trên trang báo điện tử vnExpress.net

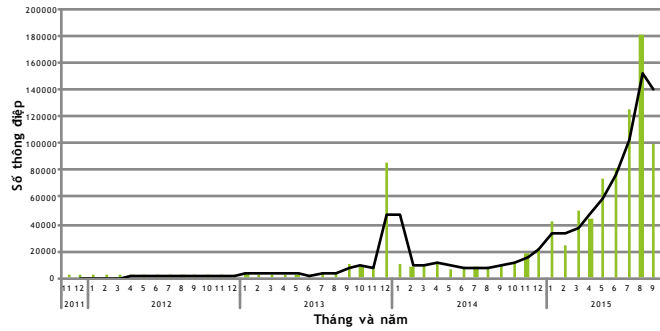
Mô tả dữ liệu thử nghiệm mô hình TART

► Dữ liệu 2. Nội dung trang báo điện tử vnExpress.net

- 1.004.396 thông điệp
- 225.009 tài khoản cá nhân



Hình 8. Trình bày lịch sử thông điệp được gửi theo từng năm trong tập ngữ liệu thu thập



Hình 9. Trình bày lịch sử số cá nhân tham gia gửi thông điệp theo từng tháng và năm từ năm 2011 đến năm 2015.

Kết quả thử nghiệm mô hình TART– Dữ liệu 1

Bảng 2. Trình bày kết quả phân tích 4 chủ đề quan tâm của cá nhân trong giai đoạn tháng 08-2014.

Cấp 0

Chủ đề 1: Hoạt động đoàn hội		Cấp 1		Cấp 1		Cấp 1	
Từ	Xác suất	Chủ đề 2: Câu lạc bộ		Chủ đề 3: Hội sinh viên		Chủ đề 4: Đoàn thanh niên	
Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất
công_tác	0.01197	Đại_học	0.01306	fanpage	0.01339	Đoàn	0.02464
chúng_ta	0.01051	học_thuật	0.01259	Lê_nin	0.01292	hoạt_động	0.02261
qui_dinh	0.01051	tham_gia	0.01053	tu_tuong	0.01194	mọi_người	0.02002
hè	0.00903	học	0.00969	phòng	0.01072	thanh_niên	0.01712
năm_học	0.00860	thể_thao	0.00928	Olympic	0.01069	sinh_viên	0.01459
học	0.00827	Bác	0.00905	Mác	0.01046	đại_hội	0.01346
sinh_viên	0.00631	Đoàn	0.00614	hội_thi	0.01003	học_sinh	0.01313
Bác	0.00616	hoạt_động	0.00579	chung_kết	0.00844	Bác	0.01274
Đoàn	0.00614	vui	0.00543	thời_đại	0.00773	chương_trình	0.01188
hoạt_động	0.00579	tình_nguyện	0.00524	công_việc	0.00752	kỹ_năng	0.01172
hội	0.00510	hội	0.00510	vòng	0.00725	tình_nguyện	0.01165
phong_trào	0.00417	giáo_tiếp	0.00581	ánh_sáng	0.00722	chuyển	0.01122
đoàn_viên	0.00373	đạt	0.00482	chúng_tôi	0.00614	ĐHQG	0.01025
Cá nhân gửi	Xác suất	ngoại_ngữ	0.00419	bản_lĩnh	0.00559	an_toàn	0.00772
97179	0.670330	chương_trình	0.00419	hành_trình	0.00515	tổ_chức	0.00767
97568	0.600000	hoạt_động	0.00414	Cá nhân gửi	Xác suất	Cá nhân gửi	Xác suất
		Cá nhân gửi	Xác suất				

71804	0.560976	78686	0.53982	79554	0.68212	67484	0.83740
79570	0.508197	72421	0.44000	78334	0.49640	69151	0.83529
71255	0.500864	79179	0.42222	64383	0.48148	67815	0.83133
77683	0.485714	64389	0.40741	80121	0.45946	68906	0.79221
64280	0.481481	70893	0.39130	71129	0.45455	68359	0.78462
79458	0.478261	73772	0.39130	69662	0.44000	68120	0.77778
97126	0.471698	79249	0.38777	71151	0.39683	70824	0.77049
74568	0.469388	79096	0.37143	64325	0.39130	68395	0.75439
Cá nhân nhận	Xác suất	79556	0.36000	72750	0.37931	69361	0.75385
97126	0.670330	69660	0.33333	64374	0.36585	68925	0.74545
77692	0.560976	Cá nhân nhận	Xác suất	Cá nhân nhận	Xác suất	Cá nhân nhận	Xác suất
81027	0.548387	72365	0.44000	90018	0.68212	64595	0.84946
67317	0.538462	90191	0.42222	73490	0.48148	72692	0.83133
76996	0.485714	72597	0.40741	96166	0.45946	71138	0.79221
78777	0.478261	90195	0.37143	74276	0.45455	75224	0.77379
97179	0.471698	90082	0.36000	70498	0.39406	68845	0.77049
84270	0.469421	76609	0.33333	64364	0.39130	77144	0.75439
73057	0.440000	66927	0.32435	80121	0.38190	77680	0.75439
74260	0.438202	67473	0.31429	71039	0.37931	72238	0.75385
		71955	0.31034	73376	0.36000	64864	0.74545
		74183	0.31212	76427	0.35484	76590	0.73585

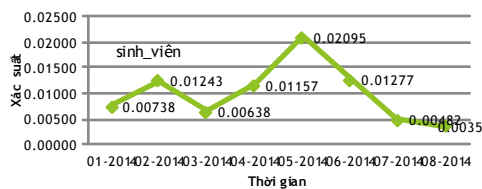
Kết quả thử nghiệm mô hình TART – Dữ liệu 1

Bảng 3. Trình bày kết quả phân tích chủ đề “Giáo dục” được cá nhân quan tâm trong giai đoạn tháng 01-2014 đến tháng 07-2014

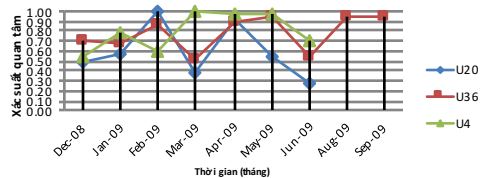
01-2014		02-2014		03-2014		04-2014		05-2014		06-2014		07-2014	
Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất
người	0.05745	tuổi_tho	0.06570	ngày_hồi	0.01252	bun	0.02095	sinh_viên	0.02095	phòng	0.02095	người	0.01746
thông_tin	0.01695	đen	0.01759	quản_tri	0.00858	minh	0.01584	một	0.01584	chỗ	0.01584	thứ_sinh	0.01728
chi_tiết	0.01506	tết	0.01523	báo_cáo	0.00776	em	0.01470	anh	0.01470	hỗ_trợ	0.01470	thông_tin	0.01502
đầu_tiên	0.01221	sinh_viên	0.01243	cắm_trai	0.00752	nhiều	0.01277	đội	0.01277	sinh_viên	0.01277	giáo_dục	0.01419
kinh_nghiệm	0.01034	xe	0.01035	Bước	0.00655	sinh_viên	0.01157	tư_vấn	0.01157	tiếp_bước	0.01157	đầu_tiên	0.01287
website	0.00871	hỏi	0.00869	sinh_viên	0.00638	hon	0.00773	chức	0.00773	hôm_nay	0.00773	kinh_nghiệm	0.00668
quy_trình	0.00801	Bơm	0.00782	hiện_tại	0.00614	món	0.00757	chương_trình	0.00757	bản_lĩnh	0.00757	bộ	0.00640
tháp	0.00783	đời	0.00765	chương_trình	0.00590	ngày	0.00742	tài_liệu	0.00742	mùa	0.00742	quy_trình	0.00630
sinh_viên	0.00738	li_xì	0.00716	khoảng_cách	0.00549	anh	0.00722	biên_bản	0.00722	thời_dại	0.00722	tháp	0.00620
hiện_tại	0.00693	ngon	0.00679	ĐHQG-HCM	0.00541	phòng	0.00645	chữ	0.00645	qui_dịnh	0.00645	học_sinh	0.00540
chuyên	0.00640	xuân_tinh_nguyên	0.00617	chúng_ta	0.00537	chương_trình	0.00616	kiến_thúc	0.00616	tình_nguyên	0.00616	hiện_tại	0.00500
Infographic	0.00617	cơ_bản	0.00596	sinh_hoạt	0.00525	kinh_tế	0.00597	điều_kiện	0.00597	bình_trình	0.00597	thì	0.00494
chính_sách	0.00595	chương_trình	0.00575	văn_bản	0.00525	Khoa	0.00582	không_biết	0.00582	fanpage	0.00582	qui_chế	0.00486
qui_chế	0.00583	thứ_sinh	0.00564	dại_diện	0.00520	lầu	0.00571	học_sinh	0.00571	thì	0.00571	sinh_viên	0.00482
công_khai	0.00550	xuân	0.00529	chỉ_hướng	0.00508	Lượt	0.00541	dại_học	0.00541	tro	0.00541	hỗ_trợ	0.00448
Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất
70464	0.85333	30257	0.93033	156036	0.88435	40957	0.85449	19626	0.93322	19626	0.93322	19626	0.95767
14652	0.82861	43941	0.93690	19325	0.86400	129579	0.85373	96561	0.94327	96561	0.94327	96561	0.95767
178518	0.71901	159072	0.93694	70262	0.85695	26286	0.82164	11028	0.93992	11028	0.93992	11028	0.95429
86559	0.71055	18960	0.93300	144011	0.85326	47736	0.79971	48455	0.94294	48455	0.94294	48455	0.95429
1182	0.68768	69993	0.93061	155430	0.85423	16264	0.79850	11050	0.94005	11050	0.94005	11050	0.95031
126092	0.65927	19702	0.92513	103628	0.85407	120626	0.79417	66030	0.94152	66030	0.94152	66030	0.95031
87327	0.65276	96714	0.85276	158531	0.85171	102482	0.77450	66252	0.93140	66252	0.93140	66252	0.94015
149073	0.65237	142267	0.79501	50733	0.84914	180396	0.76473	105841	0.92092	105841	0.92092	105841	0.92945
143228	0.64043	60976	0.65372	89787	0.84009	207296	0.73117	88967	0.85842	88967	0.85842	88967	0.86667
82496	0.63068	139920	0.59907	68962	0.83437	140619	0.72340	83619	0.85847	83619	0.85847	83619	0.86596

14652	0.82861	43941	0.93690	19325	0.86400	129579	0.85373	96561	0.94327	96561	0.94327	96561	0.95767
178518	0.71901	159072	0.93694	70262	0.85695	26286	0.82164	11028	0.93992	11028	0.93992	11028	0.95429
86559	0.71055	18960	0.93300	144011	0.85326	47736	0.79971	48455	0.94294	48455	0.94294	48455	0.95429
1182	0.68768	69993	0.93061	155430	0.85423	16264	0.79850	11050	0.94005	11050	0.94005	11050	0.95031
126092	0.65927	19702	0.92513	103628	0.85407	120626	0.79417	66030	0.94152	66030	0.94152	66030	0.95031
87327	0.65276	96714	0.85276	158531	0.85171	102482	0.77450	66252	0.93140	66252	0.93140	66252	0.94015
149073	0.65237	142267	0.79501	50733	0.84914	180396	0.76473	105841	0.92092	105841	0.92092	105841	0.92945
143228	0.64043	60976	0.65372	89787	0.84009	207296	0.73117	88967	0.85842	88967	0.85842	88967	0.86667
82496	0.63068	139920	0.59907	68962	0.83437	140619	0.72340	83619	0.85847	83619	0.85847	83619	0.86596
Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất
13426	0.70492	105830	0.89325	104116	0.88435	139366	0.84798	111253	0.91411	111253	0.91411	111253	0.91411
184607	0.70464	3836	0.64757	46167	0.86400	147443	0.80751	137588	0.88585	137588	0.88585	137588	0.88585
134347	0.69921	5253	0.46976	144665	0.85024	135408	0.78763	87188	0.88588	87188	0.88588	87188	0.88588
73649	0.69921	150587	0.40684	101460	0.85151	57198	0.78461	124448	0.84405	124448	0.84405	124448	0.84405
60406	0.65049	102851	0.39399	141234	0.85020	104265	0.77792	40329	0.84513	40329	0.84513	40329	0.84513
15949	0.64372	79827	0.38656	45740	0.84889	103255	0.77150	105677	0.65879	105677	0.65879	105677	0.65879
13771	0.64318	72037	0.37320	533	0.83959	207331	0.77146	25468	0.50521	25468	0.50521	25468	0.50521
112405	0.58252	105924	0.36401	30875	0.83332	115263	0.77082	18507	0.44825	18507	0.44825	18507	0.44825
88862	0.57282	39881	0.36039	116984	0.83037	39594	0.77053	41288	0.38501	41288	0.38501	41288	0.38501
49351	0.52239	21826	0.35836	66976	0.81707	76393	0.76519	53462	0.38376	53462	0.38376	53462	0.38376

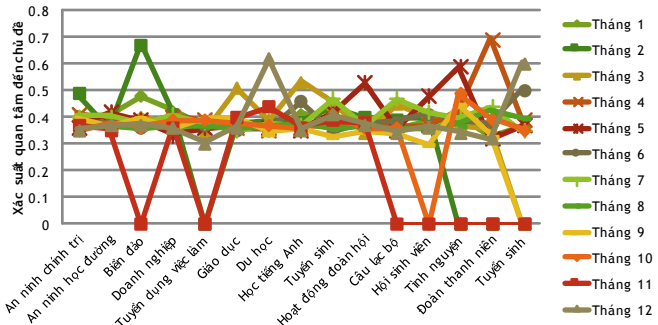
Kết quả thử nghiệm mô hình TART – Dữ liệu 1



Hình 10. Trình bày kết quả phân tích sự biến thiên xác suất xuất hiện của từ “sinh_vien” trong chủ đề “Giáo dục” qua từng giai đoạn thời gian.



Hình 11. Kết quả phân tích dựa trên xác suất chủ đề quan tâm của 3 cá nhân trong từng giai đoạn thời gian (khảo sát trên 3 chủ đề).



Hình 12. Trình bày kết quả phân tích trên 15 chủ đề trong thời gian từ tháng 01 đến tháng 12 năm 2015 (12 giai đoạn thời gian).

Kết quả thử nghiệm mô hình TART – Dữ liệu 2

Bảng 4. Trình bày kết quả phân tích 4 chủ đề quan tâm của cá nhân trong giai đoạn tháng 08-2015.

Chủ đề 1: Giáo dục		Chủ đề 2: Du học		Chủ đề 3: Học tiếng Anh		Chủ đề 4: Tuyển sinh	
Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất
học_sinh	0.01608	sinh_viên	0.02432	học	0.01927	trường	0.02563
Quốc_gia	0.01435	Anh	0.01897	tiếng	0.01867	bạn	0.01979
người	0.01407	học_sinh	0.01515	phim	0.01569	điểm	0.01676
năm	0.01161	đào_tạo	0.01437	nghe	0.01468	nộp	0.01619
phổ_biến	0.01076	Australia	0.01313	giao_tiếp	0.01342	thi	0.01573
đại_học	0.00852	tư_vấn	0.01204	nướcngoài	0.01256	thi_sinh	0.01425
đa_số	0.00818	du_học	0.01161	đúng	0.01191	xét	0.01184
trường	0.00698	thế_giới	0.01115	giỏi	0.01186	đại_học	0.01173
giảng_viên	0.00627	hỗ_trợ	0.01112	phụ_đề	0.01145	hồ_sơ	0.01099
sinh_viên	0.00415	Mỹ	0.00853	đọc_báo	0.01120	tuyển	0.01093
chương_trình	0.00373	quốc_tế	0.00719	phát_âm	0.01105	rút	0.00841
tác_phẩm	0.00373	Đại_học	0.00708	ngôn_ngữ	0.01105	mới	0.00784
phần_mềm	0.00371	học	0.00630	bổ_ích	0.01029	ngành	0.00767
thực_hành	0.00357	tổ_chức	0.00613	xem	0.00989	học_sinh	0.00761
Người gửi	Xác suất	học_phi	0.00598	đọc	0.00933571	đăng_ký	0.00704
158286	0.96567	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất
137298	0.96343	105608	0.93531	118003	0.85635	39876	0.64835
60626	0.95772	58422	0.93003	110770	0.81595	204788	0.59140
31669	0.95612	47691	0.91931	80553	0.81379	76183	0.58794
81828	0.95598						
42760	0.95463						

117098	0.94973	38108	0.91376	151667	0.80142	70833	0.58242
89872	0.94926	17919	0.90144	65834	0.78295	32949	0.57333
30717	0.94881	76940	0.87940	150337	0.77600	144665	0.55906
28007	0.94376	15145	0.87907	59749	0.75796	205021	0.55556
Người nhận	Xác suất	82605	0.87814	7278	0.75610	204669	0.54667
87517	0.97844	108764	0.87146	30809	0.73913	9523	0.53745
102862	0.96343	42527	0.86770	9614	0.73826	204116	0.52941
102967	0.95826	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất
102235	0.95772	62467	0.90760	124875	0.81379	106038	0.63964
33730	0.95658	107176	0.88415	158371	0.78295	197625	0.58794
58972	0.95657	2606	0.88314	155897	0.75796	204671	0.54667
101626	0.95598	130975	0.87940	57455	0.73913	4056	0.53745
120051	0.95463	77451	0.87907	52638	0.72477	128473	0.52294
90687	0.95426	124080	0.87814	145317	0.70968	51366	0.51724
14805	0.95247	132885	0.87146	128218	0.70787	180039	0.50554
		104570	0.86770	141342	0.70370	204737	0.49524
		79233	0.86175	177246	0.69620	89648	0.49462
		15132	0.86165	192846	0.68528	204912	0.48936

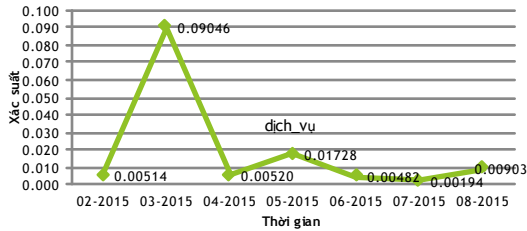
Kết quả thử nghiệm mô hình TART – Dữ liệu 2

Bảng 4.5. Trình bày kết quả phân tích chủ đề “Thương mại điện tử” được cá nhân quan tâm trong tháng 02-2015 đến tháng 06-2015.

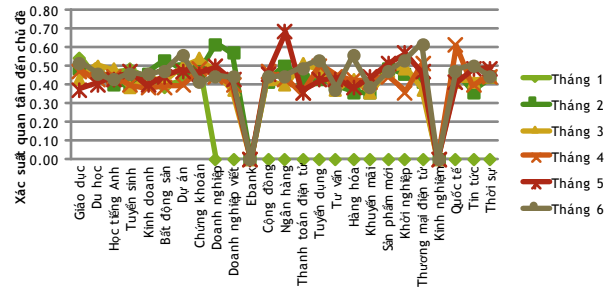
02-2015		03-2015		04-2015		05-2015		06-2015	
Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất	Từ	Xác suất
bạn	0.09046	dịch_vu	0.09046	tin_nhắc	0.01252	thanh_toán	0.01746	hạn_chế	0.01746
tài_xế	0.01951	ngân_hàng	0.01951	chuẩn	0.00858	dịch_vu	0.01728	nguy_hiểm	0.01728
người	0.01574	hình_thức	0.01574	trả_tiền	0.00776	giao_dịch	0.01502	phần_mềm	0.01502
Viber	0.01307	website	0.01307	hành_khách	0.00752	thương_mại	0.01419	sự_tin_tương	0.01419
mạng	0.01041	thông_tin	0.01041	thị_trường	0.00655	Việt_Nam	0.01287	tiền	0.01287
tiền	0.00863	Mỹ	0.00863	thương_mại	0.00638	công_ty	0.00668	chợ_đen	0.00668
mất	0.00726	Zalo	0.00726	điện_tử	0.00614	hàng	0.00640	giảm_giá	0.00640
doanh_nghiệp	0.00712	tiết_kiệm	0.00712	Internet	0.00590	dịch_vu	0.00630	tổ_chức	0.00630
TMDT	0.00651	ứng_dụng	0.00651	wifi	0.00549	đặt	0.00620	hack	0.00620
giá	0.00637	khách_hàng	0.00637	chiếm	0.00541	thời_gian	0.00540	biến_động	0.00540
Uber	0.00548	tiện_lợi	0.00548	lợi_ích	0.00537	tiền_mặt	0.00500	căng_thẳng	0.00500
taxi	0.00534	TMDT	0.00534	mua	0.00525	mua_hàng	0.00494	liên_hệ	0.00494
dịch_vu	0.00514	nướcngoài	0.00514	cắm_đổ	0.00525	di_động	0.00486	khả_năng	0.00486
vàng	0.00507	facebook	0.00507	dịch_vu	0.00520	phát_triển	0.00482	dịch_vu	0.00482
đúng	0.00466	kết_nối	0.00466	thanh_toán	0.00508	thiết_bị	0.00485	TMDT	0.00485
Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất	Người gửi	Xác suất
70464	0.82888	30257	0.95478	156036	0.90879	40957	0.87893	19626	0.95767
14652	0.81421	43941	0.95130	19325	0.87841	129579	0.86813	96561	0.95767
178518	0.70464	159072	0.95130	70262	0.87131	26286	0.83601	11028	0.95429

86559	0.69921	18960	0.94435	144011	0.86460	47736	0.81105	48455	0.95429
1182	0.67742	69993	0.94087	155430	0.86450	16264	0.80876	11050	0.95031
126092	0.65049	19702	0.93391	103628	0.86286	120626	0.80296	66030	0.95031
87327	0.64401	96714	0.86151	158531	0.86047	102482	0.78325	66252	0.94015
149073	0.64384	142267	0.80354	50733	0.85768	180396	0.77327	105841	0.92945
143228	0.63218	60976	0.66197	89787	0.84834	207296	0.73941	88967	0.86667
82496	0.62319	139920	0.60656	68962	0.84186	140619	0.73089	83619	0.86596
Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất	Người nhận	Xác suất
13426	0.70492	105830	0.91770	104116	0.90879	139366	0.82353	111253	0.93855
184607	0.70464	3836	0.66197	46167	0.87841	147443	0.79310	137588	0.90025
134347	0.69921	5253	0.48413	144665	0.86460	135408	0.77327	87188	0.90025
73649	0.69921	150587	0.41818	101460	0.86286	57198	0.77327	124448	0.85539
60406	0.65049	102851	0.40426	141234	0.86047	104265	0.76766	40329	0.85539
15949	0.64372	79827	0.39535	45740	0.85768	103255	0.76271	105677	0.66758
13771	0.64318	72037	0.38195	533	0.84834	207331	0.76271	25468	0.51397
112405	0.58252	105924	0.37255	30875	0.84186	115263	0.76228	18507	0.45679
88862	0.57282	39881	0.36864	116984	0.83862	39594	0.76228	41288	0.39326
49351	0.52239	21826	0.36585	66976	0.82456	76393	0.75770	53462	0.391288

Kết quả thử nghiệm mô hình TART - Dữ liệu 2

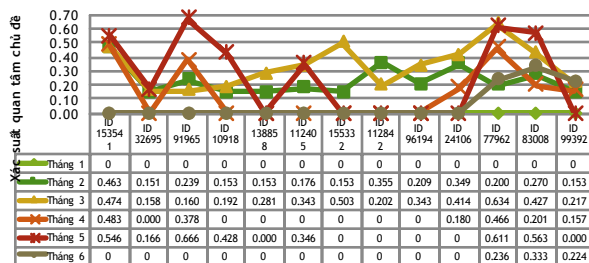


Hình 13. Trình bày kết quả phân tích sự thay đổi xác suất xuất hiện của từ "dịch_vu" trong chủ đề "Thương mại điện tử" qua từng giai đoạn thời gian.

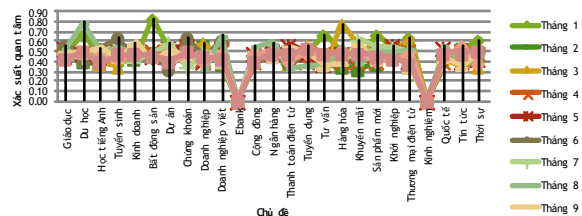


Hình 14. Trình bày kết quả phân tích trên 25 chủ đề trong thời gian từ tháng 01 đến tháng 10 năm 2015 (6 giai đoạn thời gian). Khảo sát trên 7444 cá nhân tham gia trao đổi.

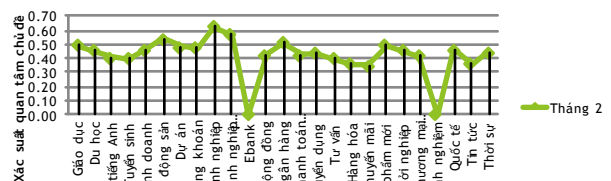
Kết quả thử nghiệm mô hình TART - Dữ liệu 2



Hình 15. Trình bày sự biến thiên việc quan tâm đến chủ đề "Thời sự" của 13 cá nhân (chọn ngẫu nhiên) trong giai đoạn từ tháng 01-2015 đến tháng 06-2015.



Hình 16. Trình bày sự biến thiên việc quan tâm đến 25 chủ đề của 7444 cá nhân quan tâm trao đổi 25 chủ đề từ tháng 01-2014 đến tháng 10-2014.



Hình 17. Trình bày sự biến thiên việc quan tâm đến 25 chủ đề của 7444 cá nhân quan tâm trao đổi 25 chủ đề trong tháng 02-2014.

Thử nghiệm và đánh giá kết quả

Mục tiêu nghiên cứu 2

Dữ liệu thử nghiệm

Bảng 6. Một số mẫu vector quan tâm chủ đề của cá nhân (actor vector đầu vào)

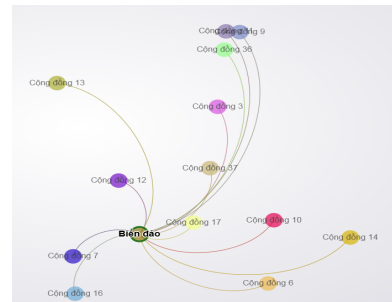
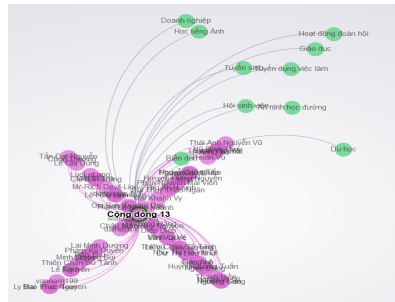
	Chủ đề quan tâm							Cá nhân
	T-0	T-1	T-2	T-3	T-4	T-5	T-6	
\vec{v}_1	(0.64444,	0.34545,	0.46826,	0.0,	0.33721,	0.0,	0.0)	Mr.tajkjd
\vec{v}_2	(0.30435,	0.44565,	0.33333,	0.30435,	0.33333,	0.52941,	0.0)	dsvantan
\vec{v}_3	(0.39601,	0.48718,	0.0,	0.35484,	0.0,	0.38462,	0.0)	nguyen.nhi.334491
\vec{v}_4	(0.34694,	0.40741,	0.0,	0.39227,	0.0,	0.36,	0.0)	trang.harry.7
\vec{v}_5	(0.0,	0.35135,	0.0,	0.41935,	0.0,	0.31429,	0.0)	anna.vy.334
\vec{v}_6	(0.0,	0.36,	0.0,	0.33333,	0.0,	0.44828,	0.40741)	haianh.nguyen.52012
\vec{v}_7	(0.48718,	0.32431	0.0,	0.0,	0.0,	0.0,	0.31034)	quyvan.pham.54
\vec{v}_8	(0.40741,	0.31034,	0.0,	0.0,	0.0,	0.41772,	0.0)	su.heo.1656
\vec{v}_9	(0.35135,	0.33333,	0.40741,	0.0,	0.30923,	0.34545,	0.0)	phuc.hanh.9678
\vec{v}_{10}	(0.64557,	0.9,	0.34884,	0.58974,	0.33354	0.0,	0.77465)	GiámĐốcTàiChinhTiềmNăng

Kết quả thử nghiệm

- ▶ Kích thước lớp ra Kohonen: 14 x 14 (25 neurons).
- ▶ Mỗi vector nhập có 15 thành phần tương ứng 15 chủ đề.
- ▶ Thời gian tháng 08/2014.
- ▶ Số cá nhân tham gia trao đổi 15 chủ đề: 2244
- ▶ Kết quả thử nghiệm 2: số cộng đồng khám phá là 41



- ☐ Danh sách các chủ đề quan tâm
 - ☐ An ninh chính trị
 - An ninh học đường(0.04454)
 - Biển đảo(0.01156)
 - ☐ Doanh nghiệp(0.00642)
 - Tuyển dụng việc làm(0.3401)
 - ☐ Giáo dục(0.02928)
 - Du học(0.00504)
 - Học tiếng Anh(0.00725)
 - Tuyển sinh(0.00688)
 - ☐ Hoạt động đoàn hội(0.01975)
 - Câu lạc bộ
 - Hội sinh viên(0.00859)
 - Tình nguyện
 - Đoàn thanh niên



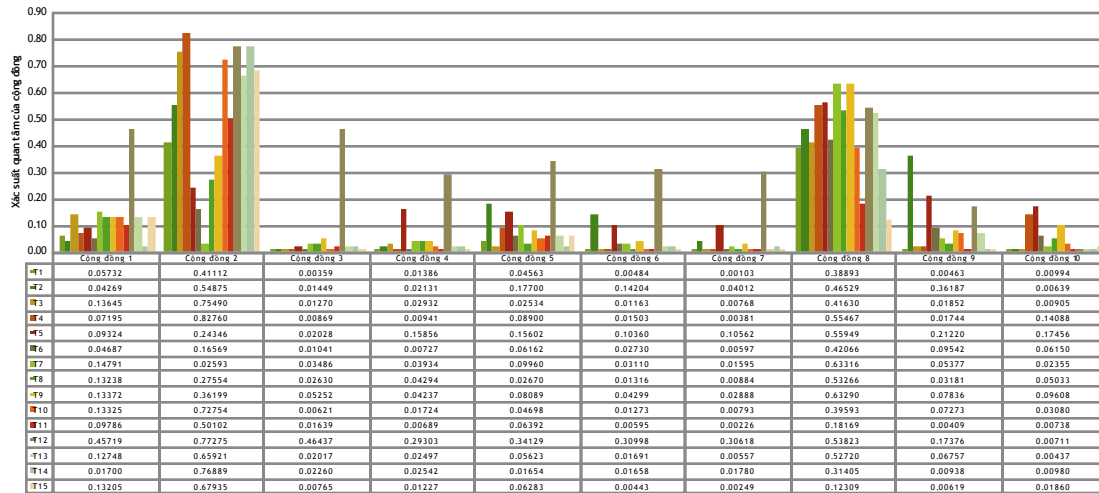
Kết quả thử nghiệm

Bảng 7. Trình bày xác suất quan tâm của từng cộng đồng đối với từng chủ đề cụ thể

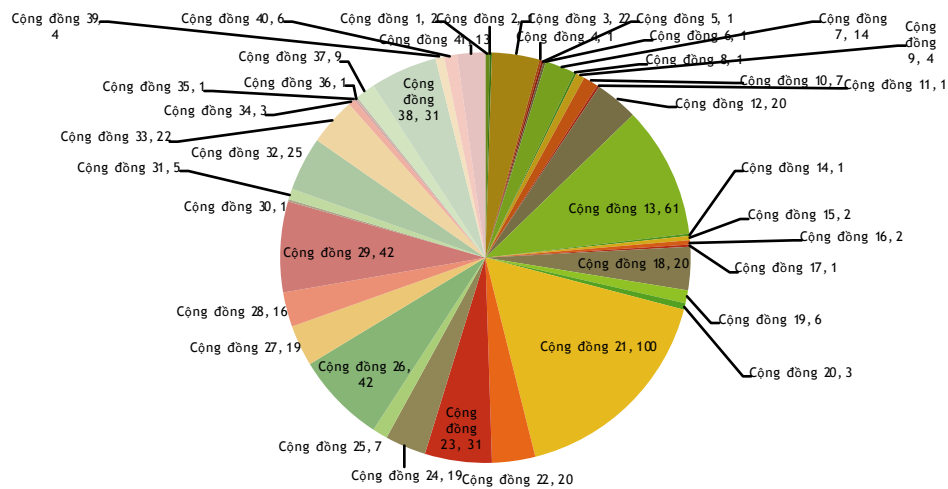
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Cộng đồng 1	0.05732	0.04269	0.13645	0.07195	0.09324	0.04687	0.14791	0.13238	0.13372	0.13325	0.09786	0.45719	0.12748	0.01700	0.13205
Cộng đồng 2	0.41112	0.54875	0.75490	0.82760	0.24346	0.16569	0.02593	0.27554	0.36199	0.72754	0.50102	0.77275	0.65921	0.76889	0.67935
Cộng đồng 3	0.00359	0.01449	0.01270	0.00869	0.02028	0.01041	0.03486	0.02630	0.05252	0.00621	0.01639	0.46437	0.02017	0.02260	0.00765
Cộng đồng 4	0.01386	0.02131	0.02932	0.00941	0.15856	0.00727	0.03934	0.04294	0.04237	0.01724	0.00689	0.29303	0.02497	0.02542	0.01227
Cộng đồng 5	0.04563	0.17700	0.02534	0.08900	0.15602	0.06162	0.09960	0.02670	0.08089	0.04698	0.06392	0.34129	0.05623	0.01654	0.06283
Cộng đồng 6	0.00484	0.14204	0.01163	0.01503	0.10360	0.02730	0.03110	0.01316	0.04299	0.01273	0.00595	0.30998	0.01691	0.01658	0.00443
Cộng đồng 7	0.00103	0.04012	0.00768	0.00381	0.10562	0.00597	0.01595	0.00884	0.02888	0.00793	0.00226	0.30618	0.00557	0.01780	0.00249
Cộng đồng 8	0.38893	0.46529	0.41630	0.55467	0.55949	0.42066	0.63316	0.53266	0.63290	0.39593	0.18169	0.53823	0.52720	0.31405	0.12309
Cộng đồng 9	0.00463	0.36187	0.01852	0.01744	0.21220	0.09542	0.05377	0.03181	0.07836	0.07273	0.00409	0.17376	0.06757	0.00938	0.00619
Cộng đồng 10	0.00994	0.00639	0.00905	0.14088	0.17456	0.06150	0.02355	0.05033	0.09608	0.03080	0.00738	0.00711	0.00437	0.00980	0.01860

Kết quả thử nghiệm

Hình 19. Trình bày kết quả phân tích cộng đồng dựa trên bảng kết quả 7.



Kết quả thử nghiệm



Hình 20. Trình bày kết quả phân tích cộng đồng trong tháng 08-2014

Đánh giá kết quả thử nghiệm gom cụm

Dưới đây là kết quả tính F-measure với $m = 5$ cụm và $k = 6$ cụm

Máy (k)/Bảng tay (m)	m_0	m_1	m_2	m_3	m_4
k_0	0.43	0.15	0.84	0.52	0.68
k_1	0.67	0.61	0.00	0.16	0.00
k_2	0.00	0.36	0.51	0.62	0.16
k_3	0.72	0.00	0.55	0.55	0.34
k_4	0.81	0.73	0.25	0.00	0.72
k_5	0.19	0.00	0.15	0.29	0.36
MAX	0.81	0.73	0.84	0.62	0.72

Tổng MAX gom cụm bằng Kohonen trên bảng: $0.81 + 0.73 + 0.84 + 0.62 + 0.72 = 3.71$

Máy (k)/Bảng tay (m)	m_0	m_1	m_2	m_3	m_4
k_0	0.51	0.61	0.23	0.41	0.00
k_1	0.49	0.39	0.41	0.17	0.83
k_2	0.32	0.28	0.51	0.76	0.14
k_3	0.68	0.87	0.48	0.29	0.36
k_4	0.79	0.71	0.51	0.23	0.28
k_5	0.00	0.25	0.72	0.00	0.68
MAX	0.79	0.87	0.72	0.76	0.83

Tổng MAX gom cụm bằng Kohonen trên bảng: $0.79 + 0.87 + 0.72 + 0.82 + 0.83 = 3.97$

Danh mục các công trình công bố chính

- ▶ [CT1] Thanh Ho and Phuc Do, *Analyzing the Changes in Online Community based on Topic Model and Self-Organizing Map*, International Journal of Advanced Computer Science and Applications (IJACSA), 6(7), 2015, ISSN: 2158-107X.
- ▶ [CT2] Thanh Ho, Duy Doan, Phuc Do (2014), *Discovering Hot Topics On Social Network Based On Improving The Aging Theory*, Advances in Computer Science: an International Journal. Volume 3, Issue 3, p. 48-53, ISSN: 2322-5157.
- ▶ [CT3] Hồ Trung Thành, Đỗ Phúc (2014), *Ontology tiếng Việt trong lĩnh vực giáo dục đại học*, Tạp chí Khoa học Công nghệ - Viện Hàn lâm Khoa học Công nghệ Việt Nam, Tập 52, số 1B, pp. 89-100, ISSN: 0866-708x.
- ▶ [CT4] Hồ Trung Thành, Đỗ Phúc (2014), *Mô hình tích hợp khám phá và gán nhãn chủ đề tiếp cận theo mô hình chủ đề*, Tạp chí Khoa học Công nghệ ĐHQG-HCM, số K4, tập 17, ISSN: 1859-0128.
- ▶ [CT5] Thanh Ho, Phuc Do (2014), *Analyzing Users' Interests with the Temporal Factor Based on Topic Modeling*, ACII DS 03-2015, Indonesia, Springer, pp. 106-115, ISSN: 0302-9743, ISBN: 978-3-319-15704-7.
- ▶ [CT6] Thanh Ho, Phuc Do (2015), *Discovering Communities of Users on Social Networks Based on the Topic Model Combined with Kohonen Network*, KSE 10/2015, UIT, Vietnam, 10/2015.

Danh mục các công trình công bố khác liên quan

- ▶ [CT7] Phan Hồ Viết Trường, **Hồ Trung Thành**, Đỗ Phúc (2013), *Phân tích tầm ảnh hưởng đối tượng theo chủ đề trong mạng xã hội*, Tạp chí Khoa học Công nghệ, Viện Hàn lâm Khoa học Công nghệ Việt Nam, tập 52, số 1B, pp. 101-111, ISSN: 0866-708x.
- ▶ [CT8] Nghe Nguyen, **Thanh Ho** and Phuc Do (2015), *Finding the Most Influential User of a Specific Topic on the Social Networks*, Advances in Computer Science : an International Journal, Volume 4, Issue 2, ISSN: 2322-5157.
- ▶ [CT9] Muon Nguyen, **Thanh Ho**, Phuc Do (2013), *Social Networks Analysis Based on Topic Modeling*, The 10th IEEE RIVF International Conference on Computing and Communication Technologies, Hanoi, pp. 119-123, ISBN: 978-1-4799-1350-3.
- ▶ [CT10] Tran Quang Hoa, Vo Ho Tien Hung, Nguyen Le Hoang, **Ho Trung Thanh**, Do Phuc (2014), *Finding the Cluster of Actors in Social Network based on the Topic of Messages*, ACIIDS 04-2014, Thailand, Springer, pp. 183-190, ISBN: 983-3-319-054756-6.

Tham gia đề tài NCKH

- ▶ Xây dựng hệ thống phân tích mạng xã hội theo chủ đề và ứng dụng vào mạng xã hội trong trường đại học. Mã số đề tài: B2013-26-02. Chủ nhiệm đề tài: PGS.TS Đỗ Phúc. Đề tài cấp ĐHQG-HCM loại B, đã nghiệm thu vào tháng 10/2015, đạt loại tốt.

Tài liệu tham khảo

**TRÂN TRỌNG CẢM ƠN
QUÍ THẦY/CÔ VÀ ANH/CHỊ!**