

Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad del País Vasco

Tesis para optar al grado de Doctor en Informática

Contribuciones al modelado semántico de las comunidades de práctica en línea

Felipe Aguilera Valenzuela

Bajo la dirección de: Dr. Manuel Graña Romay / Dr. Sebastián Ríos Pérez

agenda de presentación





introducción motivación





Chispas en plexi!

chaoo compa





introducción motivación

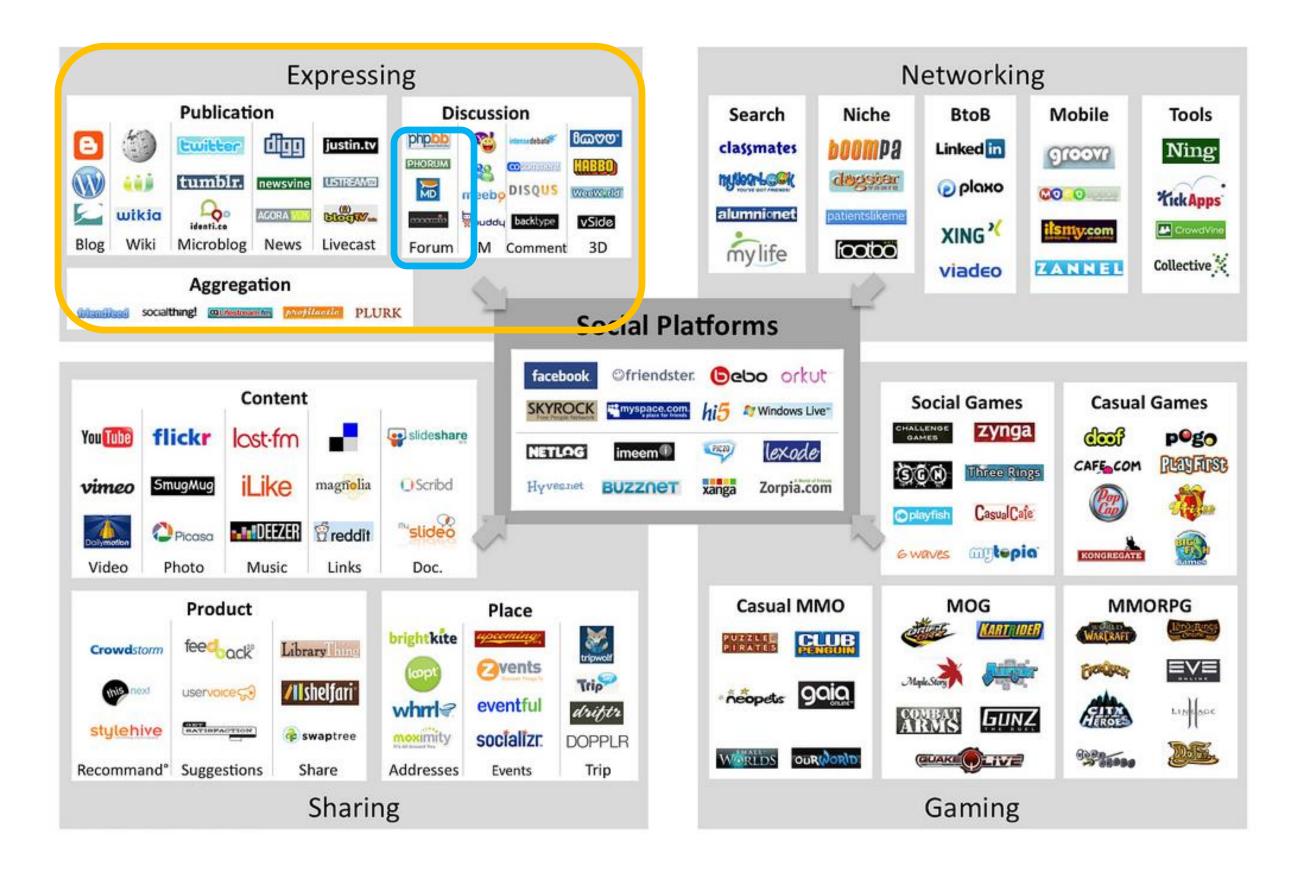
Cualquier persona puede utilizar Internet como una herramienta para agrupar personas con el mismo interés.

Tecnologías como los foros de discusión y blogs hacen mas fácil la tarea de:

- Discutir temas en conjunto con una gran cantidad de personas.
- Difundir ideas a una gran cantidad de personas en períodos de tiempo "cortos".
- Organizar eventos virtuales y eventos reales (que podrían ser buenos o malos).



introducción motivación



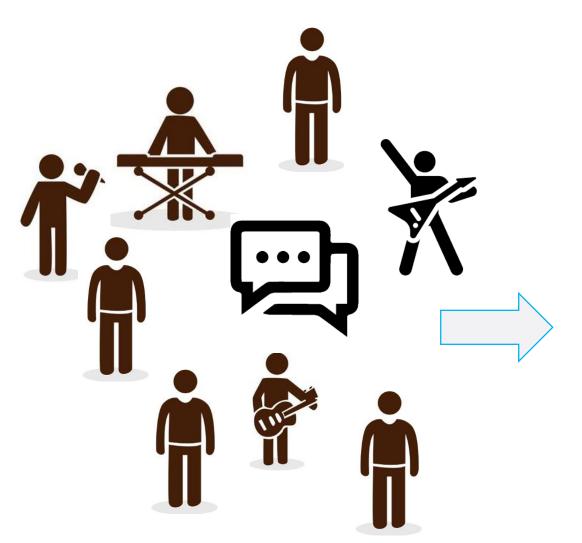


introducción foro de discusión



introducción

formación de comunidades virtuales



Grupos de personas con interés común





COMUNIDAD VIRTUAL



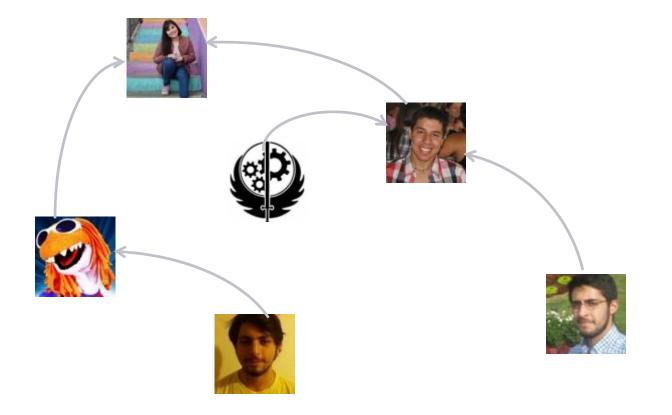


Grupo de personas compartiendo algún interés común experiencias y/o necesidades, enlazados a través de relaciones sociales por medio de las cuales se obtienen recursos importantes, desarrollan fuertes sentimientos interpersonales de pertenencia y necesidad mutua, y se da el surgimiento de un sentido de identidad compartida

-- Wellman, 2000

introducción formación de relaciones sociales





sociograma / grafo

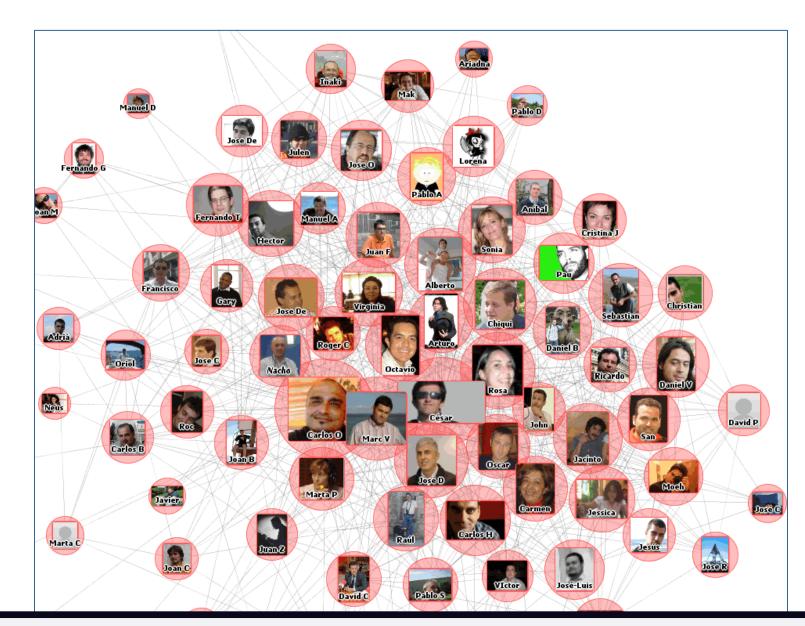


introducción

formación de relaciones sociales







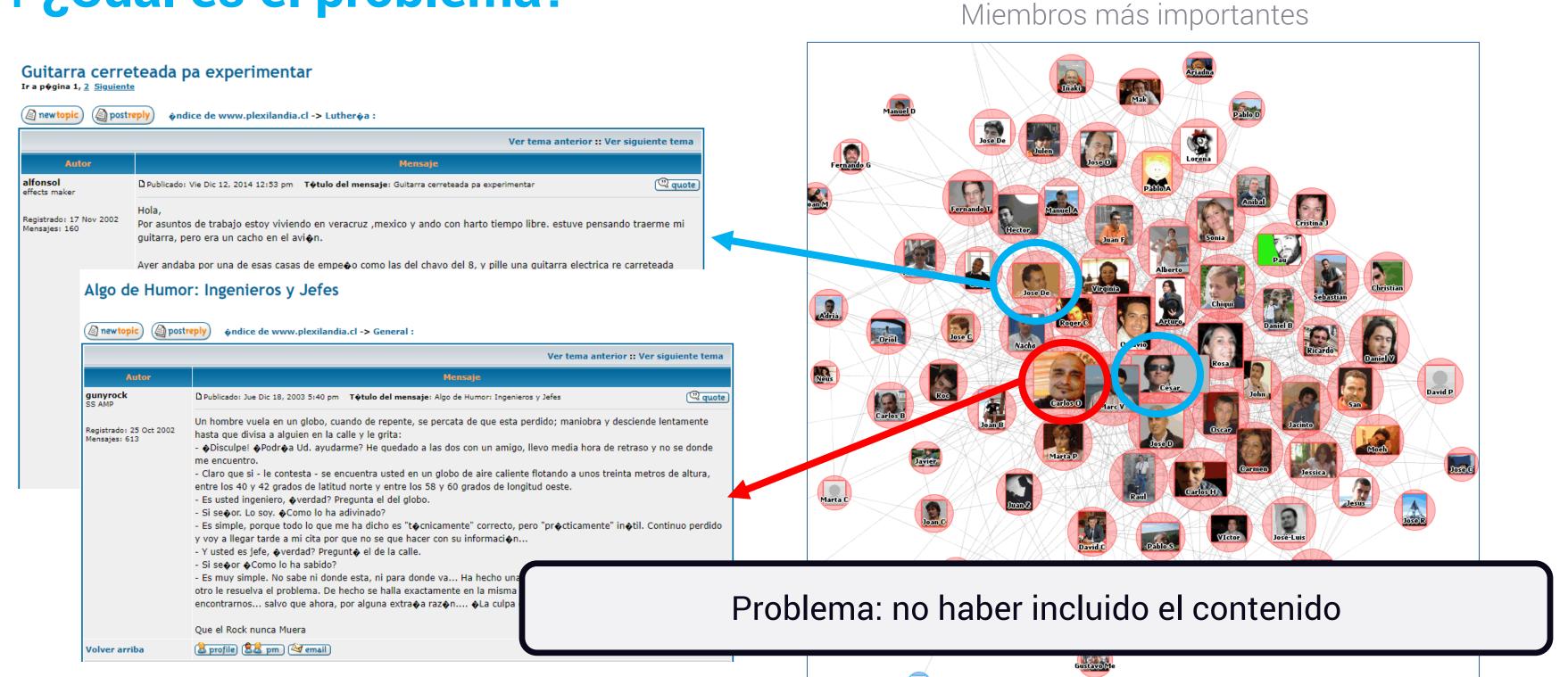
Análisis Estructural de la Comunidad Virtual ¿Qué miembros/personas son más importantes?







introducción ¿Cuál es el problema?

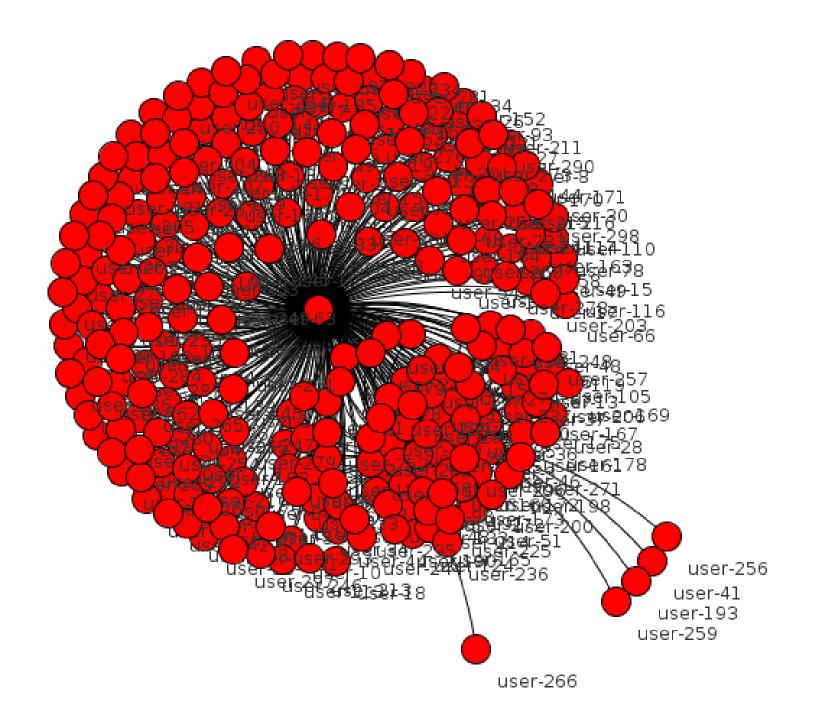




introducción ¿Cuál es el problema?

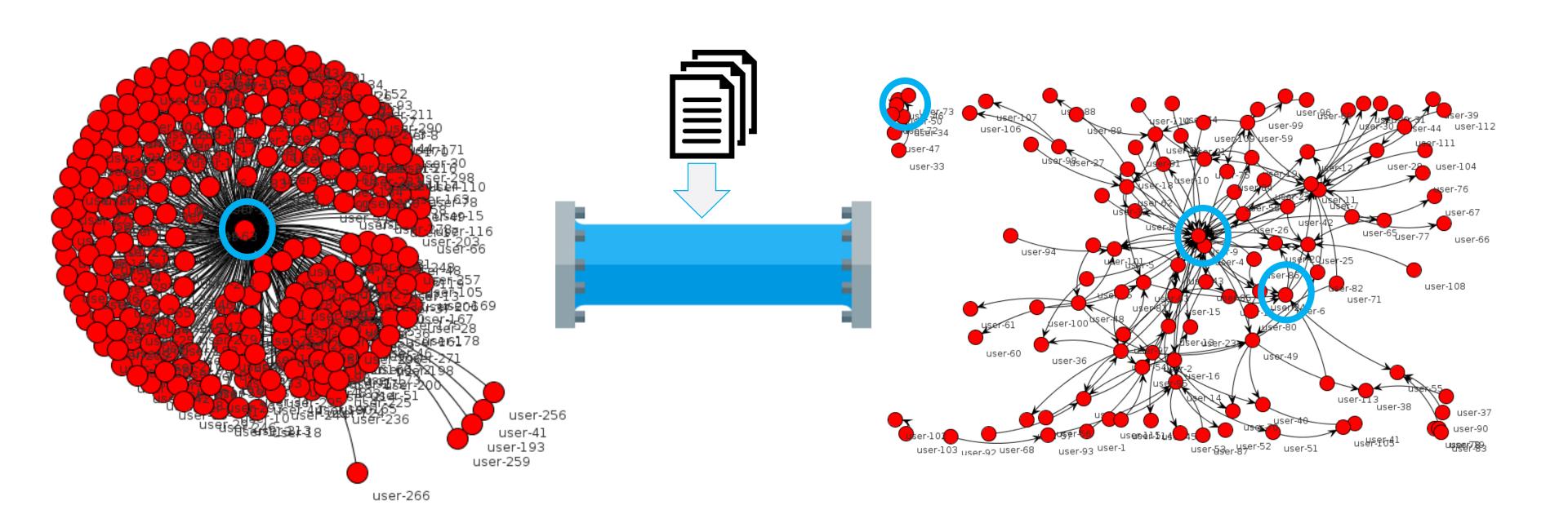


- + 2.500 miembros
- + 87.000 publicaciones
- +100.000 palabras diferentes





introducción enfoque: filtro semántico



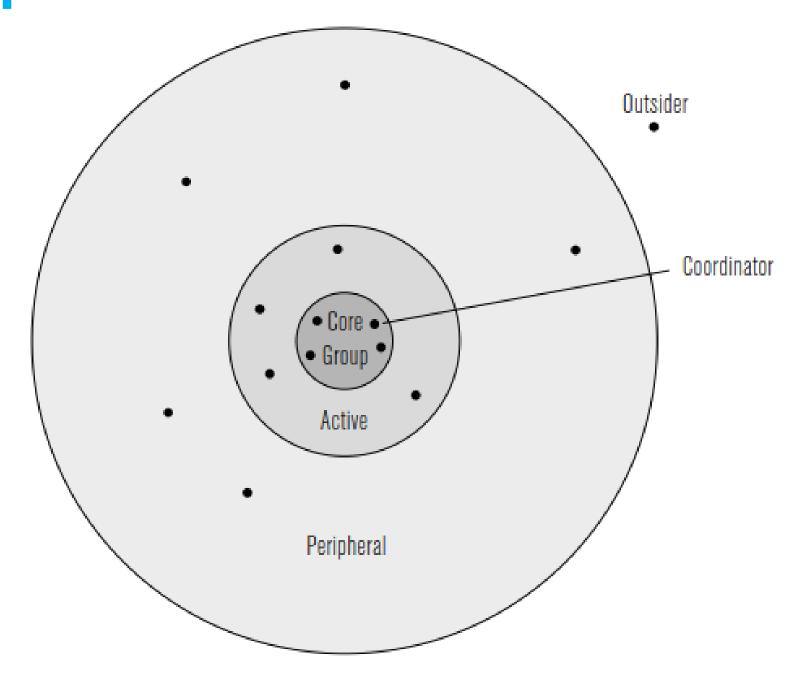


introducción tipos de estructuras sociales





introducción niveles de participación

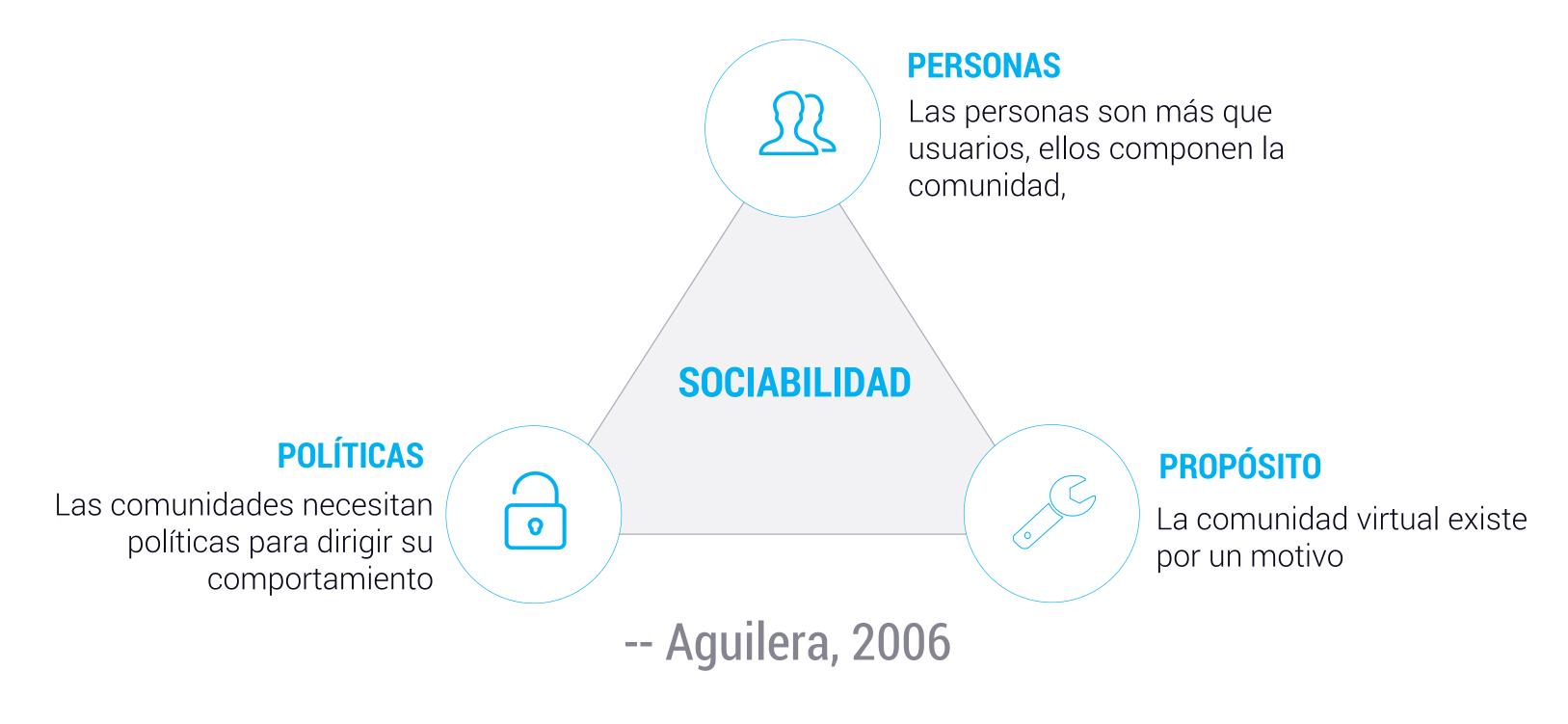


-- Wenger,2002

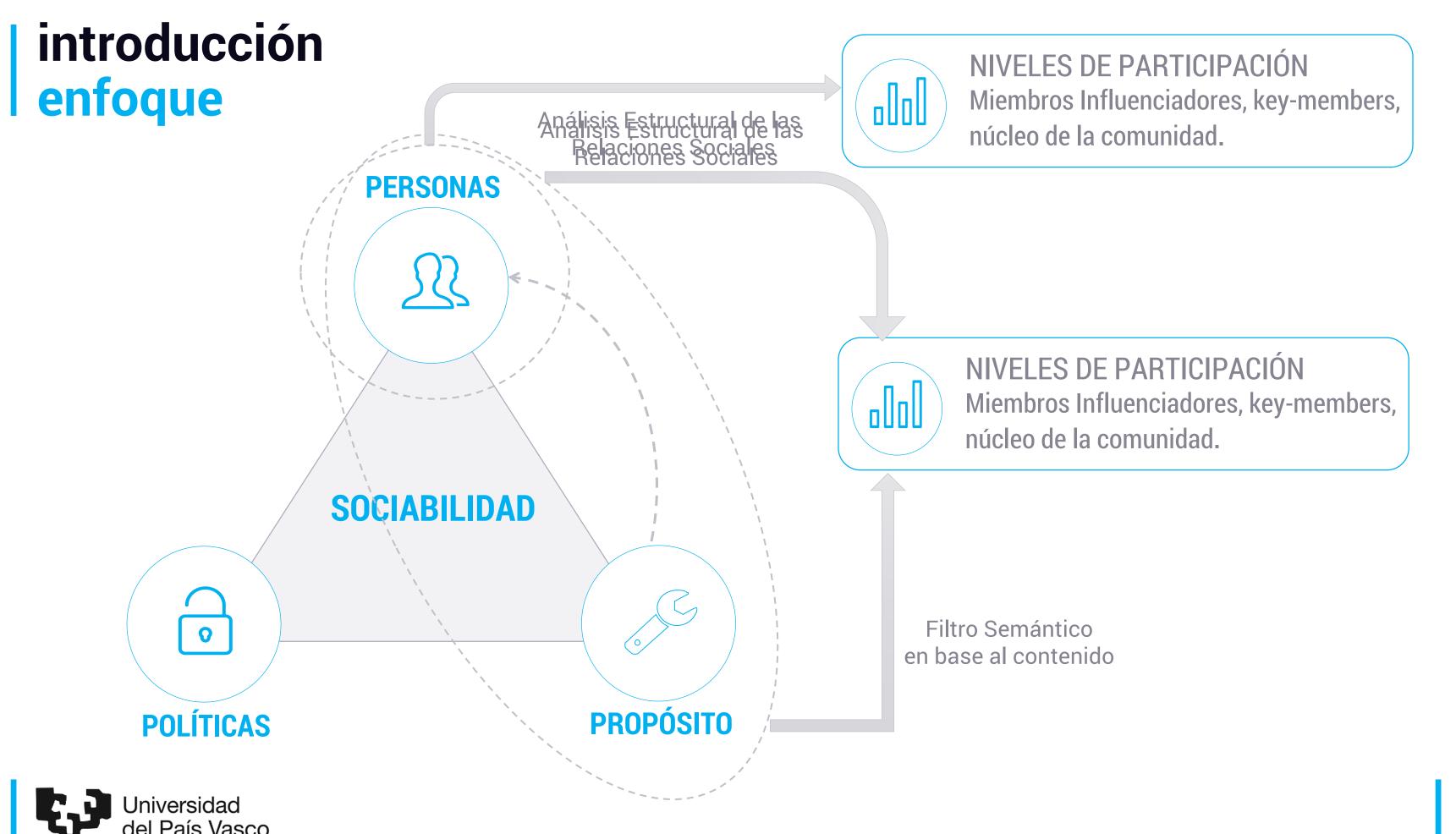




introducción concepto de sociabilidad







introducción objetivo general

Proponer nuevas técnicas que permitan construir modelos de **redes sociales** de mayor ajuste a la realidad, para poder **describir** el funcionamiento de dichas estructuras sociales en base a la **interacción social efectiva**.

Complementando el análisis estructural tradicional, basado en la representación como grafo de la estructura social, con la utilización de análisis semántico del contenido de las interacciones (mensajes) entre los miembros de una comunidad virtual.



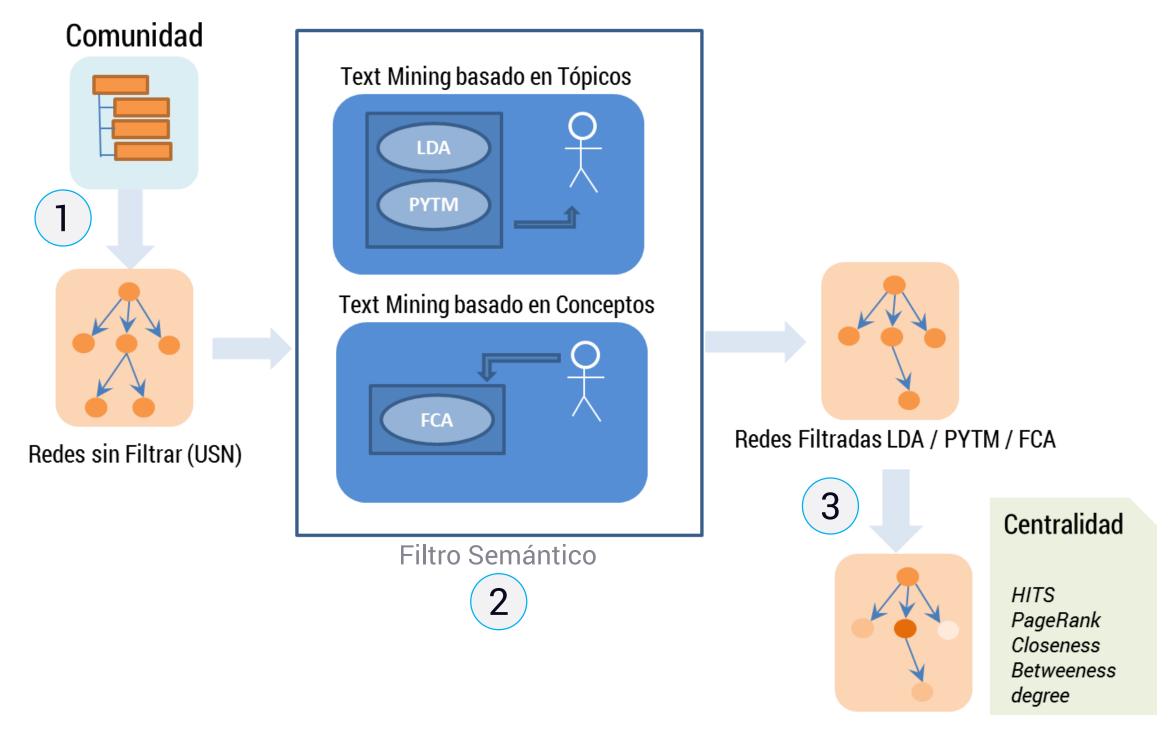
agenda de presentación





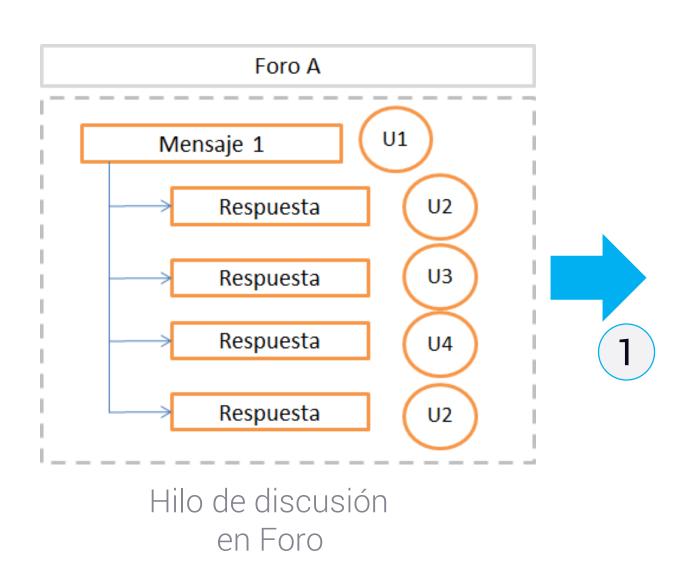
método algorítmico

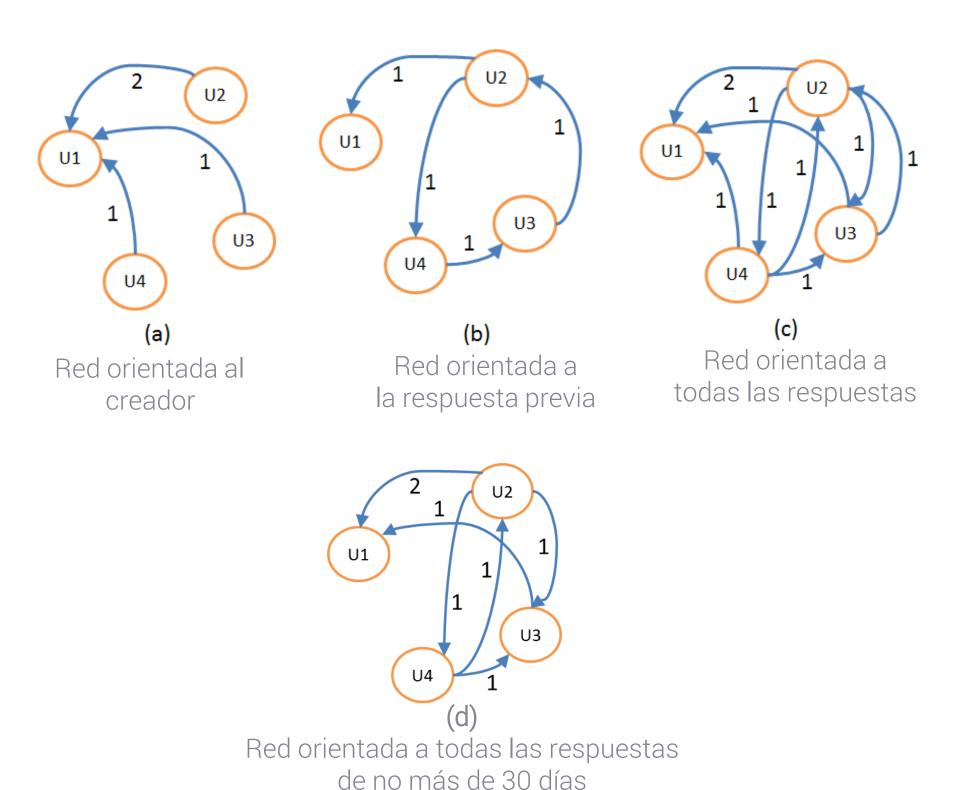
resumen





método algorítmico representación basada en grafos

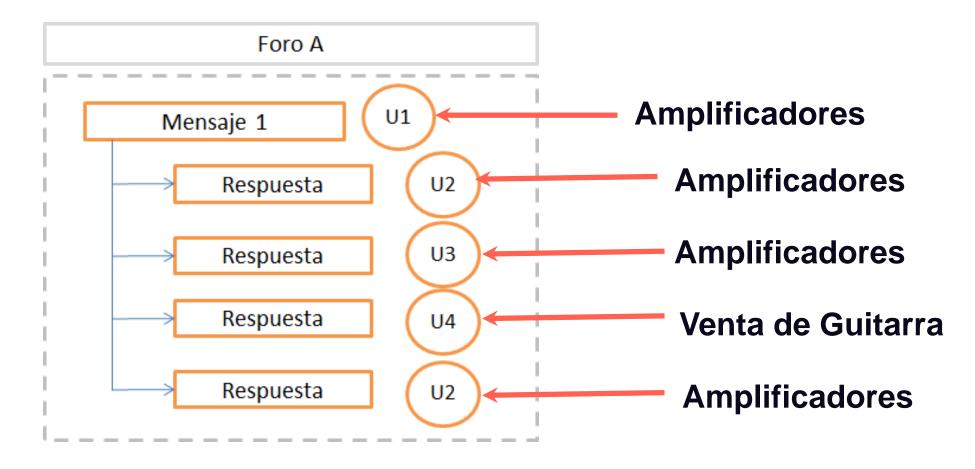




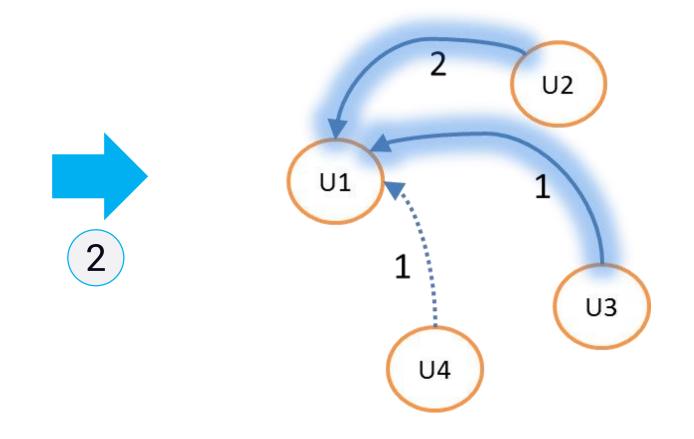


método algorítmico filtro semántico

¿De qué están hablando las personas?



La estructura del grafo se define en base a lo que se está hablando



método algorítmico filtro semántico

¿De qué están hablando las personas?

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLIÖN	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

Problema: no conocemos los tópicos ni las palabras que lo componen

debes conectarlo a una carga de 4,8 o 16 ohms del suficiente wattage. tienes que hacerle el bias y medir bien antes . si el bias no funcionara pordrias echarte algunas cosas. el amplificador debe estar conectado a tierra. deve estar cerrado es peligroso prender el amplificador con el chasis abierto.

mide los voltages primero sin tubos para estar seguro que los voltage son los adecuados.

Modelos de Semántica Latente

Tópico 5 (Guitarras)		Tópico 6 (Placas)		
cable	0.0261	placa	0.0185	
ruido	0.0251	papel	0.0120	
cables	0.0170	queda	0.0097	
tierra	0.0142	caja	0.0095	
problema	0.0133	placas	0.0093	
tiene	0.0114	aluminio	0.0079	
guitarra	0.0112	son	0.0073	
gracias	0.0088	agua	0.0072	
volumen	0.0082	después	0.0068	
jack	0.0082	pintura	0.0068	



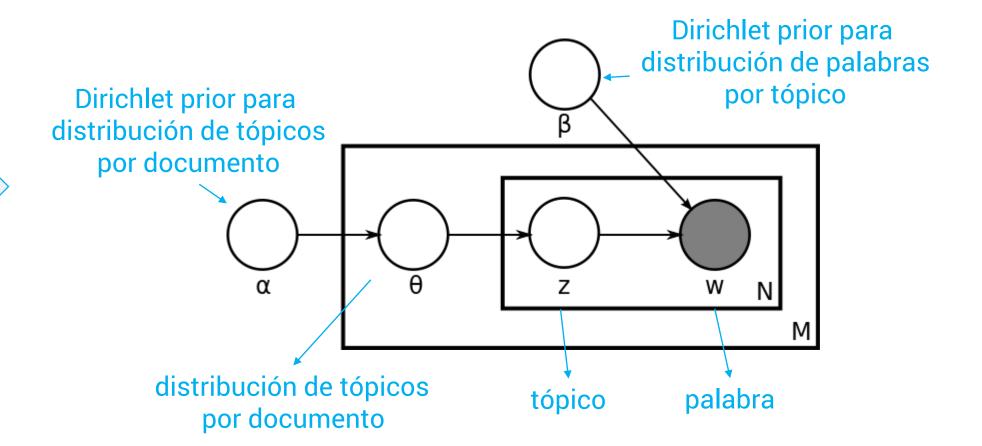
método algorítmico filtro semántico - LDA

Modelos de Semántica Latente

- Latent Dirichlet Allocation (LDA)
- Pitman-Yor Topic Model (PYTM)

Son modelos generativos de tópicos, que intentan imitar el proceso en el cual un documento es escrito.

Identifican conjuntos de palabras, los cuales son llamados tópicos.





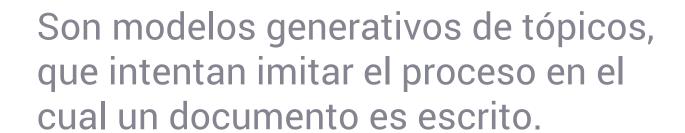
Inferencia a través de Gibbs sampling $p(w_n|z_n,\beta)$



método algorítmico filtro semántico - PYTM

Modelos de Semántica Latente

- Latent Dirichlet Allocation (LDA)
- Pitman-Yor Topic Model (PYTM)



Identifican conjuntos de palabras, los cuales son llamados tópicos.

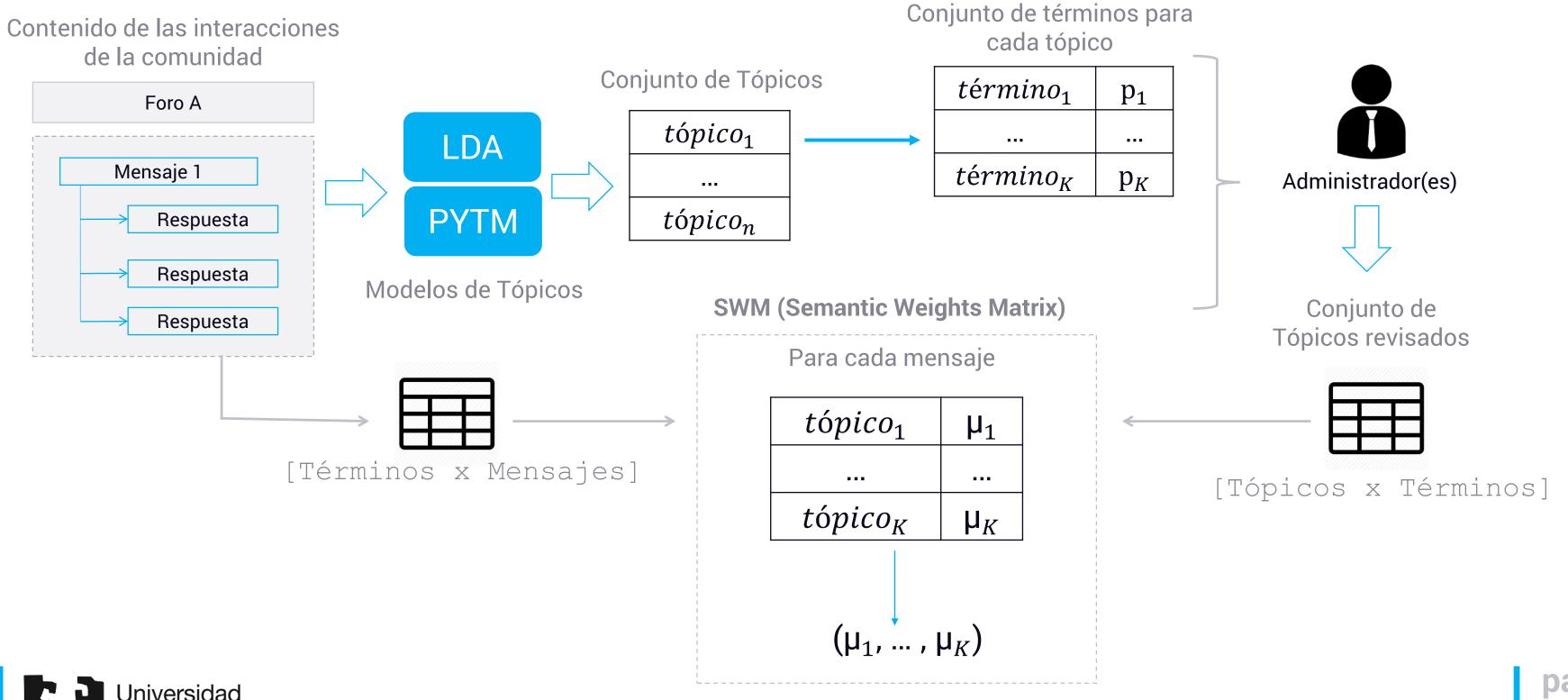
Es una generalización del modelo LDA

Está basado en Chinese Restaurant Process (CPR)

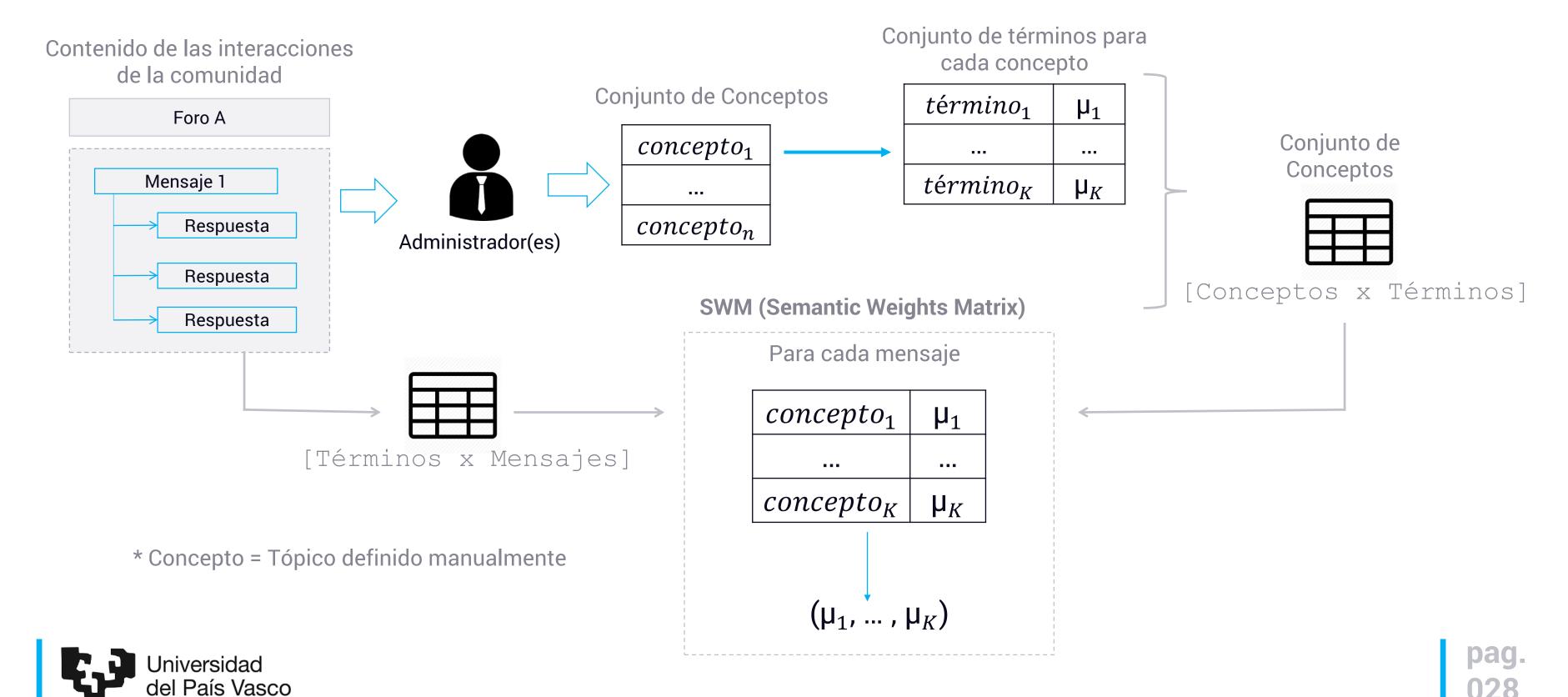
• Permite capturar la distribución de ley de potencia (power law) que siguen las palabras.



método algorítmico filtro semántico - tópicos

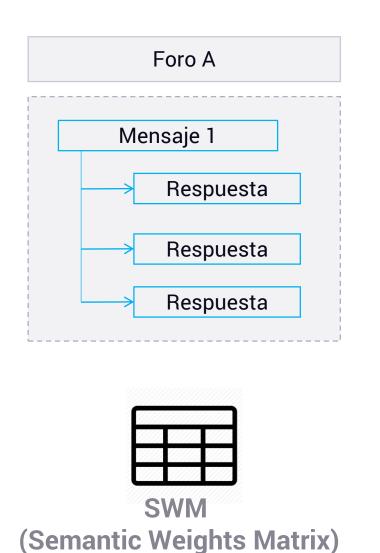


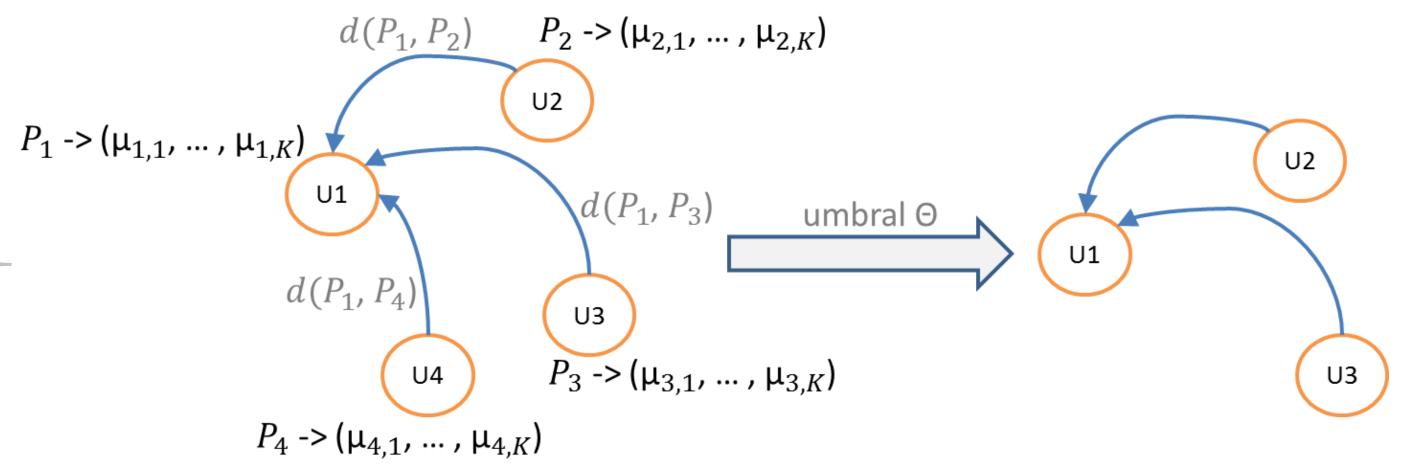
método algorítmico filtro semántico - conceptos

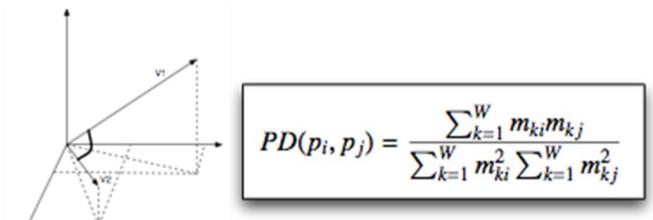


método algorítmico

filtro semántico - proceso de filtrado de los grafos



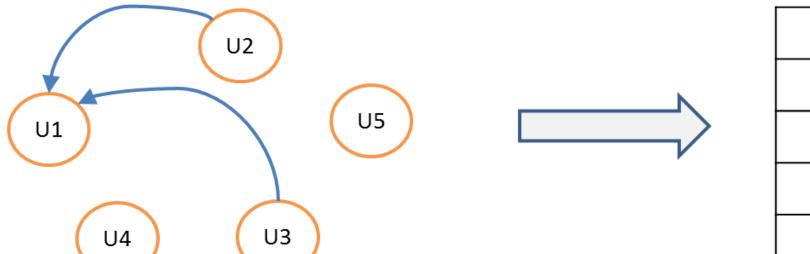




Se filtran también los mensajes cuyos componentes no son relevantes en al menos un cierto nivel ρ , en donde $\rho \in (0, 0,2]$.



método algorítmico aplicación de algoritmo de centralidad



U_1	$puntaje_1$		
U_2	puntaje ₂		
U_3	puntaje ₃		
U_4	$puntaje_4$		
U_5	puntaje ₅		

HITS
degree
pagerank
closeness
betweeness



agenda de presentación





resultados resumen



Grupo de personas que se han reunido en torno a la construcción de efectos de música. Utilizan un foro de discusión.

- +2.500 miembros.
- +14 años funcionamiento.
- +87.000 publicaciones.
- 2-3 administradores



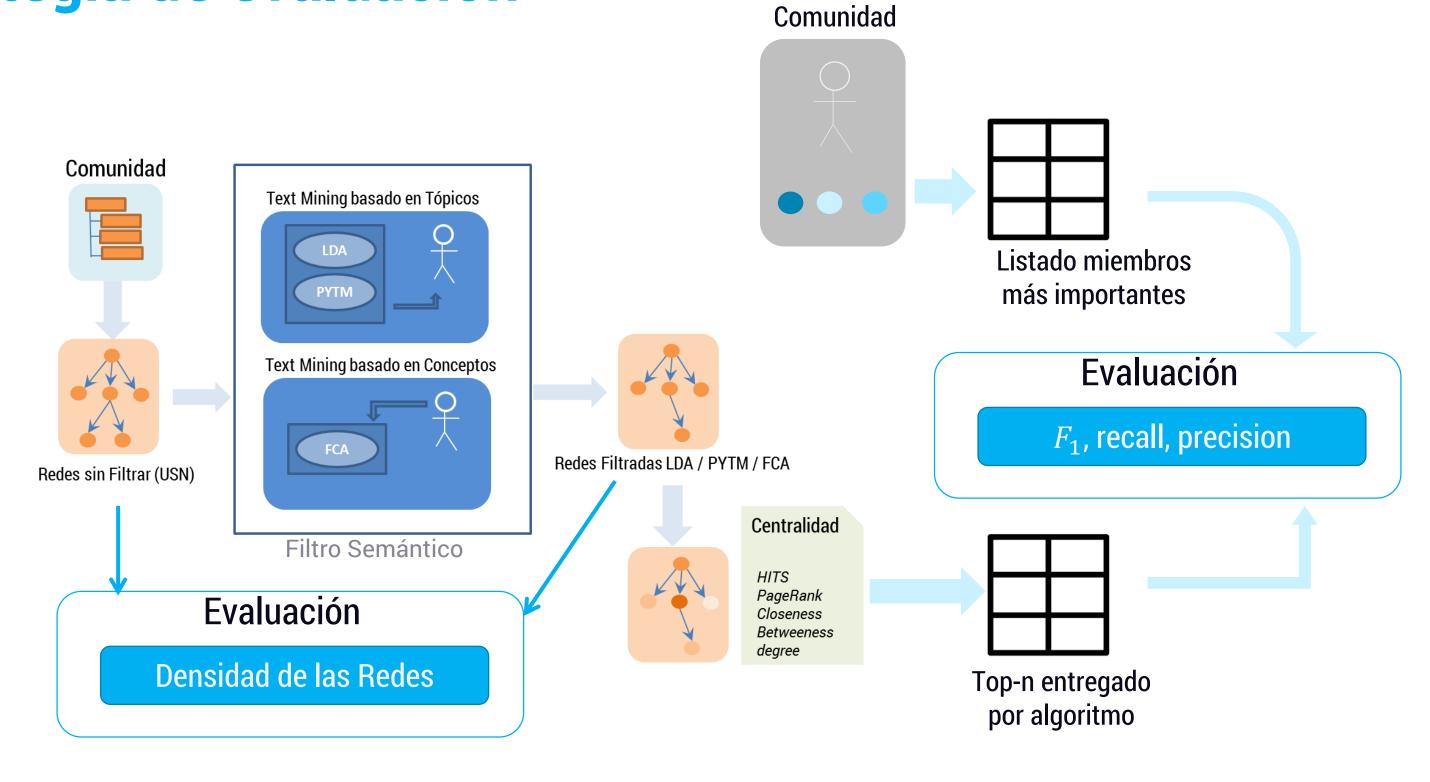
Se encuentra al interior de una empresa. Utilizan un sistema web para compartir ideas de innovación.

- +500 miembros
- +3 años funcionamiento
- +1.000 publicaciones
- 1 Administrador



resultados

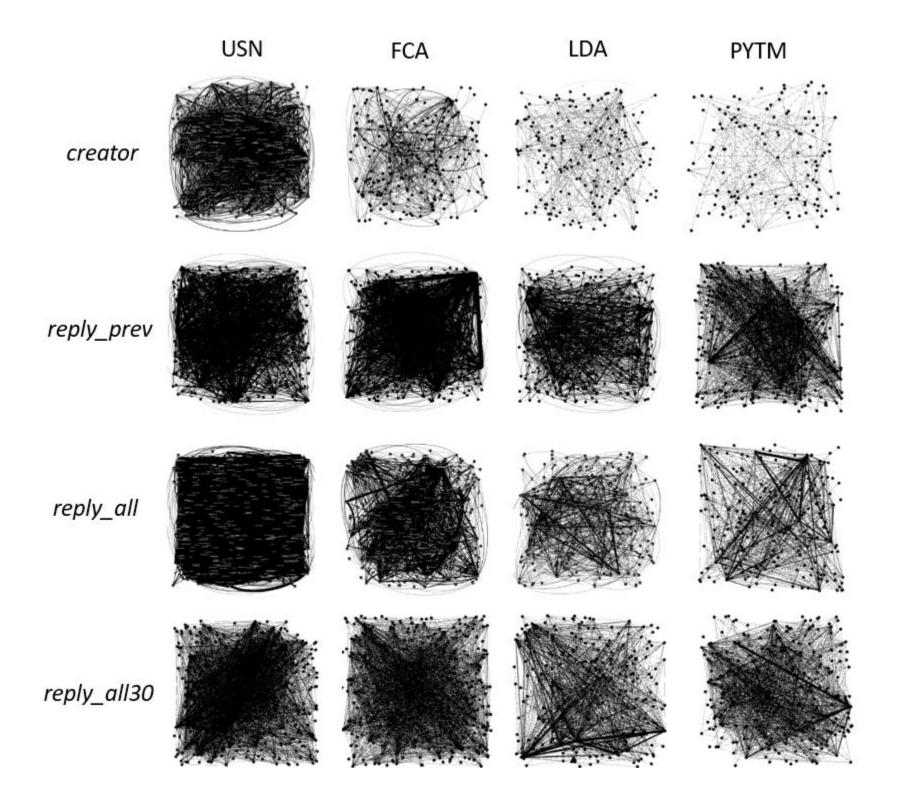
estrategia de evaluación



Administradores

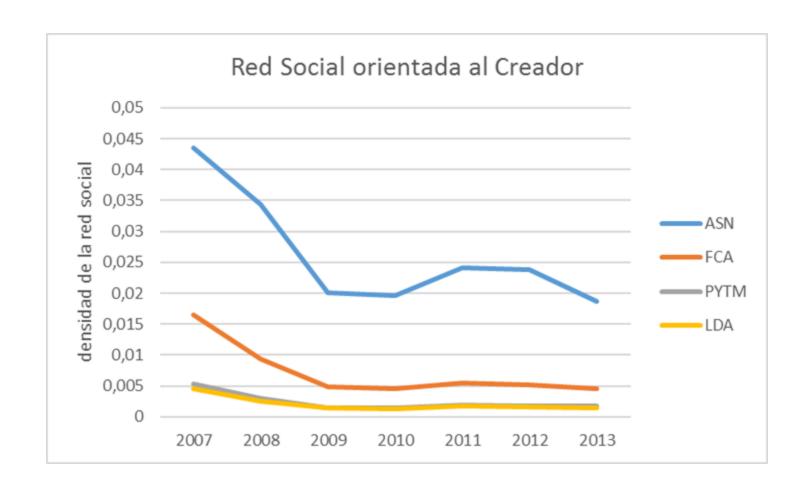


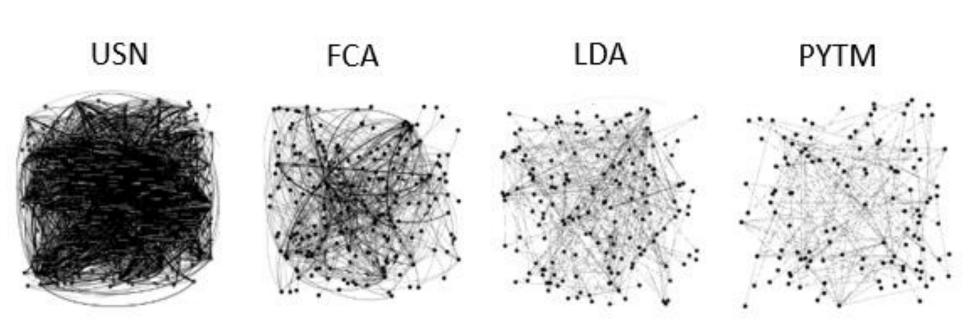
introducción representación en forma de grafo





resultados reducción de densidades



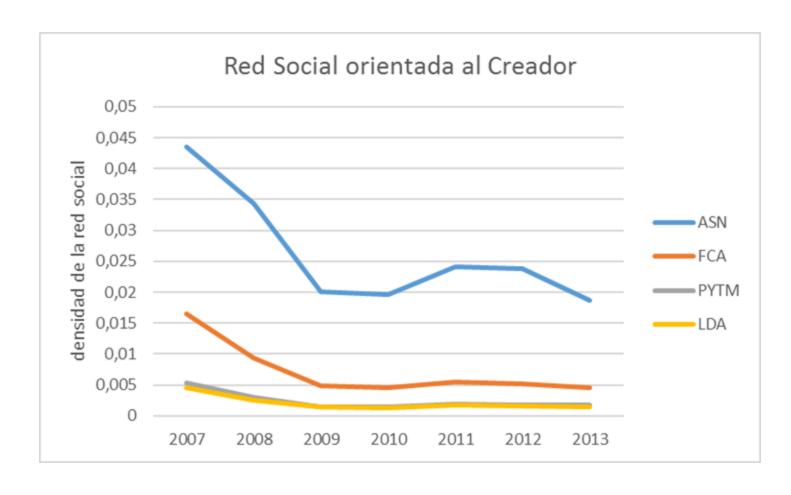


Densidad de la red social

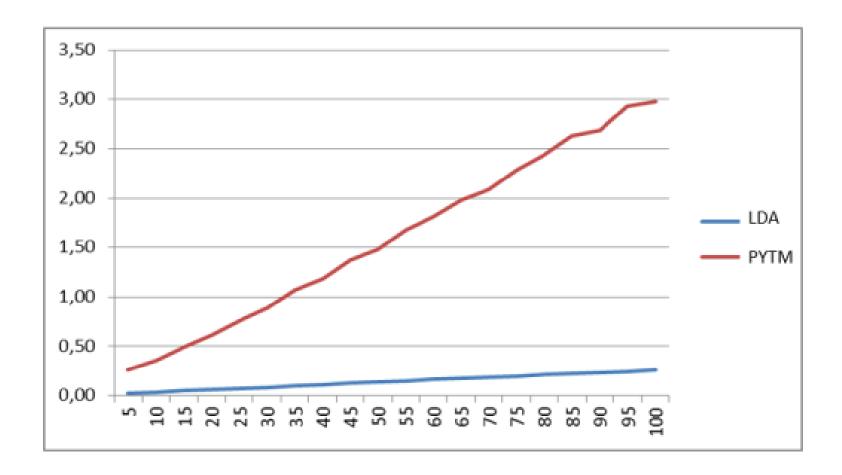
Hay reducción significativa de la densidad de la red



resultados reducción de densidades



Densidad de la red social



Tiempo de Procesamiento



resultados

datos entregados por administradores

usuario	tipo
user1	Α
user24	Α
user56	Α
user289	В
user23	В
user12	С
•••	

66 en total Corresponde al 3% de todos los miembros **Tipo A**: corresponden a los miembros más importantes de la comunidad.

Tipo B: corresponden a miembros menos importantes que los del Tipo A, sin embargo siguen siendo miembros clave.

Tipo C: corresponden a miembros que fueron importantes en el pasado, y que generalmente participaron desde el comienzo de la red social, sin embargo en la actualidad no participan mucho

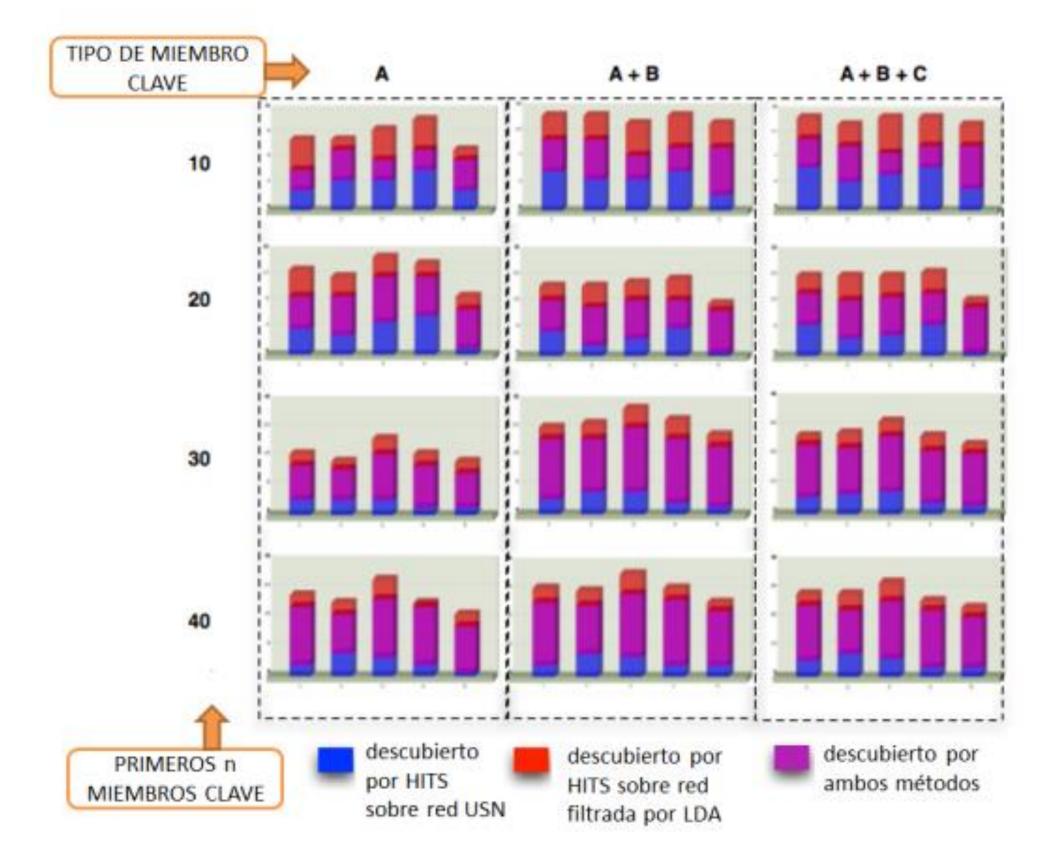
resultados intersección de resultados

Top-10 v/s Miembros de Tipo A 11 Total miembros detectados Detectados sólo por red filtrada Detectados por red sin filtrar y red filtrada Detectados sólo por red sin filtrar 08-2013 09-2013 10-2013 11-2013 12-2013 Detectan miembros de forma diferente



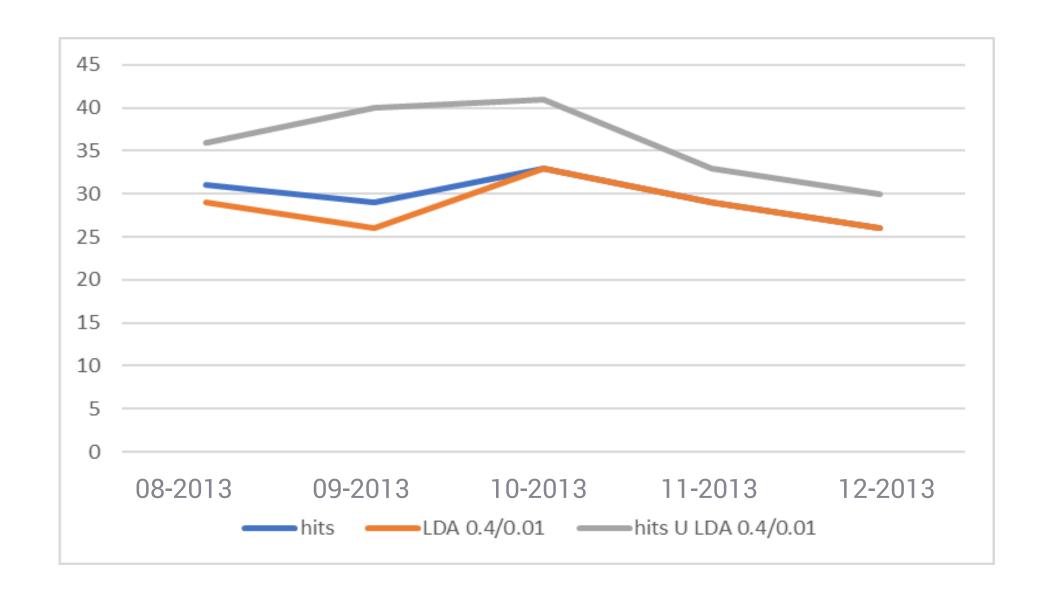
resultados

intersección de resultados





resultados unión de resultados



Resultados de red sin filtrar se complementa con red filtrada



resultados análisis

Cuadro 5.2: Detección de miembros claves para diferentes métodos y configuración de parámetros, para los últimos 5 meses del año 2013, utilizando la

red creator

r <u>ed creator</u>						
Tipo de Experto	Método	08-2013	09-2013	10-2013	11-2013	12-2013
A	USN	18	16	20	18	13
	FCA 0.8/0.2	16	12	20	14	13
	USN ∪ FCA 0.8/0.2	22	19	23	18	16
	PYTM 0.8	18	12	20	15	13
	USN U PYTM 0.8	21	18	25	18	15
	LDA 0.4/0.01	18	13	20	16	15
	$\mathbf{USN} \cup \mathbf{LDA} 0.4/0.01$	21	19	25	19	16
A+B	USN	26	25	29	27	23
	FCA 0.8/0.2	23	20	28	24	21
	USN \cup FCA $0.8/0.2$	32	29	34	30	26
	PYTM 0.8	26	21	29	26	20
	$USN \cup PYTM 0.8$	30	29	36	30	25
	LDA 0.4/0.01	27	22	29	27	22
	$\mathbf{USN} \cup \mathbf{LDA} 0.4/0.01$	31	30	36	31	26
A+B+C	USN	31	29	33	29	26
	FCA 0.8/0.2	27	25	34	26	23
	USN \cup FCA $0.8/0.2$	38	34	40	33	29
	PYTM 0.8	28	26	32	29	26
	$USN \cup PYTM 0.8$	35	40	40	33	30
	LDA 0.4/0.01	29	26	33	29	26
	$\mathbf{USN} \cup \mathbf{LDA} 0.4/0.01$	36	40	41	33	30



resultados análisis

Cuadro 5.5: Rendimiento de la detección de miembros clave de Tipo A+B+C utilizando red USN y la combinación con filtro LDA

Periodo	1	2	3	4	5
Precision USN	0.78	0.73	0.83	0.73	0.65
Precision USN \cup LDA	0.71	0.65	0.79	0.66	0.58
Recall USN	0.48	0.45	0.51	0.45	0.40
Recall USN \cup LDA	0.55	0.55	0.63	0.51	0.46
F USN	0.59	0.55	0.63	0.55	0.50
$F USN \cup LDA$	0.62	0.60	0.70	0.57	0.51



resultados análisis

Se evaluó:

- 2 comunidades
- 12 meses cada una
- 4 topologías de representación de red
- 6 métricas de centralidad
- 2 estrategias de filtrado con semántica latente
- 1 estrategia basada en conceptos
- 20 tamaños diferentes de total de tópicos
- 9 valores diferentes del prior de las distribuciones
- 16 valores diferentes de umbral de filtrado

...Muchos experimentos...

Mejor rendimiento

- Red orientada al creador
- Centralidad medida por HITS (hub)
- LDA, alfa=0.4, y beta=0.1
- 15 tópicos para el período de estudio
- Filtro $\Theta = 0.4 \text{ y p} = 0.01$

Resultados analizados y validados por los administradores.

Resultados están siendo utilizados para la refundación de la comunidad 1, y para las mejoras de la comunidad 2.

agenda de presentación





conclusiones

Se ha logrado crear un nuevo método que otorga información para:

- (a) Identificar a los miembros más importantes en la vida de la comunidad, esto es, aquellos que contribuyen de forma más abundante y creativa a los propósitos de la comunidad, y
- (b) Comprender la dinámica de participación al interior de la comunidad, y cómo los miembros mas influyentes generan participación en torno al propósito de la comunidad.



trabajo futuro

Extensión de las técnicas propuestas a otro tipos de comunidades, como por ejemplo, a las redes de recomendación.

Estudio de falsos influenciadores dentro de estructuras sociales en línea, especialmente en comunidades virtuales de otros tipos.

Extender el análisis incluyendo el estudio de propagación de información dentro de una comunidad virtual.



FIN

Muchas gracias

¿Preguntas?

