

# **SISTEMA DE RECOMENDACIONES BASADO EN CONFIANZA PARA UN ENTORNO DE PUBLICIDAD UBICUA**



Trabajo de Grado

**Juan Camilo Ospina Quintero**

Director: PhD. Ing. Gustavo Adolfo Ramírez González  
Asesor: PhD(c). Ing. Francisco Orlando Martínez Pabón

**Universidad del Cauca**  
**Facultad de Ingeniería Electrónica y Telecomunicaciones**  
**Departamento de Telemática**  
**Línea de Investigación en Aplicaciones sobre Internet**  
**Popayán, Junio de 2016**



# CONTENIDO

	<b>Pág.</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Contexto . . . . .	1
1.2 Problema y motivación . . . . .	3
1.3 Objetivos . . . . .	4
1.3.1 Objetivo general . . . . .	4
1.3.2 Objetivos específicos . . . . .	4
1.4 Metodología . . . . .	4
1.4.1 WP1. Generación de la base inicial de conocimiento . . . . .	4
1.4.2 WP2. Definición de parámetros de redes sociales . . . . .	4
1.4.3 WP3. Diseño del filtro colaborativo basado en confianza . . . . .	5
1.4.4 WP4. Construcción del piloto y evaluación del filtro colaborativo basado en confianza . . . . .	5
1.4.5 WP5. Publicación . . . . .	5
1.5 Estructura del documento . . . . .	5
<b>2 Estado del arte</b>	<b>7</b>
2.1 Metodología de vigilancia tecnológica . . . . .	7
2.2 Trabajos relacionados . . . . .	12
2.2.1 Sistemas de recomendaciones para publicidad ubicua . . . . .	12

2.2.2	Inferencia de confianza . . . . .	14
2.2.3	Sistemas de recomendaciones basados en confianza . . . . .	17
2.3	Brechas . . . . .	19
<b>3</b>	<b>Sistema de recomendaciones basado en confianza</b>	<b>21</b>
3.1	Introducción . . . . .	21
3.2	Apache Mahout™ . . . . .	21
3.3	Componentes . . . . .	22
3.4	Arquitectura . . . . .	24
3.4.1	Tecnologías relacionados . . . . .	25
3.4.2	Descripción de la arquitectura . . . . .	28
3.5	Descripción del dataset de publicidad . . . . .	29
3.6	Modelo matemático . . . . .	31
<b>4</b>	<b>Diseño e implementación del algoritmo de inferencia de confianza</b>	<b>35</b>
4.1	Introducción . . . . .	35
4.2	Parámetros de confianza en redes sociales . . . . .	35
4.3	Graph API de Facebook . . . . .	37
4.3.1	Introducción . . . . .	37
4.3.2	Formato de peticiones y respuestas . . . . .	38
4.3.3	Permisos . . . . .	39
4.3.4	Recolección de información . . . . .	39
4.4	Representación de la información de confianza . . . . .	40

## CONTENIDO

---

4.4.1	Introducción . . . . .	40
4.4.2	Redes de confianza . . . . .	40
4.4.3	Dataset de confianza . . . . .	40
4.5	Algoritmo de inferencia de confianza . . . . .	43
4.5.1	Procesamiento de información de Facebook . . . . .	43
4.5.2	Modelo matemático . . . . .	44
4.6	Restricciones del algoritmo de inferencia de confianza . . . . .	47
<b>5</b>	<b>Experimentación</b>	<b>49</b>
5.1	Primer piloto: sistema de recomendaciones para entornos de publicidad ubicua . . . . .	49
5.2	Segundo piloto: sistema de recomendaciones basado en confianza . . . . .	50
5.2.1	Tecnologías y herramientas . . . . .	51
5.2.2	Arquitectura . . . . .	52
5.2.3	Pruebas offline . . . . .	52
5.2.4	Pruebas online . . . . .	64
<b>6</b>	<b>Conclusiones y trabajo futuro</b>	<b>69</b>
6.1	Contribuciones del trabajo . . . . .	69
6.1.1	Implementación de referencia . . . . .	69
6.1.2	Datasets . . . . .	69
6.1.3	Prototipos funcionales . . . . .	69
6.1.4	Publicaciones . . . . .	69
6.2	Conclusiones . . . . .	70

6.2.1	Conclusiones de la implementación del algoritmo de inferencia de confianza . .	70
6.2.2	Conclusiones de la implementación del TBRS . . . . .	70
6.2.3	Análisis final . . . . .	71
6.3	Trabajo futuro . . . . .	72
<b>A Anexo A. Guía de instalación de Mahout</b>		<b>73</b>
A.1	Preparación del entorno . . . . .	73
A.2	Convertir el proyecto Maven a Eclipse . . . . .	75
A.3	Configurar variables del sistema . . . . .	75
A.4	Importar el proyecto a Eclipse . . . . .	76
<b>B Anexo B. Recolección de Datos - Piloto Smarta</b>		<b>77</b>
<b>C Publicaciones</b>		<b>81</b>
<b>Bibliografía</b>		<b>85</b>

## LISTA DE TABLAS

	<b>Pág.</b>
2.1 Resultado total de referencias . . . . .	8
2.2 Publicaciones por país . . . . .	8
2.3 Estratificación de la confianza . . . . .	15
2.4 Dimensiones de la confianza . . . . .	16
2.5 Brechas . . . . .	19
3.1 Comparación user-user vs item-item . . . . .	27
3.2 Resumen del dataset . . . . .	32
4.1 Parámetros de confianza . . . . .	36
4.2 Parámetros seleccionados . . . . .	37
4.3 Información recolectada de Facebook . . . . .	39
4.4 Resumen del dataset de confianza . . . . .	41
4.5 Estadísticas del grafo de confianza . . . . .	42
5.1 Herramientas utilizadas en el segundo piloto . . . . .	51
5.2 Resultados encuesta de parámetros de confianza . . . . .	53
5.3 Plan de pruebas iniciales para el algoritmo de confianza . . . . .	54
5.4 Resultados pruebas ground truth . . . . .	58
5.5 Test t-student PCA . . . . .	58

## LISTA DE TABLAS

---

5.6	Plan de pruebas para el sistema de recomendaciones . . . . .	59
5.7	Escenarios de prueba para el sistema de recomendaciones . . . . .	61
5.8	Resultados de novedad . . . . .	63
5.9	Test t-student novedad . . . . .	64
5.10	Representación de interacciones del usuario en una escala de 1 a 5 . . . . .	65
5.11	Test t-student precisión online . . . . .	66
5.12	Análisis novedad online . . . . .	67

# LISTA DE FIGURAS

	<b>Pág.</b>
1.1 Descomposición en paquetes de trabajo . . . . .	6
2.1 Publicaciones trust recommender systems . . . . .	9
2.2 Publicaciones trust in social networks . . . . .	9
2.3 Publicaciones trust inference algorithm . . . . .	10
2.4 Publicaciones pervasive advertising . . . . .	10
2.5 Publicaciones social recommender systems . . . . .	11
2.6 Palabras clave . . . . .	11
3.1 Clasificación de los sistemas de recomendaciones . . . . .	23
3.2 Dataset de preferencias . . . . .	24
3.3 Modelo de datos . . . . .	25
3.4 Componentes de un sistema de recomendaciones . . . . .	26
3.5 Precisión para Pearson, Spearman y distancia Euclidiana . . . . .	27
3.6 Diagrama de clases de la solución . . . . .	28
3.7 Diagrama de actividad de la solución . . . . .	30
3.8 Diagrama de componentes de la solución . . . . .	31
3.9 Diagrama de despliegue de la solución . . . . .	32
4.1 Grafo social de la Fundación Innovagen . . . . .	41

## LISTA DE FIGURAS

---

4.2	Representación en JSON del muro y las fotos . . . . .	43
4.3	Representación en JSON del inbox . . . . .	43
4.4	Representación en JSON de las etiquetas en fotos . . . . .	44
4.5	Organización de información JSON . . . . .	44
4.6	Flujo del algoritmo para generar dataset . . . . .	45
4.7	Dataset de confianza . . . . .	46
4.8	Interacciones entre usuarios . . . . .	46
5.1	Arquitectura del primer piloto . . . . .	50
5.2	Arquitectura segundo piloto . . . . .	52
5.3	Encuesta de parámetros de confianza . . . . .	53
5.4	Prueba 1 PCA . . . . .	54
5.5	Prueba 2 PCA . . . . .	55
5.6	Prueba 3 PCA . . . . .	55
5.7	Prueba 4 PCA . . . . .	55
5.8	Prueba 5 PCA . . . . .	56
5.9	Prueba 6 PCA . . . . .	56
5.10	Algoritmo para generar top-10 amigos de Facebook . . . . .	57
5.11	Test de precisión: Escenario 1 . . . . .	62
5.12	Test de precisión: Escenario 2 . . . . .	62
5.13	Test de precisión: Escenario 3 . . . . .	62
5.14	Test de precisión: Escenario 4 . . . . .	63

## **LISTA DE FIGURAS**

---

5.15 Resultados de precisión pruebas online . . . . .	66
---	----



# Capítulo 1

## Introducción

### 1.1. Contexto

Es indudable que la publicidad juega un rol importante en muchas de las actividades de la sociedad actual; grandes compañías y pequeñas empresas recurren a este medio para promocionar sus productos, mejorar sus ventas y promocionar ideas, servicios o cualquier ítem susceptible de ser negociado. Grandes multinacionales como Google generan grandes ingresos a partir de la publicidad e incluso en Colombia en el año 2015 la inversión en publicidad digital alcanzó la cifra de 380 mil millones de pesos <sup>1</sup>.

La publicidad según se define en [1] es “una relación que permite a los proveedores compartir información con los consumidores buscando un beneficio para ambos, pues educa a estos últimos acerca de los productos y servicios disponibles que pueden satisfacer sus necesidades y ayuda a los vendedores a crear un ambiente agradable para su acercamiento con los clientes”.

Dentro del mundo de la publicidad, existe un nuevo paradigma que está tomando fuerza conocido como publicidad ubicua; ésta se define en términos simples como el uso de la computación ubicua con fines publicitarios [2]. La computación ubicua nos presenta un mundo donde las soluciones computacionales y los servicios de comunicación están disponibles en todo momento y en todo lugar [3], permitiendo la construcción de sistemas que pueden adaptarse a los constantes cambios de su entorno y soportando las actividades diarias de sus usuarios [4]. En este orden de ideas, las características esenciales de este paradigma computacional (ubicuidad, interactividad y automatización) son ideales para transformar los esquemas de publicidad tradicionales a través de mecanismos de comunicación simétrica, personalización, medición precisa de audiencia, adaptación al contexto y persuasión.

Con base en lo anterior, la publicidad ubicua promueve una interacción con los usuarios en un nivel más íntimo y personalizado, que permite entregar información relevante en el momento oportuno [5]. En la actualidad, los dispositivos móviles modernos (Smartphones) son claros candidatos para hacer frente a las limitaciones propias de la publicidad tradicional

---

<sup>1</sup>Reporte IAB Colombia, 2015

como la falta de personalización e interacción. Varios trabajos de investigación han abordado sus propuestas de publicidad ubicua apoyándose en las características inherentes de este tipo de dispositivos, como son, por ejemplo, la disponibilidad de información de perfiles, interacción en redes sociales, sensores, entre otras.

Cabe destacar que una de las características fundamentales de la publicidad ubicua es su sensibilidad al contexto. La información contextual, según Schillit [6], “se compone de tres aspectos importantes: dónde estás, con quién estás y con qué recursos cuentas”, como puede apreciarse, divide el contexto en tres partes: localización, naturaleza social, y las herramientas para percibir datos del usuario; en el tema que aborda esta investigación, es de especial interés el contexto social del usuario dado que el 69 % de las personas buscan opiniones para tomar decisiones de compra o para informarse sobre algún producto, incluso el 54 % ha manifestado que compraría un producto con opiniones negativas en internet, si se los recomienda un amigo o familiar cercano [7].

En este sentido, las redes sociales en internet, definidas como: “servicios web que permiten a los usuarios (i) construir un perfil público o semipúblico, (ii) articular una lista de usuarios con los que comparte una conexión y (iii) ver y conectar su lista de conexiones con aquellas creadas por otros” [8], surgen como una alternativa interesante para conocer el contexto social del usuario, debido a que la mayoría de éstas animan a las personas a mantener sus perfiles actualizados con información demográfica (género, edad, ciudad de residencia), preferencias (ideología política, religión) y gustos (tipo de música, de películas, de libros, etc.). Adicionalmente, los usuarios participan activamente en la generación de contenido subiendo fotos, videos, escribiendo opiniones o comentando opiniones de otros, convirtiendo a estas redes en plataformas colaborativas [9]; es en esta información donde se centra la presente investigación, específicamente en la personalización de anuncios publicitarios tomando como base el contexto social del usuario, principalmente sus redes sociales.

De acuerdo a lo anterior, el proceso de sugerir anuncios basados en información contextual social debe ser apoyado en sistemas de recomendaciones (SR), los cuales utilizan técnicas de búsqueda y filtrado de información para guiar a un usuario en un camino personalizado de acuerdo a sus intereses en un gran espacio de opciones [10]. En [11] se presenta una completa conceptualización de los SR y además se los clasifica en tres (3) grupos: 1) Basados en contenido: los cuales utilizan las características de los elementos que le han gustado al usuario en el pasado para realizar las recomendaciones; 2) Filtrado Colaborativo: que se basan en la similitud de usuarios y sus ratings para realizar las recomendaciones; 3) Híbridos: los cuales son una combinación de los dos anteriores. Para esta investigación son de especial interés los SR de filtrado colaborativo, específicamente aquellos que, además de los ratings, agregan una dimensión adicional que incluye la fortaleza de la relación entre usuarios, información que puede derivarse a partir de las redes sociales. Este tipo de sistemas se denominan sistemas de recomendaciones basados en confianza [12]. Es importante recalcar la importancia que tienen estos sistemas en cuanto a la personalización de anuncios en ambientes de publicidad ubicua.

## 1.2. Problema y motivación

Como se mencionó en la sección 1.1, la publicidad juega un papel trascendental en el comercio moderno; empresas grandes, medianas y pequeñas arriesgan gran parte de sus ingresos en estrategias que permitan promocionar sus productos y/o servicios para encontrar los nichos de clientes adecuados a sus necesidades. Tradicionalmente se han utilizado medios masivos como prensa, radio, televisión, publicidad impresa y últimamente publicidad web y móvil; sin embargo hay factores comunes en todos estos métodos: la falta de personalización a los gustos del usuario y a su contexto y la falta de interacción, lo cual ha llevado a los usuarios a tener poco interés en los anuncios que se le presentan y a buscar, cada vez más, productos para bloquear la publicidad. La publicidad ubicua ofrece un nuevo paradigma que busca transformar estos esquemas tradicionales pero trae implícitos una serie de retos que serán analizados de acuerdo a las brechas identificadas en el estado del arte.

De acuerdo al estado del arte se puede concluir que los SR son el enfoque más apropiado para generar recomendaciones personalizadas. Empero, los enfoques tradicionales de los SR usan aproximaciones que solo toman en cuenta el espacio usuario x ítem y en entornos de publicidad ubicua el contexto juega un papel muy importante, por lo tanto, se debe ampliar ese espacio para que incluya el contexto. Si bien la información contextual que se puede obtener de un usuario es bastante amplia, para el tema que ocupa esta investigación es de mayor relevancia el contexto social. En los sistemas de recomendaciones tradicionales que usan enfoques colaborativos, generalmente se trabaja con la similitud de usuarios anónimos y en base a esto se generan las recomendaciones; sin embargo, en entornos publicitarios tiene un alto impacto la opinión de las personas cercanas. Si bien diversas investigaciones han trabajado en la inferencia de confianza a partir de información en redes sociales para determinar el círculo social cercano del usuario, aún la inclusión de este tipo de datos en el razonamiento de los algoritmos de recomendación para entornos de publicidad ubicua es incipiente.

Como ejemplo, se puede pensar en un usuario que llega a una tienda de tecnología buscando cambiar su smartphone. Es posible que la persona tenga en mente algunas características que quiera de su celular, no obstante, la gran variedad de dispositivos dificultan tomar la decisión de compra. En este caso, un sistema de recomendaciones basado en confianza puede disminuir la carga cognitiva del usuario para tomar la decisión sobre qué smartphone comprar. Se puede pensar en un caso hipotético donde el usuario tiene instalada en su celular una aplicación ofrecida por la tienda que, tomando como base los gustos del usuario, proporciona recomendaciones de smartphones privilegiando las opiniones de sus amigos cercanos.

Con base en lo expuesto anteriormente, se plantea la siguiente pregunta para orientar el proceso de investigación del presente trabajo: ¿Cómo procesar la información de confianza que se puede inferir de las redes sociales para incorporarla en el razonamiento de un sistema de recomendaciones para entornos de publicidad ubicua?

## **1.3. Objetivos**

### **1.3.1. Objetivo general**

Desarrollar un sistema de recomendaciones basado en confianza para entornos de publicidad ubicua.

### **1.3.2. Objetivos específicos**

- Definir los parámetros que se utilizarán para inferir la confianza entre usuarios a partir de la información disponible en redes sociales.
- Diseñar un filtro colaborativo basado en confianza para entornos de publicidad ubicua.
- Construir un piloto en el contexto de publicidad ubicua con el fin de evaluar la calidad del filtro colaborativo desarrollado.

## **1.4. Metodología**

La estructura básica de las actividades propuestas para el desarrollo del presente trabajo toma como referencia la descomposición jerárquica WBS (Work Breakdown Structure) sugerida por el Project Management Institute (PMI), a través de su metodología PMBOK (Project Management Base of Knowledge), específicamente en el área de gestión del alcance (Scope Management) [13]. La Figura 1.1 muestra la descomposición en paquetes de trabajo. Los procesos de ingeniería del software requeridos serán abordados a través de un extracto de la metodología UP (Unified Process) [14].

### **1.4.1. WP1. Generación de la base inicial de conocimiento**

- Revisión del estado del arte.
- Síntesis.
- Construcción de la base teórica.

### **1.4.2. WP2. Definición de parámetros de redes sociales**

- Caracterización de perfiles de usuario.
- Identificación y caracterización de patrones de interacción.

## 1.5. Estructura del documento

---

### 1.4.3. WP3. Diseño del filtro colaborativo basado en confianza

- Exploración de frameworks y algoritmos de filtrado colaborativo.
- Formalización matemática de los parámetros de redes sociales dentro de un algoritmo de filtrado colaborativo.
- Implementación del filtro colaborativo basado en confianza.

### 1.4.4. WP4. Construcción del piloto y evaluación del filtro colaborativo basado en confianza

- Definición del entorno del piloto.
- Puesta en marcha del piloto.
- Estructuración del plan de pruebas offline y online <sup>2</sup>.
- Recolección de datos y análisis.

### 1.4.5. WP5. Publicación

- Socialización en seminarios.
- Elaboración de artículos.
- Elaboración de la monografía.

## 1.5. Estructura del documento

### Capítulo 1. INTRODUCCIÓN

Se define el problema, los objetivos de la investigación y la estructura del documento

### Capítulo 2. ESTADO DEL ARTE

Se presenta un resumen de los trabajos relacionados con sistemas de recomendaciones para publicidad ubicua, confianza y sistemas de recomendaciones basados en confianza.

### Capítulo 3. SISTEMA DE RECOMENDACIONES BASADO EN CONFIANZA

Se describe la arquitectura de un sistema de recomendaciones clásico y el proceso de adaptación para incluir confianza.

---

<sup>2</sup>Las pruebas offline son aquellas que se realizan con un conjunto predefinido de datos para tratar de simular el comportamiento de los usuarios con un sistema de recomendaciones y se utilizan para evaluar diferentes algoritmos en un entorno controlado. Por otro lado las pruebas online se realizan con usuarios que interactúan en tiempo real con el sistema y permiten evaluar el rendimiento del sistema a medida que cambia el comportamiento de los usuarios [15]

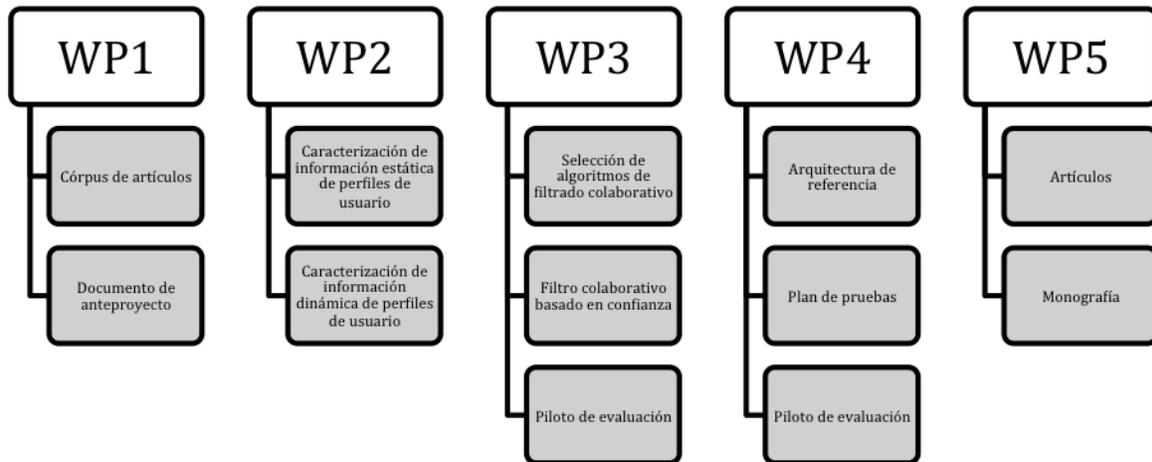


Figura 1.1: Descomposición en paquetes de trabajo. Fuente propia.

#### **Capítulo 4. DISEÑO E IMPLEMENTACIÓN DEL ALGORITMO DE INFERENCIA DE CONFIANZA**

Se describen los parámetros seleccionados para inferir confianza en redes sociales y la implementación del algoritmo utilizando PCA.

#### **Capítulo 5. EXPERIMENTACIÓN**

Se describen los pilotos implementados para la evaluación de los algoritmos implementados, así como los resultados obtenidos

#### **Capítulo 6. CONCLUSIONES Y TRABAJO FUTURO**

Se analizan los resultados obtenidos, se detallan las principales contribuciones y se expone un conjunto de recomendaciones para posibles trabajos futuros.

# Capítulo 2

## Estado del arte

### 2.1. Metodología de vigilancia tecnológica

La revisión del estado actual del conocimiento se abordó a través de una perspectiva de vigilancia tecnológica, la cual se define como "la búsqueda, detección, análisis y comunicación de informaciones orientadas a la toma de decisiones sobre amenazas y oportunidades externas en el ámbito de la ciencia y tecnología" [16]. En términos generales, se abordaron cuatro fases: i) planeación, donde fue definido el objetivo de la vigilancia, se identificaron las fuentes de información más relevantes (ACM Digital Library, IEEE Xplore Digital Library, Google Scholar y Author Mapper de Springer) y las palabras clave que orientaron el proceso de búsqueda de información; ii) búsqueda y captación, a través de la cual se obtiene un corpus compuesto por los papers más relevantes sobre la temática del proyecto; iii) análisis, en donde se realiza un estudio sobre la información del corpus construido en la fase anterior, de acuerdo a los objetivos planteados para el esquema de vigilancia tecnológica. En las tablas 2.1, 2.2 se muestran algunos resultados del proceso, relacionando con la identificación de artículos y países con mayor producción científica en las áreas seleccionadas. En las figuras 2.1, 2.2, 2.3, 2.4 y 2.5 se muestra la cantidad de publicaciones con cada término de búsqueda y su evolución desde el año 2000 hasta el año 2016. Finalmente, en la figura 2.6 se muestran las palabras clave que surgieron durante el proceso de vigilancia tecnológica. Finalmente, en la fase iv) se socializan los resultados obtenidos.

<b>Término de búsqueda</b>	<b>Author Mapper</b>	<b>IEEE Xplore Digital Library</b>	<b>Google Scholar</b>	<b>ACM Digital Library</b>
Trust recommender systems	1449	330	14400	272
Trust in social networks	9532	1744	54700	696
Trust inference algorithm	1832	89	19700	46
Pervasive advertising	780	93	677	64
Social recommender systems	3683	1207	19300	2464

Tabla 2.1: Resultado total de referencias por término de búsqueda. Fuente propia.

<b>País</b>	<b>Publicaciones</b>
Estados Unidos	3133
Reino Unido	1741
China	1549
Alemania	1528
Italia	985
Australia	141

Tabla 2.2: Número de publicaciones por país. Fuente. Author Mapper.

## 2.1. Metodología de vigilancia tecnológica

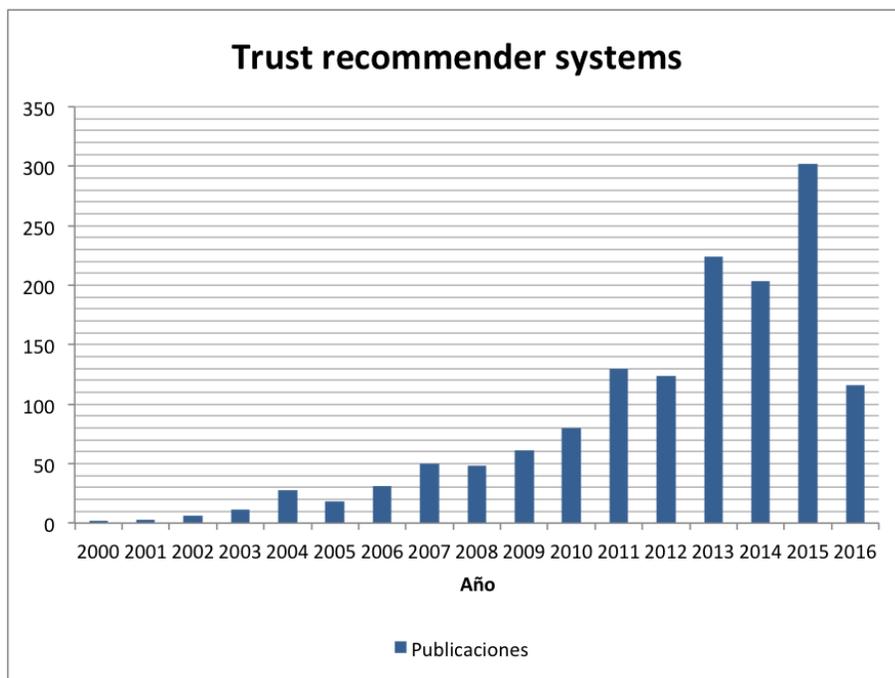


Figura 2.1: Número de publicaciones por año para la frase "trust recommender systems". Fuente. Author Mapper.

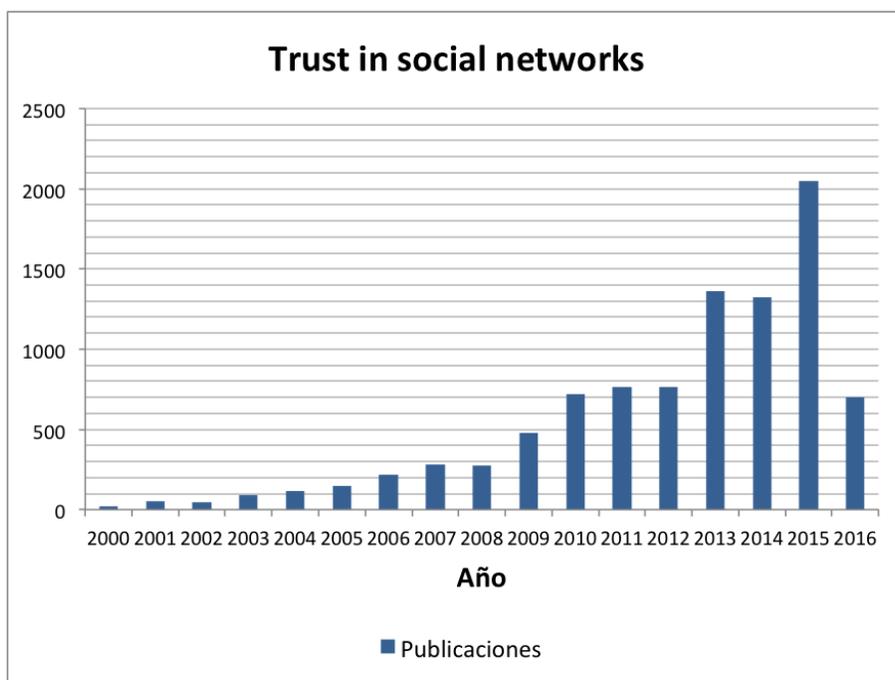


Figura 2.2: Número de publicaciones por año para la frase "trust in social networks". Fuente. Author Mapper.

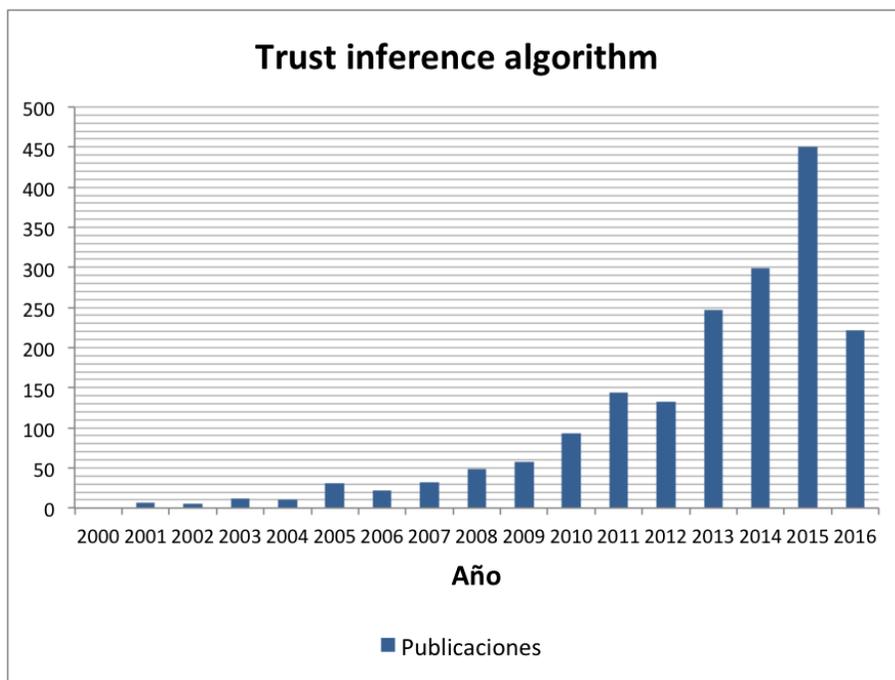


Figura 2.3: Número de publicaciones por año para la frase "trust inference algorithm". Fuente. Author Mapper.

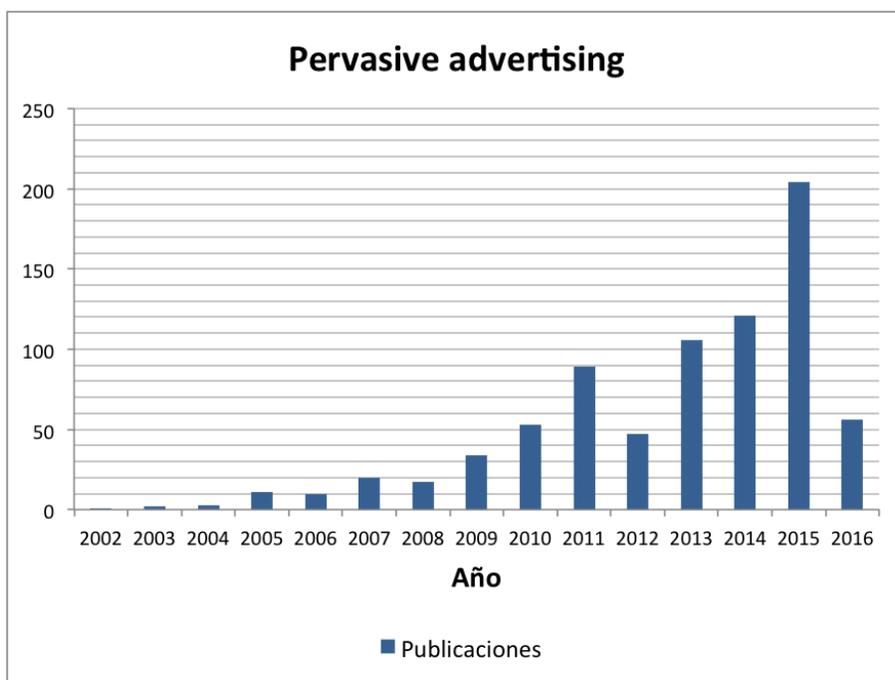


Figura 2.4: Número de publicaciones por año para la frase "pervasive advertising". Fuente. Author Mapper.

## 2.1. Metodología de vigilancia tecnológica

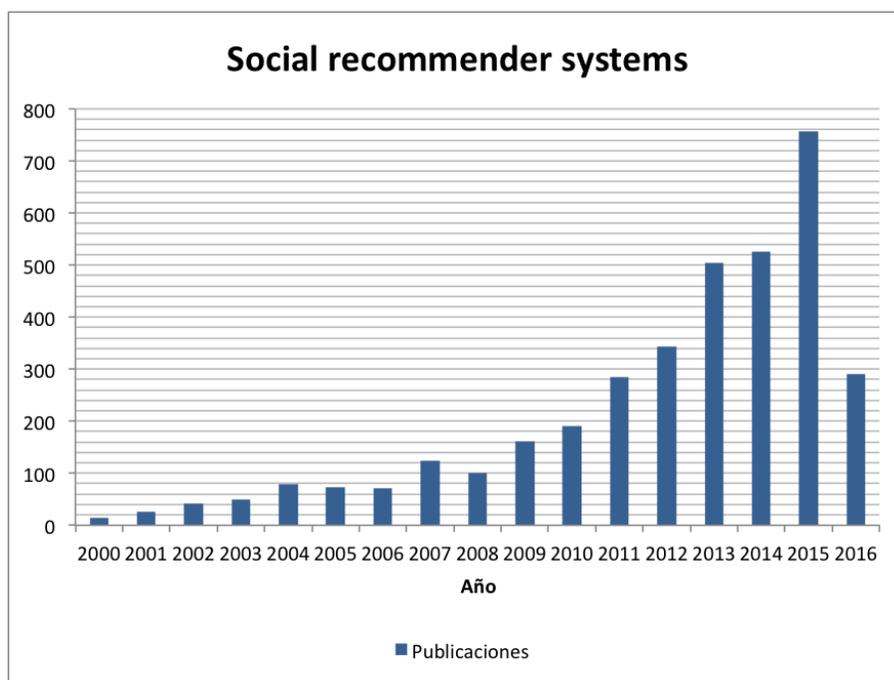


Figura 2.5: Número de publicaciones por año para la frase "social recommender systems".  
Fuente. Author Mapper.



Figura 2.6: Palabras clave asociadas con la vigilancia tecnológica. Fuente. Author Mapper.

De manera general, los resultados del proceso de vigilancia tecnológica se pueden resumir en los siguientes puntos:

- Los sistemas de recomendaciones que incluyen información contextual -en este caso de confianza- son un área de investigación actual y que va en aumento.
- Modelar la confianza en sistemas computacionales ha sido un área de investigación tan grande como los sistemas de recomendaciones. Múltiples definiciones se han abordado en ámbitos tan dispares como métodos de autenticación basados en confianza de agentes artificiales, hasta confianza entre personas en redes sociales.
- Inferir confianza a partir de la información de redes sociales online, es un tema que ha crecido de forma exponencial en los últimos 3 años. La ubicuidad de este tipo de servicios y la cantidad de información suministrada por los usuarios, ha convertido a las redes sociales en el foco de varias investigaciones.
- Hasta el momento muchas investigaciones se han centrado en modelar confianza en una OSN, sin embargo son pocas las que se han implementado en entornos con usuarios reales.
- La publicidad ubicua es un campo reciente de investigación que ha tenido un gran crecimiento en los últimos años debido a la gran variedad de dispositivos interconectados entre sí y al surgimiento del paradigma conocido como Internet de los Objetos (IoT).
- Los países con el mayor número de publicaciones en el campo de sistemas de recomendaciones basados en confianza son Estados Unidos y Reino Unido, sin embargo, en los dos últimos años han aumentado considerablemente las investigaciones en este campo en China, debido principalmente a la existencia de grandes redes sociales como Weibo (222 millones de usuarios) o Qzone (645 millones de usuarios) y la información que se puede obtener de las mismas.
- Si bien existen trabajos que integran la confianza en sistemas de recomendaciones, no se encontró un trabajo que se enfocara en la inferencia de confianza a partir de redes sociales en un entorno de publicidad ubicua.

## 2.2. Trabajos relacionados

De acuerdo al contexto de investigación presentado anteriormente, los trabajos relacionados se agruparon en tres áreas temáticas principales para facilitar su análisis y presentación: sistemas de recomendaciones para publicidad ubicua, confianza y sistemas de recomendaciones basados en confianza.

### 2.2.1. Sistemas de recomendaciones para publicidad ubicua

Existen varias investigaciones relacionadas con sistemas de recomendaciones aplicados a diferentes dominios, incluyendo recomendaciones de libros, películas, noticias e incluso planes vacacionales [17]-[20]. En este sentido, los sistemas de recomendaciones no son un tema

## 2.2. Trabajos relacionados

---

nuevo de investigación, de hecho, varios proyectos han aprovechado los beneficios de estos durante casi dos décadas, para llegar de una manera mucho más personalizada a sus usuarios y facilitar el acceso a información relevante. Este tipo de sistemas han sido utilizados en un sin número de campos, desde la recomendación de noticias como se observa en [20], hasta en la realización de sistemas para la solución de preguntas como Yahoo! Answers [21], incluyendo también los sistemas de recomendación de publicidad. [9] es un excelente ejemplo de ello, sin embargo en este último no se ha tenido en cuenta el uso de publicidad ubicua. Otro de los trabajos relevantes en este campo, es el realizado por [11], el cual presenta una conceptualización muy completa de los sistemas de recomendaciones y un estudio del estado del arte de los mismos. Cabe aclarar que ese trabajo no se centra en un sistema de recomendaciones en especial y por supuesto los sistemas para publicidad ubicua no son considerados. [22] clasifica los sistemas de recomendaciones según el tipo de datos que se necesitan para generar las predicciones; en su trabajo plantea que dado un set de usuarios  $U$  y un set de ítems  $I$ , se distinguen cinco tipos de sistemas: demográficos [23], [24], basados en utilidad [25], basados en conocimiento, y basados en contenido y filtrado colaborativo que son los más comunes [26]-[30]. También se destaca el trabajo realizado en [31], el cual se convierte en un punto de partida importante para todos los trabajos relacionados con los sistemas de recomendaciones, pues combina una serie de artículos producto de investigaciones enfocadas en este tipo de sistemas, pero no específicamente en el contexto de la publicidad, actualmente se encuentra en su segunda edición.

Si bien puede parecer que las investigaciones realizadas hasta el momento en cuanto a sistemas de recomendaciones son suficientes para abordar cualquier dominio, [32] señala que “...el desarrollador de un sistema de recomendaciones para un dominio de aplicación específico debería entender los requerimientos particulares del dominio, sus desafíos y sus limitaciones de aplicación”; a partir de esta aseveración se puede concluir que, aunque hay un núcleo común entre todos los sistemas de recomendaciones, el dominio de aplicación conlleva consigo dificultades propias que pueden convertirse en objeto de nuevas investigaciones.

En ese sentido, hay algunas investigaciones que se han enfocado en caracterizar el dominio de la publicidad ubicua. Al respecto, [2] y [33] ofrecen una conceptualización importante sobre esta área tanto desde el punto de vista de sus características y posibles aplicaciones como en la definición de los retos y posibilidades a los que se enfrenta. Ya en cuanto a entornos reales de publicidad ubicua, [34] presenta una plataforma de publicidad que utiliza información extraída del perfil de Facebook para presentar anuncios en entornos ubicuos, usando una conexión Bluetooth que reconoce al usuario en un contexto específico; sin embargo, este trabajo no utiliza una técnica formal de recomendaciones, simplemente se basa en análisis de intereses del usuario y se limita al uso de Bluetooth como tecnología para identificarlo. Igualmente, en [1] se ofrece un marco de referencia sobre el tipo de información que se puede obtener de los usuarios para entregar anuncios personalizados, pero su enfoque está más dirigido a la caracterización de usuarios que a la recomendación de anuncios.

Al analizar el conjunto de trabajos en área, se puede concluir que todavía existen brechas que se deben superar, siendo las principales: la migración de un espacio usuario-ítem a un espacio usuario-ítem-contexto en entornos de publicidad ubicua, y el considerar las redes

sociales como punto de partida para acceder a la información de usuarios conocidos e inferir el grado de confianza entre estos.

### 2.2.2. Inferencia de confianza

La confianza ha sido un área de investigación muy llamativa tanto desde las ciencias sociales como desde las ciencias de la computación. Actualmente existen varias definiciones sobre el concepto de confianza y cada una varía dependiendo del contexto en el cual se aplique [35], [36]. Desde el punto de vista de las ciencias sociales, [37] argumenta que "la confianza ocurre cuando un sujeto percibe un camino ambiguo, cuyo resultado puede ser bueno o malo, y la ocurrencia de este resultado depende de las acciones de otra persona, si se recorre ese camino se considera que hay confianza, de lo contrario no"; por su parte [38] divide la confianza en 4 aspectos fundamentales: competencias (o capacidades), benevolencia, integridad y previsibilidad. [39]-[41] mencionan la importancia de la confianza en el funcionamiento de una sociedad y, finalmente, [42] argumenta que la confianza es un aspecto necesariamente social. Desde el punto de vista de las ciencias de la computación, hay varias investigaciones que ofrecen definiciones para la confianza, por ejemplo para Mui et al., la confianza es "una expectativa subjetiva que un agente tiene sobre el comportamiento futuro de otro basado en la historia de sus encuentros" [43]; [44] y [45] han propuesto definiciones similares. Por su parte, [46]-[48] presentan la confianza como la base que permite a un sujeto (trustee) usar o manipular recursos de otro sujeto (trustor) o que puede influenciar la toma de decisiones para realizar una compra o utilizar un servicio. Finalmente, Grandison et al. señala que en el contexto de la computación, la confianza se puede entender como: reputación, seguridad, calidad de datos o servicios, credenciales o gestión de riesgos. Para el tema que aborda esta investigación, son de especial interés las definiciones propuestas en [37] y [43] puesto que están relacionadas con la naturaleza de los sistemas de recomendaciones, los cuales toman decisiones basados en el historial de acciones de un usuario y de los usuarios parecidos a este (similitud de usuarios).

Uno de los principales desafíos en este campo de investigación, ha sido cómo representar la confianza en un entorno computacional debido a la cantidad de factores que la afectan, por lo tanto, para utilizarla como un valor numérico entre personas, la definición debe enfocarse y simplificarse de acuerdo a las características del sistema [49], [50]; al respecto, [51] propone un formalismo que se puede utilizar como parte de la estructura de un agente de confianza artificial, racional e inteligente y además proporciona mecanismos para entender la confianza de una forma sencilla. Kosko e Isaka, establecen una escala numérica que permite determinar la confianza entre dos agentes como se puede ver en la tabla 2.3, esta forma de representar la confianza también se propone en [51]-[54]. Otro de los acercamientos a la representación de confianza, es el que se presenta en [55]; en este trabajo se definen 6 dimensiones de la confianza las cuales se encuentran resumidas en la tabla 2.4.

Por otro lado, también es de interés referenciar aquellos trabajos que se han centrado en métodos para calcular la confianza entre dos agentes. Al respecto, son varias las investigaciones realizadas, por ejemplo, en [57], [58] ofrecen un enfoque que permite combinar la

## 2.2. Trabajos relacionados

<b>Valor de confianza</b>	<b>Tipo de relación</b>
+1	Confianza total
> 0.9	Confianza muy alta
0.75 a 0.9	Confianza alta
0.5 a 0.75	Confianza media-alta
0.25 a 0.5	Confianza media-baja
0 a 0.25	Confianza baja
-0.25 a 0	Desconfianza baja
-0.5 a -0.25	Desconfianza media-baja
-0.75 a -0.5	Desconfianza media-alta
-0.9 a -0.75	Desconfianza alta
< -0.9	Desconfianza muy alta
-1	Desconfianza total

Tabla 2.3: Estratificación de la confianza. Fuente. Fuzzy Thinking: The New Science of Fuzzy Logic [56]

información de la reputación individual y la de otras personas, mientras se tiene en cuenta el contexto, esto permite a un agente determinar en quién confiar y para qué se puede confiar. [59] proporciona métodos para calcular grados de confianza en presencia de información conflictiva. En [60] y [61] se presentan trabajos que describen cómo es la reputación y cómo esta afecta a los sistemas de recomendaciones. [62] y [63] utilizan ontologías para expresar la información de confianza y reputación y logran cuantificarla para su uso en algoritmos que permiten tomar decisiones entre dos entidades; esta cuantificación de confianza y sus algoritmos asociados se denominan métricas de confianza. [64] sugiere que se puede calcular la confianza a partir de las interacciones que dos agentes mantienen en una red de confianza, en relación con este trabajo, varias investigaciones en el campo de la sociología han argumentado que entre más transacciones haya entre dos individuos, mayor es la confianza presente entre ellos ([65]-[69]), y [70], [71] manifiestan que se puede calcular la confianza entre personas en internet siempre y cuando las plataformas soporten aspectos sociales. Finalmente, en [54], [57], [72]-[75] abordan los conceptos de confianza global, entendida como la reputación de un agente en la red; y confianza local, entendida como la relación entre dos agentes particulares.

Otro aspecto fundamental en el área de la confianza, son las redes sociales en internet (OSN). [76] manifiesta la relación existente entre la confianza en redes offline y aquella presente en las OSN. Con respecto a las redes sociales offline, uno de los mayores aportes viene de las investigaciones de Granovetter; en su trabajo "The Strength of Weak Ties" [77], pre-

Destino	Representación	Método	Gestión	Computación	Propósito
Agentes	Credenciales	Identificación y autenticación	Uso de una tercera parte confiable	Valores discretos (poco confiable, confiable, neutro)	<i>Humanos</i>
Redes	Firmas digitales	Intercambio de credenciales	<i>Sistemas descentralizados</i>	Valores continuos	Agentes software
Servicios	<i>Historial de interacciones pasadas</i>	Reputación	Mecanismos de votación	<i>Algoritmos para calcular eigenvalores o matrices de adyacencia de grafos</i>	Proveedores de servicios
	<i>Redes de confianza</i>			<i>Confianza local</i>	Canales de comunicación
	Ontologías			<i>Confianza global</i>	Protección de datos
					Encontrar información confiable

Tabla 2.4: Dimensiones de la confianza desde el punto de vista computacional. Aquellas subrayadas son las de mayor relevancia para esta investigación dado el contexto de publicidad ubicua y sistemas de recomendaciones Fuente. propia

senta el concepto de "fortaleza de la relación" y realiza un experimento donde se mide la calidad de las ofertas de trabajo ofrecidas por personas conocidas y por personas cercanas, además determina 4 factores que contribuyen a la fortaleza de la relación: tiempo, intensidad emocional, intimidad y servicios recíprocos. De acuerdo a sus resultados, los lazos débiles (los conocidos) son una mejor fuente de información que los lazos fuertes (los cercanos), debido a que los círculos en los que se mueven son distintos y por lo tanto tienen acceso a información que no se obtendría preguntando únicamente en su grupo cercano. Sin embargo, en un entorno online, este razonamiento no es necesariamente cierto debido a que, si bien puede que el grupo de personas cercanas frecuente los mismos espacios físicos, esto no implica que accedan a los mismos espacios en internet. [78] presenta un estudio sobre preguntas y respuestas en redes sociales, los resultados demuestran que hay poca correlación entre la calidad de las respuestas y la fortaleza de los lazos y que, de hecho, las respuestas proporcionadas por las personas cercanas generalmente se calificaron como más interesantes y con mayor contribución a la solución de la pregunta realizada.

Con respecto a las redes sociales online, varias investigaciones se han enfocado en inferir confianza a partir de una OSN. Otros incluso han propuesto marcos conceptuales a partir de los cuales se puede calcular la confianza que hay en estos entornos. En [79] se crea un dataset a partir de calificaciones que daban los usuarios a parámetros predefinidos de un grupo de amigos escogido al azar; a partir de este dataset proponen una lista de atributos clasificados en 7 categorías y que, según la investigación, son los más relevantes para calcular la confianza entre dos usuarios (En la sección 4.2 se detallan todos los parámetros utilizados para calcular confianza en redes sociales). En [80], los autores sugieren un modelo que permite determinar la fortaleza de la relación a partir de las interacciones entre usuarios tomando como base el conjunto total de interacciones y no solo las que ocurren entre dos personas, además proponen un conjunto de variables que consideran son las más relevantes para la fortaleza de la relación, y prueban este modelo utilizando datos reales de Facebook y LinkedIn. En [81] demuestran que la frecuencia de contacto en Facebook es el factor más relevante a la hora

## 2.2. Trabajos relacionados

---

de determinar la fortaleza de la relación entre dos personas. [82] establece la relación entre confianza y fortaleza de la relación definiendo la última como "confianza basada en conocimiento y experiencias". [83] y [84] siguen un enfoque similar e introducen el concepto de transacciones, que al igual que el los dos casos anteriores, desarrolla modelos que permiten inferir la fortaleza de la relación a partir de unas variables predefinidas. Cabe resaltar que todos los trabajos en los que se hace inferencia de confianza a partir de redes sociales, toman como base un conjunto común de variables que sirven como insumo para las investigaciones futuras en esta área. Finalmente, en [85] se utilizan datos de Facebook recolectados mediante una aplicación llamada iTrust; la información de los usuarios se procesa y se genera una lista de los amigos más confiables. Uno de los principales aportes es la utilización de una técnica estadística llamada Principal Component Analysis (PCA), la cual permite determinar la relevancia de las diferentes variables de confianza de forma personalizada para cada usuario.

Es importante recalcar el aporte de los trabajos anteriormente mencionados específicamente en cuanto a los marcos conceptuales que proponen para entender la confianza dentro de los sistemas computacionales y los mecanismos de inferencia en distintos entornos incluidas las redes sociales, sin embargo, no se encontró un modelo de inferencia de confianza desarrollado específicamente para el contexto de la publicidad ubicua.

### 2.2.3. Sistemas de recomendaciones basados en confianza

En el contexto publicitario es de vital importancia tener en cuenta el entorno social del usuario. Considerando que, de acuerdo con [1], uno de los principales retos de la publicidad es la correcta caracterización de los usuarios, surgen las redes sociales como una alternativa que permite conocer en gran detalle los gustos, disgustos e incluso los intereses de los usuarios. Sin embargo, la información acerca de las preferencias de las personas no es suficiente para ofrecer anuncios lo suficientemente personalizados para atraer la atención del usuario. En ese sentido, combinar un sistema de recomendaciones con dicha información proveniente de redes sociales puede resultar en recomendaciones más relevantes y pertinentes para los usuarios [86].

No son pocos los trabajos que han incluido la confianza dentro de los sistemas de recomendaciones [49], [87]-[98], y en años recientes ha aumentado el interés en esta área de investigación [99]; trabajos como [87] y [100] demostraron que existe una correlación entre confianza y la similitud utilizada en los sistemas de recomendaciones; otros han demostrado que la credibilidad de una recomendación está directamente relacionada con la experticia y la confianza [101]-[109], además se ha demostrado que incluir confianza permite mejorar problemas como el cold-start [110]-[113] y la calidad de las recomendaciones [114], [115].

De acuerdo con [8], los sistemas de recomendaciones basados en confianza se pueden clasificar en dos grandes grupos: 1) aquellos que requieren que la confianza se exprese de forma explícita por los usuarios y 2) aquellos que la calculan de forma automática.

En cuanto al primer grupo, [116] propone reemplazar el mecanismo de predicción de ratings de los sistemas de recomendaciones tradicionales basados en filtrado colaborativo con el uso de un algoritmo capaz de propagar la confianza en una red y de estimar el peso de ésta

para utilizarlo en lugar de la similitud. Adicionalmente, realizan una comparación sobre un dataset y demuestran que utilizar un sistema de recomendaciones basado en confianza es más efectivo en términos de precisión que los sistemas tradicionales, especialmente para usuarios que han proporcionado pocos ratings. [86] presenta FilmTrust, un sitio web que utiliza la confianza obtenida de una red social propietaria para ofrecer recomendaciones de películas. El trabajo se enfoca en cómo crear interfaces para representar las conexiones de los usuarios con base en la información que ellos mismos han proporcionado de forma espontánea. Adicionalmente se propone TidalTrust, un algoritmo para inferir confianza que se utiliza para generar ratings personalizados para cada usuario; los resultados demuestran un rendimiento superior de este algoritmo sobre los algoritmos de filtrado colaborativo basados en similitud. Sobre este primer grupo cabe resaltar la incorporación del dominio de confianza en los algoritmos tradicionales de filtrado colaborativo, sin embargo requieren que los usuarios expresen de forma explícita los valores de confianza con otros miembros de la red y no se enfocan en el contexto de la publicidad ubicua.

En cuanto al segundo grupo, el trabajo más sobresaliente es [117], el cual propone un framework de recomendación para redes sociales basado en filtrado colaborativo y confianza. En esta investigación, son definidos dos métodos para calcular la similitud entre usuarios a partir de la información disponible en Facebook: la primera se basa en la similitud de perfiles (información demográfica, intereses, aplicaciones, etc.), y la segunda se basa en la confianza (patrones de interacción entre usuarios). A partir de estos dos valores se obtiene un único valor que denominan fortaleza de la relación, el cual se utiliza posteriormente en los algoritmos de recomendaciones tradicionales. Además se propone un framework de recomendación donde se especifican los pasos necesarios para generar una recomendación a partir de Facebook. Las principales brechas en este trabajo es que toman como base un dataset ya construido con información de Facebook del 2006, no toman en cuenta la red actual y toda la información adicional que permite obtener, además se enfocan en la recomendación de eventos o aplicaciones y no en el dominio específico de publicidad ubicua.

## 2.3. Brechas

### 2.3. Brechas

Una vez realizado el proceso de análisis del estado actual del conocimiento, se pudieron identificar algunas brechas en las cuales se enmarca esta investigación, estas se consignan en la Tabla 2.5.

Área temática	Brecha a superar
Sistemas de recomendaciones para publicidad ubicua	Migración de un espacio usuario-ítem a un espacio usuario-ítem-contexto en entornos de publicidad ubicua.
	La caracterización del dominio es vital para los sistemas de recomendaciones de filtrado colaborativo, por lo cual es necesario contar con un dataset apropiado. En la revisión del estado del arte no se encontró con un dataset específico para el contexto de la publicidad ubicua.
Confianza	No existe un modelo de inferencia de confianza aplicado a sistemas de recomendaciones en entornos de publicidad ubicua.
Sistemas de recomendaciones basados en confianza	Aunque la publicidad ubicua no es un área nueva de investigación, el uso de confianza y específicamente de sistemas de recomendaciones basados en confianza no ha sido un tema ampliamente explorado.
	Algunas investigaciones se han enfocado en tratar de inferir el grado de confianza entre usuarios a partir de la información de redes sociales como Facebook o Twitter, sin embargo no se integra este tipo de información en un algoritmo de recomendaciones para el dominio de la publicidad ubicua.

Tabla 2.5: Brechas identificadas. Fuente propia.



## Capítulo 3

# Sistema de recomendaciones basado en confianza

### 3.1. Introducción

La presente investigación aborda 2 aspectos principales: inclusión de confianza en sistemas de recomendaciones e inferencia automática de confianza en redes sociales. De acuerdo a lo presentado en el capítulo 2, los sistemas de recomendaciones basados en confianza (TBRS) son una área de investigación de creciente interés, debido a que se ha demostrado que las recomendaciones que provienen de amigos tiene un gran impacto al momento de tomar decisiones de compra.

En este capítulo se describe la arquitectura de un sistema de recomendaciones clásico y el proceso de adaptación para incluir confianza. El capítulo se encuentra organizado de la siguiente manera: en la sección 3.2 se presenta Apache Mahout™, el motor de recomendaciones que soporta la presente investigación; en la sección 3.3 se ilustran los componentes que conforman Mahout; en la sección 3.4 se presenta la arquitectura del sistema propuesto; en la sección 3.5 se describe el dataset que se utilizó en el proceso de recomendación; finalmente, el modelo matemático que soporta al sistema se presenta en la sección 3.6.

### 3.2. Apache Mahout™

Mahout es una librería de machine learning de código abierto soportada por Apache. Este proyecto inició en el 2008 como un subproyecto de Apache Lucene, el cual proporciona algoritmos para búsqueda, minería de texto y recuperación de información. Inicialmente, Mahout estaba enfocado en técnicas de clustering y clasificación, sin embargo después de fusionarse con el proyecto Taste, incluyó algoritmos para implementar filtrado colaborativo.

Una de las características principales de Mahout, es su facilidad de adaptación; su estructura en componentes permite a los desarrolladores de sistemas de recomendaciones adaptar los algoritmos a sus propias necesidades simplemente extendiendo las implementaciones ya existentes, además, al estar escrito en Java, permite la portabilidad a distintos sistemas ope-

rativos. Adicionalmente, cuenta con una amplia documentación que se actualiza de forma constante; el código del proyecto se encuentra disponible para descarga directamente desde <https://github.com/apache/mahout>. En <http://mahout.apache.org/> se encuentra el detalle de los algoritmos implementados e información sobre las últimas versiones y también cuenta con un libro enfocado en los desarrolladores de sistemas de machine learning [118]. Estas características que hacen de Mahout uno de los frameworks más robustos para la implementación de sistemas de recomendaciones, lo han llevado a ser ampliamente usado tanto en entornos académicos como en la industria y fue el framework seleccionado para esta investigación.

En el anexo A se encuentra una guía detallada para instalar y configurar Mahout en Ubuntu.

### 3.3. Componentes

De acuerdo a la clasificación presentada por [11], los sistemas de recomendaciones (RS) se dividen en 3: Basados en contenido (CB), basados en filtrado colaborativo (CF) y los híbridos que son una combinación de ambos (ver figura 3.1). Los sistemas de CF realizan recomendaciones al usuario activo en base a los ítems que a usuarios similares les han gustado en el pasado. La similitud entre usuarios se calcula en base al historial de ratings [31]. Este tipo de sistemas también recibe el nombre de "sistemas de recomendaciones basados en vecindad" o "sistemas basados en correlación de personas" [119]. Por otro lado, los sistemas CB son aquellos que recomiendan ítems similares a aquellos que al usuario activo le han gustado en el pasado [31]. Esta similitud se calcula con base en las características asociadas de los ítems. Por ejemplo, si a un usuario le gustó una película del género comedia, el sistema aprende que puede recomendar otras películas de este género. Este comportamiento si bien puede tener un alto grado de precisión, carece de un componente fundamental para el contexto de la publicidad ubicua: la novedad. Debido a esto el enfoque de filtrado colaborativo se presenta como una buena alternativa para el contexto que abarca esta investigación.

Mahout se enfoca en los sistemas de CF y proporciona un framework flexible, basado en componentes para implementarlos y con un motor de recomendaciones que contiene diferentes implementaciones de los algoritmos más utilizados en este campo.

La mejor forma de comprender las recomendaciones en Mahout es con un ejemplo sencillo [118]. Todo RS necesita datos de entrada en los cuales basarse para realizar la recomendación; estos toman la forma de *preferencias*, las cuales son asociaciones entre ítems y usuarios (un ítem puede ser un libro, una película, una comida, etc). Cada preferencia consiste de un identificador de usuario (userID), un identificador de ítem (itemID) y un valor numérico que representa qué tanto le gusta el ítem al usuario, por ejemplo, en una escala de 1 a 5 (siendo 1 la representación de "no me gusta" y 5 la representación de "me gusta mucho"). Al conjunto de preferencias se le denomina *dataset*. En la figura 3.2 se puede ver la representación de un dataset.

Otra forma de representar el dataset se puede ver en la figura 3.3, esta abstracción se conoce como *modelo de datos (DM)* y es uno de los componentes principales de Mahout; las

### 3.3. Componentes

---

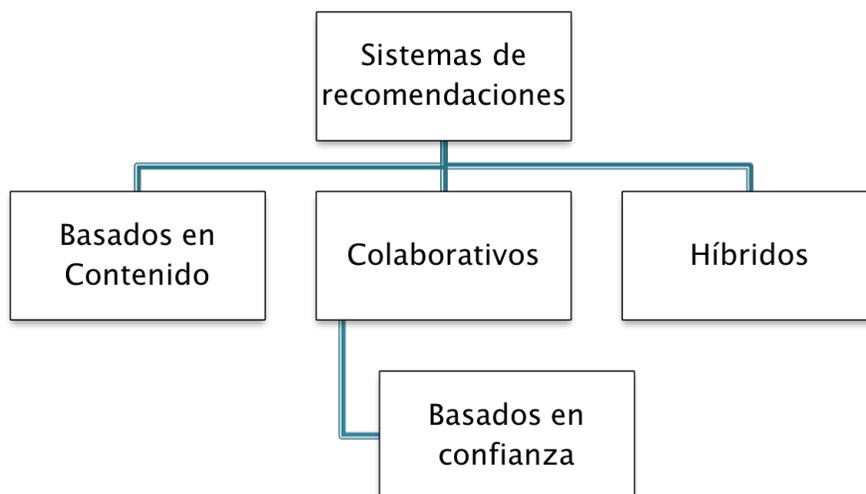


Figura 3.1: Clasificación de los sistemas de recomendaciones. Fuente propia.

líneas continuas representan asociaciones positivas mientras que las punteadas representan las negativas. Analizando el DM de la figura, se pueden obtener las siguientes tendencias:

1. Los usuarios 1 y 5 tienen gustos similares. A ambos les gusta el ítem 1, un poco menos el ítem 2 y no les gusta el ítem 3
2. Los usuarios 1 y 4 tienen gustos similares. A ambos les gusta el ítem 1 y tienen un gusto similar con el ítem 3.
3. Los usuarios 1 y 2 tienen gustos opuestos.
4. Los usuarios 1 y 3 no tienen gustos opuestos, sin embargo no se puede establecer una similitud entre sus gustos.

El segundo componente involucrado en el proceso de recomendación, es la *similitud*. Este componente calcula un valor numérico que determina qué tan parecidos son dos usuarios, utilizando algoritmos como la correlación de Pearson; en el ejemplo anterior, se puede realizar este razonamiento simplemente observando el DM, sin embargo para sistemas reales donde el dataset tiene miles o millones de preferencias, es necesario contar con este tipo de técnicas. Una vez calculada la similitud entre usuarios, aparece el tercer componente: *la vecindad (N)*, el cual se encarga de seleccionar el top-k de usuarios más similares. Para el ejemplo anterior, el top-2 para el usuario 1, estaría conformado por los usuarios 5 y 4. Finalmente, el *recomendador* toma la información de similitud, vecindad y ratings para generar una lista de ítems interesantes y desconocidos para el usuario que solicita la recomendación; siguiendo con el ejemplo anterior, para el usuario 1 la lista estaría conformada por los ítems 104, 105 y 106.

1, 101, 5.0	4, 101, 5.0
1, 102, 3.0	4, 103, 3.0
1, 103, 2.5	4, 104, 4.5
2, 101, 2.0	4, 106, 4.0
2, 102, 2.5	5, 101, 4.0
2, 103, 5.0	5, 102, 3.0
2, 104, 2.0	5, 103, 2.0
3, 101, 2.5	5, 104, 4.0
3, 104, 4.0	5, 105, 3.5
3, 105, 4.5	5, 106, 4.0
3, 107, 5.0	

Usuarios  
Items  
Ratings

Figura 3.2: Dataset de preferencias. Fuente. Mahout in Action [118].

La figura 3.4 ilustra los 4 componentes de un sistema de recomendaciones en Mahout y sus relaciones.

En la siguiente sección se ilustra la arquitectura del sistema propuesto en esta investigación y en la sección 3.6 se presenta el modelo matemático de un sistema de recomendaciones en Mahout y los cambios realizados para incluir confianza.

### 3.4. Arquitectura

Una vez comprendido el funcionamiento de Mahout y sus componentes; y de revisar el estado del arte, se procedió a plantear la arquitectura de un sistema de recomendaciones basado en confianza. En términos de desarrollo de software, la arquitectura se refiere a las representaciones de alto nivel de un sistema, la disciplina para crear tales representaciones y la documentación de las mismas. Estas estructuras permiten entender el sistema y las relaciones entre sus componentes [120].

## 3.4. Arquitectura

---

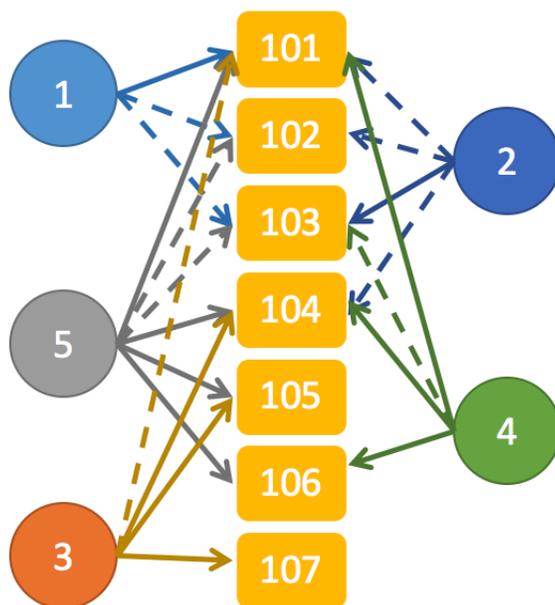


Figura 3.3: Modelo de datos. Fuente. Mahout in Action [118].

Esta sección se divide en dos partes: en la primera se presentan las tecnologías seleccionadas para implementar el TBRS; en la segunda parte se propone una arquitectura de referencia que incluye los siguientes diagramas según lo establecido en el Proceso Unificado de Desarrollo (RUP) [14]:

- Diagrama de componentes: muestra los módulos de sistema, sus dependencias y relaciones.
- Diagramas de actividades: representa los procesos realizados en el sistema para generar las recomendaciones basadas en confianza.
- Diagrama de clases: muestra las clases del sistema, sus atributos, operaciones y sus relaciones.
- Diagrama de despliegue: muestra la distribución física del sistema.

### 3.4.1. Tecnologías relacionados

En la sección 3.2 se presentó Mahout, el framework de recomendaciones que se utilizó para implementar el TBRS. A continuación se detalla el proceso de selección de los algoritmos y tecnologías utilizadas para desarrollar el sistema.

#### 3.4.1.1. Java

Es un lenguaje de programación de propósito general, orientado a objetos y diseñado específicamente para tener tan pocas dependencias como sea posible. Su intención es permitir a

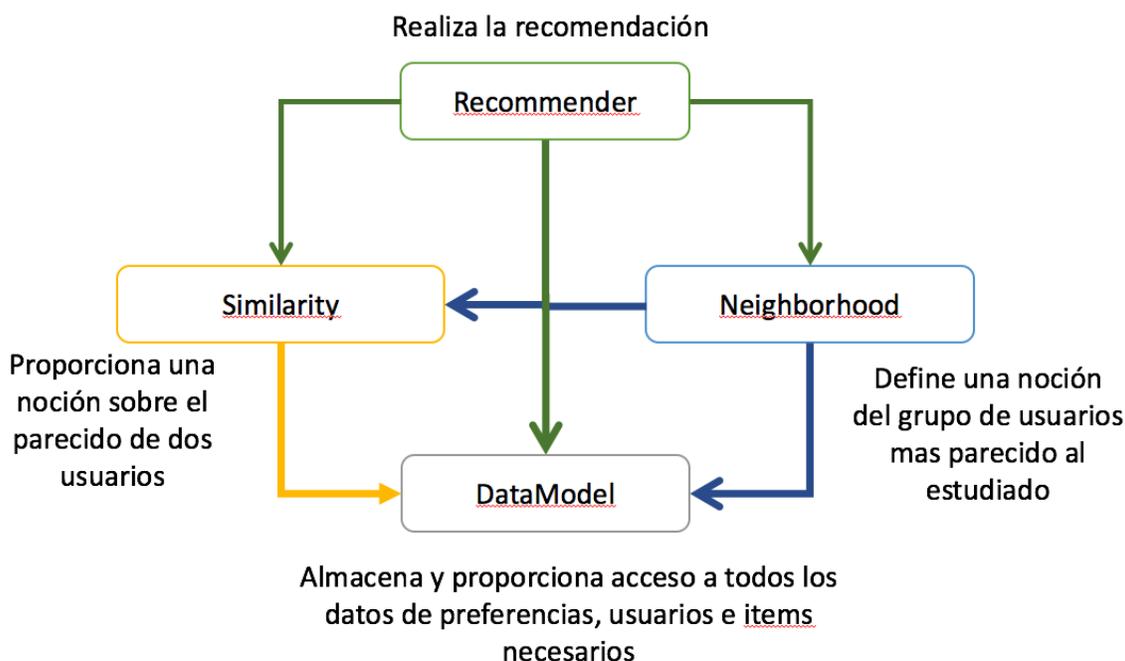


Figura 3.4: Componentes de un sistema de recomendaciones en Mahout. Fuente. Mahout in Action [118].

los desarrolladores "escribir una vez, ejecutar donde sea" (write once, run anywhere), es decir, que el código compilado se puede ejecutar en cualquier plataforma. Como se mencionó anteriormente, es el lenguaje en el cual está desarrollado Mahout.

### 3.4.1.2. Algoritmo de similitud

Una de las primeras decisiones para implementar cualquier sistema de recomendaciones, es la selección del algoritmo de similitud. Inicialmente se debe definir si se utilizará una similitud basada en ítems o basada en usuarios, en la tabla 3.1 se comparan los dos enfoques.

De acuerdo a las características del dataset (más detalles en la sección 3.5), inicialmente se tiene un entorno con más ítems (anuncios publicitarios) que usuarios, sin embargo la base de usuarios es mucho más dinámica que la de anuncios y esta tendencia puede cambiar. Según lo anterior, y para un sistema de recomendaciones clásico, sería preferible un enfoque basado en ítems, empero, para un TBRS las relaciones entre los usuarios son el núcleo del sistema, además, para el contexto de la publicidad ubicua, la novedad es algo deseable. Por estas razones se escogió el enfoque basado en usuarios.

Mahout incluye implementaciones de varios algoritmos enfocados en la similitud de usuarios, entre esos Pearson, Spearman y Distancia Euclidiana. Para escoger el más adecuado se realizaron pruebas de precisión y recall sobre una versión inicial del dataset presentado en la sección 3.5. La precisión y el recall son las métricas más utilizadas para evaluar la calidad de

### 3.4. Arquitectura

un sistema de recomendaciones, la primera se define como la cantidad de ítems recomendados que son relevantes mientras que la segunda se define como la cantidad de ítems relevantes que son recomendados. La figuras 3.5 muestran los resultados de la evaluación.

Criterio	Descripción
Precisión	Para sistemas con muchos más ítems que usuarios, se prefiere un método basado en usuario, en el caso contrario se prefiere uno basado en ítems
Eficiencia	Cuando el número de usuarios excede el número de ítems, un enfoque basado en ítems requiere menos memoria y tiempo para calcular las similitudes
Estabilidad	El enfoque basado en ítems es preferible en entornos más estáticos (cuando la lista de ítems es estática en comparación a la de usuarios)
Novedad	Un enfoque basado en usuarios permite el descubrimiento de diferentes tipos de ítems

Tabla 3.1: Comparación entre los enfoques basados en usuario y aquellos basados en ítem  
Fuente. propia

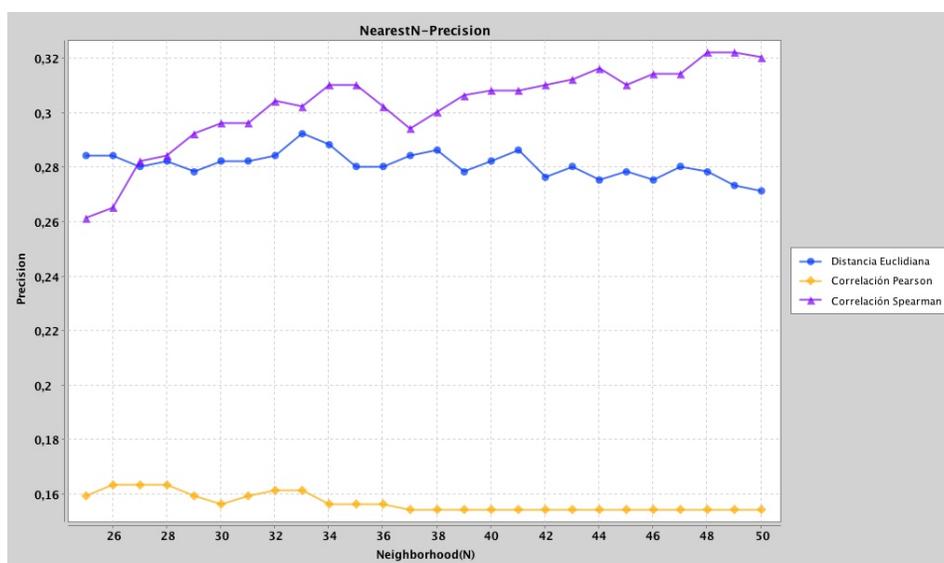


Figura 3.5: Comparación de precisión para los algoritmos de Pearson, Spearman y distancia Euclidiana. Fuente propia.

En base a la figura anterior se decidió utilizar la correlación de Spearman como algoritmo de similitud.

### 3.4.2. Descripción de la arquitectura

#### 3.4.2.1. Diagrama de clases

En la figura 3.6 se presenta el diagrama de clases del TBRS propuesto.

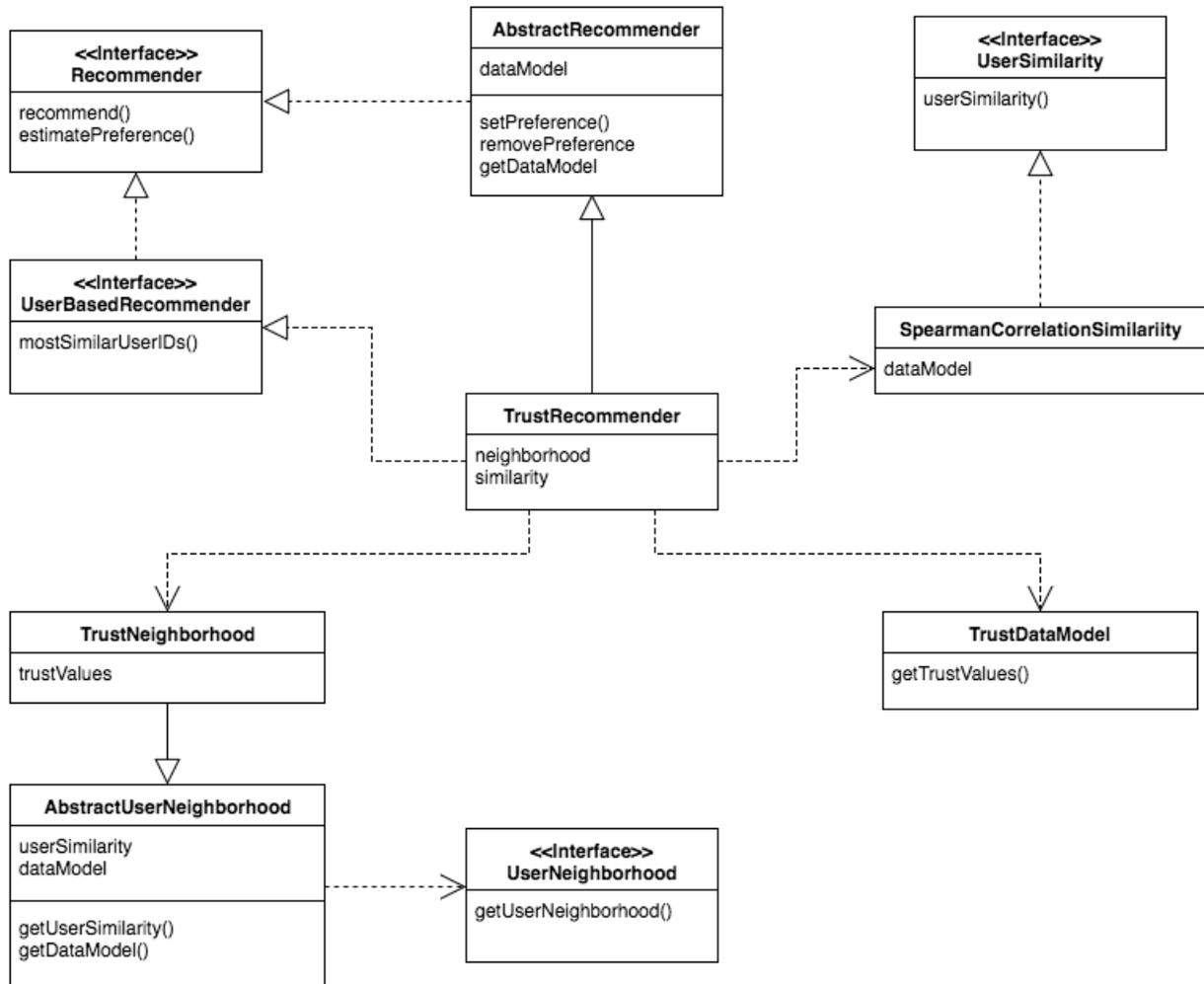


Figura 3.6: Diagrama de clases de la solución. Fuente propia

Como se mencionó anteriormente, una de las fortalezas de Mahout es su facilidad de adaptación. Las clases `TrustRecommender` y `TrustNeighborhood` extienden la funcionalidad ya implementada en Mahout y añaden la dimensión de confianza a los cálculos de vecindad y recomendaciones (`AbstractRecommender`, `UserBasedRecommender` y `AbstractUserNeighborhood`). Para la similitud de usuarios se utiliza la implementación `SpearmanCorrelationSimilarity` que ya viene incluida en Mahout.

## 3.5. Descripción del dataset de publicidad

---

### 3.4.2.2. Diagrama de actividades

En la figura 3.7 se presenta el flujo que sigue el sistema para generar una recomendación utilizando información de confianza; el flujo del lado izquierdo representa un sistema de recomendaciones clásico y en azul se encuentran resaltados los procesos que se añadieron para incluir la información de confianza. En el capítulo 4 se detalla el proceso para generar un dataset de confianza a partir de información de Facebook.

### 3.4.2.3. Diagrama de componentes

En la figura 3.8 se muestran los componentes del sistema y sus relaciones. A diferencia del RS clásico donde el data model solo contiene información de preferencias, en este caso también proporciona información de confianza entre usuarios. Otra diferencia importante, es que el algoritmo de vecindad no toma la información de preferencias del data model sino que accede directamente a la información de confianza y al top-k de usuarios proporcionado por el algoritmo de similitud.

### 3.4.2.4. Diagrama de despliegue

Finalmente, en la figura 3.9 se presenta el despliegue del sistema. Adicional a los elementos mencionados en las secciones anteriores, aquí se presenta el módulo REST API, el cual se implementó para permitir a cualquier tipo de cliente conectarse al sistema. En el capítulo 5 se presentan dos pilotos realizados para evaluar el TBRS propuesto donde los cliente son una aplicación móvil para dispositivos Android y una aplicación para Smart TV.

## 3.5. Descripción del dataset de publicidad

Al pensar en publicidad ubicua, se deben tener en cuenta algunos aspectos importantes sobre la caracterización del dominio. En la introducción al libro "Recommender Systems Handbook", Ricci et al. ([32]) señalan que "...el desarrollador de un sistema de recomendaciones para un dominio de aplicación específico debería entender los requerimientos particulares del dominio, sus desafíos y sus limitaciones de aplicación"; en este sentido, una de las principales tareas para caracterizar el dominio de la publicidad ubicua era tener un dataset de anuncios publicitarios.

Uno de los principales desafíos fue la falta de un dataset apropiado para el contexto que aborda esta investigación. Muchos de los trabajos relacionados utilizan datasets públicos como el de MovieLens <sup>1</sup>, Epinions <sup>2</sup> o Last.fm <sup>3</sup> para probar sus algoritmos, sin embargo ninguno es aplicable al dominio de la publicidad.

Para solventar esta carencia, se construyó una aplicación web donde estudiantes de la Universidad del Cauca podían publicar anuncios, luego, se realizaron diversas sesiones donde,

---

<sup>1</sup>Disponible para descarga en <http://grouplens.org/datasets/movielens/>

<sup>2</sup>Disponible para descarga en <https://snap.stanford.edu/data/soc-Epinions1.html>

<sup>3</sup>Disponible para descarga en <http://grouplens.org/datasets/hetrec-2011/>

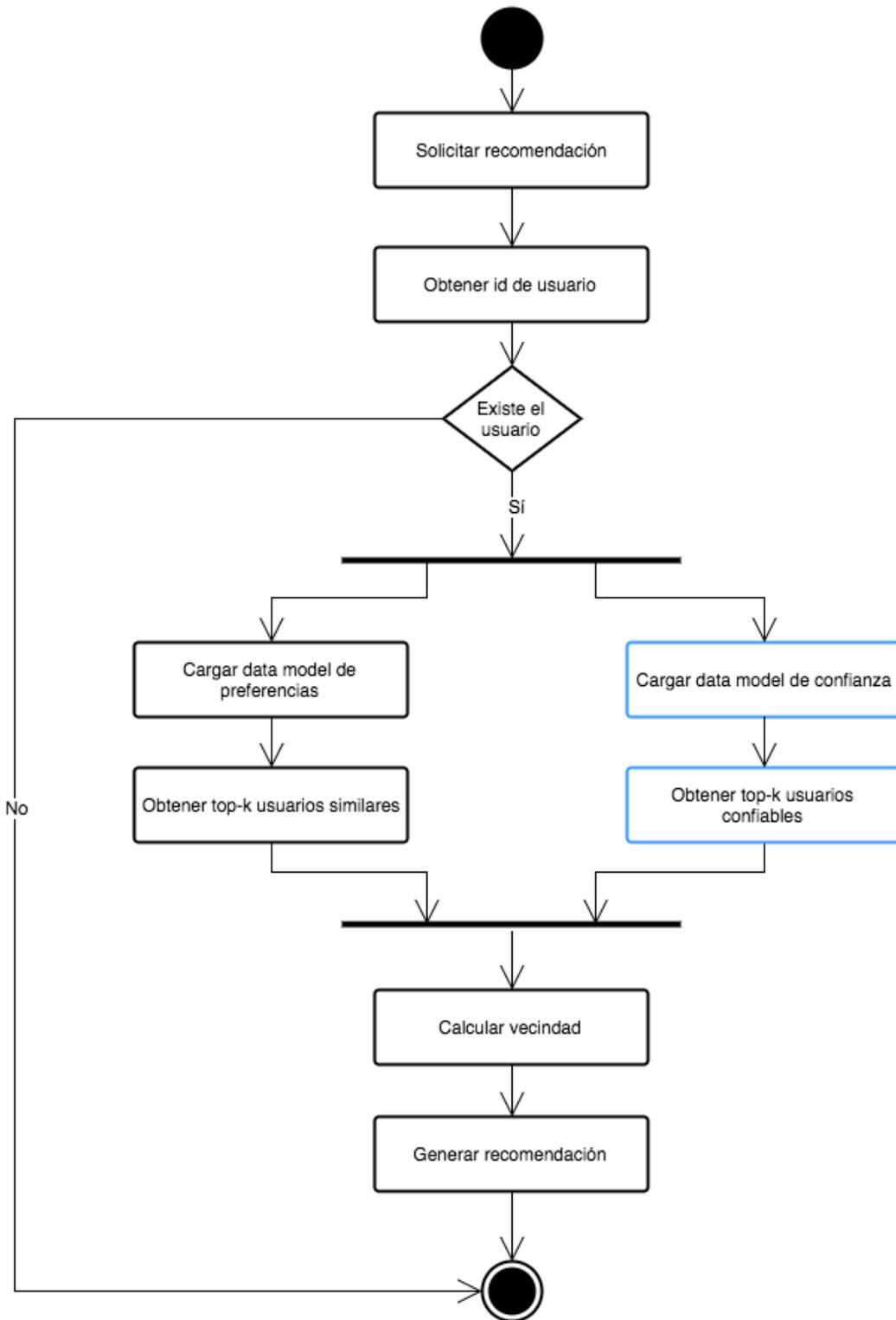


Figura 3.7: Diagrama de actividad de la solución. Fuente propia

### 3.6. Modelo matemático

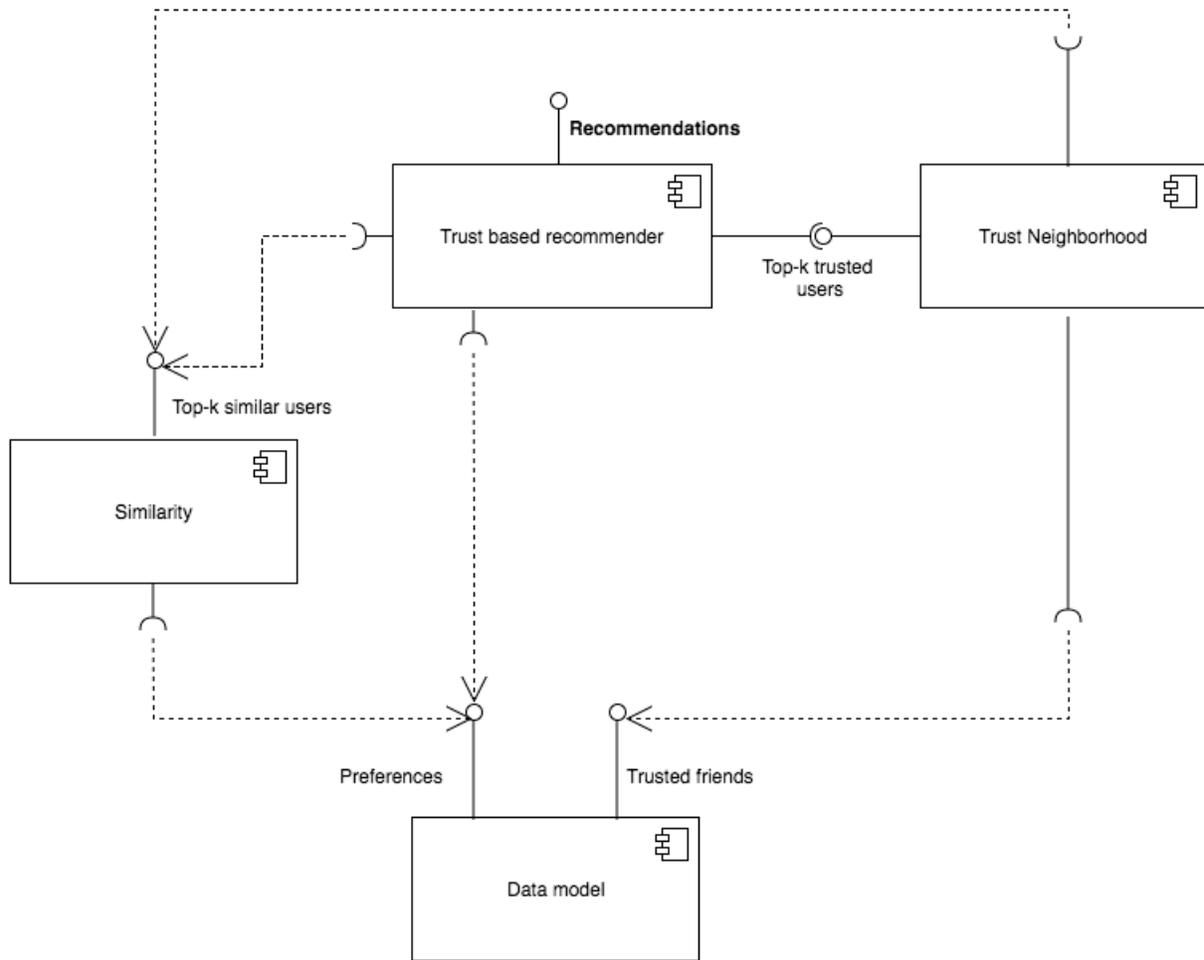


Figura 3.8: Diagrama de componentes de la solución. Fuente propia

mediante la misma aplicación, se les pedía a los participantes calificar al menos 10 anuncios de distintas categorías. La primera versión de este dataset permitió evaluar distintos algoritmos de similitud y escoger el más apropiado para el dominio de la publicidad ubicua (ver sección 3.4.1). En la tabla 3.2 se encuentra el resumen del dataset.

### 3.6. Modelo matemático

El filtrado colaborativo es un algoritmo de recomendación que basa sus predicciones en las calificaciones o comportamiento de otros usuarios en el sistema, es decir, encuentra usuarios cuyas calificaciones previas son similares a las de un usuario particular, y utiliza esa información para predecir lo que le puede gustar a este usuario [121]. Tradicionalmente, los RS de filtrado colaborativo utilizan una métrica de similitud para encontrar la vecindad del usuario (el top-k de usuarios similares) y, basándose en las preferencias de esa vecindad, computa una predicción para un ítem. El algoritmo de filtrado colaborativo está definida por la siguiente

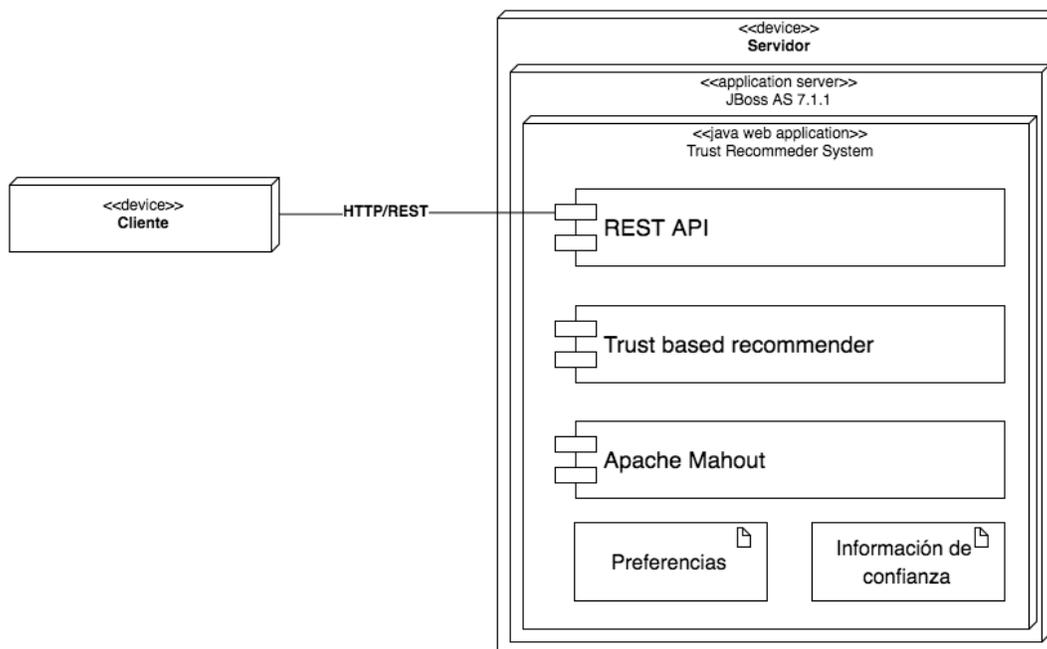


Figura 3.9: Diagrama de despliegue de la solución. Fuente propia

Resumen dataset	
Usuarios	176
Anuncios	127
Ratings	10128

Tabla 3.2: Resumen del dataset. Fuente propia.

función de agregación conocida como la fórmula de Resnick [122]:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * S_{a,u}}{\sum_{u=1}^n S_{a,u}} \quad (3.1)$$

donde  $p_{a,i}$  es la calificación predicha para un usuario  $a$  sobre un ítem  $i$ ,  $\bar{r}_a$  representa la calificación promedio para un usuario  $a$ ,  $n$  es el tamaño de la vecindad y  $S_{a,u}$  es una métrica de similitud entre usuarios, generalmente calculada con Pearson (PCC) o Spearman (SCC). En la práctica, los RS de filtrado colaborativo muestran falencias relacionadas con la naturaleza dispersa de los datos (los usuarios generalmente califican solo una pequeña porción de los ítems disponibles), el problema del cold-start (usuarios nuevos que no han calificado suficientes ítems para calcular una métrica de similitud) o, incluso más importante, la filosofía misma del algoritmo de calcular similitud con usuarios anónimos. En varias investigaciones se ha demostrado que las personas tienden a confiar más en las recomendaciones de personas en las que confían que en aquellas de usuarios anónimos [101]-[104], [106]-[109], [112], [114], [115], [123], [124]; esto es una gran motivación para incorporar confianza en un filtro

### 3.6. Modelo matemático

colaborativo tradicional. De acuerdo a la revisión de la literatura, se identificaron las dos formas más comunes para incluir confianza en el algoritmo de recomendación: media ponderada basada en confianza y filtrado colaborativo basado en confianza.

El primero redefine la estrategia de recomendación al calcular una media ponderada de confianza; en vez de calcular la calificación promedio de un ítem para cada usuario similar, se incluyen valores de confianza  $t_{a,u}$  de tal forma que los usuarios con mayores valores de confianza tengan más peso en la recomendación final.

$$p_{a,i} = \frac{\sum_{u \in n} t_{a,i} * r_{u,i}}{\sum_{u \in n} t_{a,u}} \quad (3.2)$$

[86] propone un algoritmo llamado Tidal Trust, el cual utiliza la ecuación 3.2. De acuerdo a sus resultados, la estrategia no ofreció un beneficio claro sobre el filtrado colaborativo clásico en términos de precisión.

El segundo enfoque introduce una alternativa para el cálculo de la similitud ( $S_{a,u}$  en la ecuación 3.1), la cual intenta inferir un valor ( $T_{a,u}$ ) en base a las relaciones del usuario en una red de confianza (usando técnicas de propagación o agregación) y reemplaza la similitud calculada mediante PCC o SCC por este valor de confianza.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * T_{a,u}}{\sum_{u=1}^n T_{a,u}} \quad (3.3)$$

[87] utiliza esta estrategia donde reemplaza la similitud por un valor obtenido mediante la comparación de perfiles en una red de confianza. De acuerdo a sus resultados, este enfoque permite dar una mayor credibilidad a las recomendaciones.

Aunque ambos enfoques muestran mejoras en las predicciones especialmente para usuarios controversiales y, ayudan a mejorar el problema de cold start, otras investigaciones como [61], [123] y [116] han demostrado que una combinación de los valores de similitud (obtenida mediante PCC o SCC) y los valores de confianza, permite tener sistemas más robustos. Este último es el enfoque escogido para esta investigación.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * WTS_{a,u}}{\sum_{u=1}^n W_{a,u}} \quad (3.4)$$

En este caso  $WTS_{a,u}$  no representa similitud o confianza sino una combinación de ambos valores, como se muestra en la ecuación 3.5:

$$WTS_{a,u} = \alpha S_{a,u} + \beta T_{a,u} \quad (3.5)$$

donde  $\alpha + \beta = 1$ ,  $T_{a,u}$  representa la confianza entre dos usuarios<sup>4</sup> y  $W_{a,u}$  representa la similitud entre usuarios calculada utilizando SCC. El procedimiento que se realiza es calcular la similitud para un usuario y luego calcular el valor de confianza para ese mismo usuario, una alternativa sería tomar primero la lista de usuarios confiables y luego calcular similitud

<sup>4</sup>en el capítulo 4 se describe la implementación del algoritmo de confianza

### Capítulo 3. Sistema de recomendaciones basado en confianza

---

como lo propone [61], sin embargo esto requiere un grupo de usuarios altamente conectado para asegurar que con cada uno se pueda calcular un valor de similitud.

Desde la perspectiva de experimentación, los valores  $\alpha$  y  $\beta$  son pesos que permiten calibrar el algoritmo variando la relevancia que se le da a  $S_{a,u}$  o a  $T_{a,u}$ .

En el capítulo 4 se presenta el algoritmo utilizado para calcular el valor  $T_{a,U}$  entre dos usuarios tomando como base sus interacciones en Facebook.

# Capítulo 4

## Diseño e implementación del algoritmo de inferencia de confianza

### 4.1. Introducción

En el capítulo 3 se presentó la arquitectura y modelo matemático del TBRS propuesto. En este capítulo se describe el proceso realizado para inferir un valor de confianza a partir de la información obtenida de Facebook. El capítulo se encuentra organizado de la siguiente manera: en la sección 4.2 se presenta un conjunto de parámetros obtenidos de redes sociales que son relevantes para calcular confianza entre dos usuarios; en la sección 4.3 se introduce el API para obtener información de Facebook; el grafo de confianza para el TBRS propuesto y el dataset de confianza se ilustran en la sección 4.4; en la sección 4.5 se describe la implementación realizada para calcular confianza a partir de Facebook, la técnica estadística PCA y el modelo matemático del algoritmo de inferencia; finalmente, en la sección 4.6, se detalla el alcance y las restricciones del algoritmo propuesto.

### 4.2. Parámetros de confianza en redes sociales

En la sociedad humana, la confianza depende de una variedad de factores que no son fáciles de modelar en un sistema computacional, por lo tanto, para utilizar la confianza como un valor numerico, la definición debe enfocarse y simplificarse de acuerdo a las características del sistema [49]. En el caso particular que aborda esta investigación, es de especial interés la confianza entre personas en una red social, al respecto [55], [71], [76], [125] indican que la confianza se puede manifestar mediante interacciones entre usuarios siempre y cuando las plataformas mediante las cuales se comuniquen soporten aspectos sociales; en ese sentido, para inferir confianza en una red social, es necesario identificar aquellos parámetros que representan interacciones entre usuarios.

En investigaciones previas ([79], [80], [83], [84]) se han definido algunos conjuntos de parámetros que pueden representar la confianza en una red social. La tabla 4.1 resume los más relevantes encontrados en la literatura y sus correspondientes categorías.

## Capítulo 4. Diseño e implementación del algoritmo de inferencia de confianza

<b>Intensidad</b>
Palabras intercambiadas en el muro Posts en el muro Mensajes intercambiados por inbox Actualizaciones de estado Comentarios en fotos
<b>Intimidad</b>
Número de amigos Días desde la última comunicación Cantidad de palabras de intimidad en el muro o inbox Apariciones juntos en fotos Distancia entre hogares Estado de relación
<b>Duración</b>
Días desde la primera comunicación
<b>Servicios recíprocos</b>
Links intercambiados en el muro Aplicaciones en común
<b>Estructurales</b>
Amigos en común Grupos en común Similitud de intereses Número de redes en común
<b>Soporte emocional</b>
Cantidad de palabras positivas Cantidad de palabras negativas Regalos intercambiados
<b>Distancia social</b>
Diferencia de edad Diferencia de empleo Diferencia educativa Diferencias políticas y religiosas
<b>Demográficas</b>
Género Número de aplicaciones instaladas Número de mensajes en inbox Número de posts en el muro Número de comentarios en fotos Número de comentarios en fotos

Tabla 4.1: Parámetros de confianza en redes sociales Fuente. propia

### 4.3. Graph API de Facebook

---

Son varios los trabajos que se han apoyado en este conjunto de trabajos para sus propias investigaciones. En [117], el framework propuesto por Chen y Fong, utiliza posts en el muro, mensajes intercambiados por inbox, número de amigos, amigos en común y grupos en común. En [78], por ejemplo, utilizan palabras intercambiadas en el muro, días desde la última comunicación, palabras de intimidad en el muro o inbox, días desde la primera comunicación, amigos en común y número de palabras positivas.

Para esta investigación se seleccionó un subconjunto de 7 variables tomando como base el trabajo realizado por Xiaoming Li ([85]), quien propuso un framework para inferir confianza de un usuario A hacia un usuario B. La tabla 4.2 resume los parámetros seleccionados.

Mensajes intercambiados por inbox entre A y B
Likes (de B para A)
Etiquetas realizadas por A donde menciona a B
Etiquetas realizadas por B donde menciona a A
Etiquetas en común (publicaciones donde los dos usuarios aparecen etiquetados)
Comentarios (de B para A)
Publicaciones en el muro de A por parte de B

Tabla 4.2: Parámetros de confianza seleccionados Fuente. propia

En la sección 4.5 se presenta el algoritmo utilizado para inferir confianza a partir de los parámetros seleccionados.

## 4.3. Graph API de Facebook

### 4.3.1. Introducción

La información de perfil disponible en un sistema de recomendaciones tradicional puede proporcionar valores de similitud entre personas, sin embargo investigaciones previas han demostrado que la información de interacción es una de las fuentes más importantes para determinar la confianza en una OSN. Para esta investigación se seleccionó la red social Facebook. Otros trabajos han optado por desarrollar sus propias redes sociales, sin embargo, al ser Facebook la red social más utilizada en la actualidad, permite acceder a información más detallada y precisa que aquella que se le pueda pedir a un usuario en un entorno experimental.

Facebook proporciona a los desarrolladores una API con la cual se pueden consultar datos, publicar nuevas historias, subir fotos y realizar la mayoría de tareas que se pueden hacer directamente desde su plataforma: La Graph API. Su nombre viene de la idea de un "grafo social" el cual representa toda la información de la red social. Este grafo está compuesto por:

- **Nodos:** elementos como usuarios, fotos, páginas y comentarios.
- **Aristas:** las conexiones entre esos elementos, por ejemplo, comentarios de una foto.

- **Propiedades:** información sobre los elementos, por ejemplo el cumpleaños de una persona o el nombre de una página.

Esta API está basada en HTTP, por lo tanto funciona con cualquier lenguaje de programación que tenga una librería HTTP. Además ofrece una herramienta denominada Graph API Explorer, la cual permite realizar peticiones directamente desde un navegador web sin necesidad de implementar ninguna lógica adicional.

### 4.3.2. Formato de peticiones y respuestas

Cada nodo está identificado con un ID único el cual es utilizado para accederlo desde la Graph API. El formato de petición para un nodo es el siguiente:

```
GET graph.facebook.com/{node-id}
```

En el caso de que se quiera acceder a la información de una arista:

```
GET graph.facebook.com/{node-id}/{edge-name}
```

Para acceder a ciertas propiedades de un nodo o arista se utiliza el siguiente formato de petición:

```
GET graph.facebook.com/{node-id}?fields=<first-level>{<second-level>}
```

Donde el parámetro <first-level> se refiere a propiedades o aristas del nodo padre (node-id) y <second-level> se refiere a propiedades o aristas del nodo <first-level>.

En cuanto al formato de respuestas, la Graph API utiliza JSON para entregar la información solicitada. Aunque la respuesta varía dependiendo del nodo o arista que se desee leer, el formato general es el siguiente:

```
{  
  "fieldname": {field-value},  
  ...  
}
```

Por ejemplo para la petición `GET graph.facebook.com/me?fields=name,id`, se obtiene la respuesta

```
{  
  "name": "Juan Ospina",  
  "id": "576381956"  
}
```

Donde `me` representa el nodo al cual se le hace la petición (`me` representa al usuario actual) y `name` e `id` representan propiedades de este nodo.

### 4.3. Graph API de Facebook

---

#### 4.3.3. Permisos

La mayoría de peticiones deben estar firmadas con un "access token", el cual determina los permisos que tiene una aplicación para acceder a cierta información de Facebook. Desde el Graph API Explorer se pueden solicitar permisos para acceder a información como correo electrónico, cumpleaños, fotos, entre otros.

#### 4.3.4. Recolección de información

Para la presente investigación se utilizó el Graph API Explorer para recolectar información de 20 voluntarios de la Fundación Innovagen en Popayán. En total se recolectó la información consignada en la tabla 4.3:

<b>Nodo</b>	<b>Descripción</b>
Me	Nombre e id del usuario
Feed	El feed representa el muro del usuario en Facebook. Se recolectó información de likes, comentarios, autor de la publicación y fecha de las últimas 500 publicaciones.
Fotos en las que ha sido etiquetado	Se recolectó información de likes, comentarios y etiquetas de las últimas 500 fotos en las que aparece el usuario
Fotos subidas por el usuario	Se recolectó información de likes, comentarios y etiquetas de las últimas 500 fotos en las que aparece el usuario
Inbox	Se recolectó información de la cantidad de mensajes que el usuario ha intercambiado con otros mediante el inbox de Facebook.

Tabla 4.3: Información recolectada de Facebook Fuente propia

Es importante aclarar que toda la información recolectada fue con autorización de los usuarios, además, todos los datos fueron anonimizados, de tal forma que solo los ID de los usuarios se utilizaron, además, solo se tuvo en cuenta la cantidad de interacciones y no el contenido.

En la sección 4.4 se describe el dataset generado con la información de Facebook, además en el anexo B se encuentra la guía para generar y descargar la información directamente desde el Graph API Explorer.

## 4.4. Representación de la información de confianza

### 4.4.1. Introducción

Frente al problema de cómo representar la confianza en un sistema computacional, varias investigaciones han abordado múltiples enfoques. Es común encontrar trabajos donde la confianza es sinónimo de reputación, seguridad, calidad de datos o servicios, credenciales, gestión de riesgo, entre otros [44]. Cada uno de estos enfoques tiene distintas maneras de representar la confianza; como esta investigación toma el concepto de confianza como la fortaleza de la relación entre dos personas, la mejor forma de representarla es mediante grafos [49], [55], [92].

En matemáticas y ciencias de la computación, un grafo es un conjunto de objetos llamados nodos que están unidos por enlaces llamados aristas. Generalmente son utilizados para estudiar las relaciones entre unidades que interactúan entre sí [126]. En una red social, como Facebook, las conexiones entre usuarios se pueden representar como un grafo, donde los nodos representan a los usuarios y las aristas la existencia de una conexión de amistad, además, se puede añadir un peso a cada arista; los grafos donde se incluyen pesos son denominados grafos ponderados [127].

### 4.4.2. Redes de confianza

Las aplicaciones web sociales generalmente permiten a las personas expresar su opinión sobre otros usuarios del sistema, a este tipo de aplicaciones se les denomina redes sociales; un tipo especial de estas redes, son aquellas en las que existe un valor de confianza o desconfianza entre usuarios, a estas se les denomina redes de confianza [8]. Un ejemplo de este tipo de redes es CouchSurfing, una plataforma para ofrecer alojamientos; los usuarios pueden crear un perfil e indicar si están dispuestos a ofrecer hospedaje, otros pueden buscar entre los perfiles y determinar quiénes son más confiables; para esto, CouchSurfing proporciona varias herramientas como referencias personales y relaciones de amistad, además, después de un alojamiento, cada usuario puede indicar qué tanto confía o desconfía en el otro creando así una gran red de confianza.

Facebook, por su parte, a pesar de ser una red social con millones de usuarios, no ofrece de forma directa un valor de confianza entre sus usuarios, sin embargo, la información intercambiada entre las personas mediante esta red social es un insumo suficiente para generar una red de confianza entre sus usuarios.

### 4.4.3. Dataset de confianza

Antes de generar el grafo ponderado para Facebook, es necesario construir un dataset con información entre las conexiones de los usuarios y la confianza entre ellos. Para generar este dataset se procesaron los datos obtenidos mediante el Graph API Explorer, las características del dataset se resumen en la tabla 4.4. En la siguiente sección se detalla el proceso para obtener un valor de confianza entre dos usuarios.

#### 4.4. Representación de la información de confianza

Resumen dataset	
Usuarios	20
Conexiones de confianza entre usuarios	175

Tabla 4.4: Resumen del dataset de confianza. Fuente propia.

En la figura 4.1, se muestra el grafo generado a partir del dataset. Para este propósito se utilizó Gephi, una herramienta libre para el análisis de grafos. Gephi permite al usuario interactuar con la representación de los datos, manipular las estructuras, personalizar los nodos y aristas con colores, y aplicar algoritmos para determinar conectividad, modularidad, densidad, diámetro de la red y para detectar comunidades.

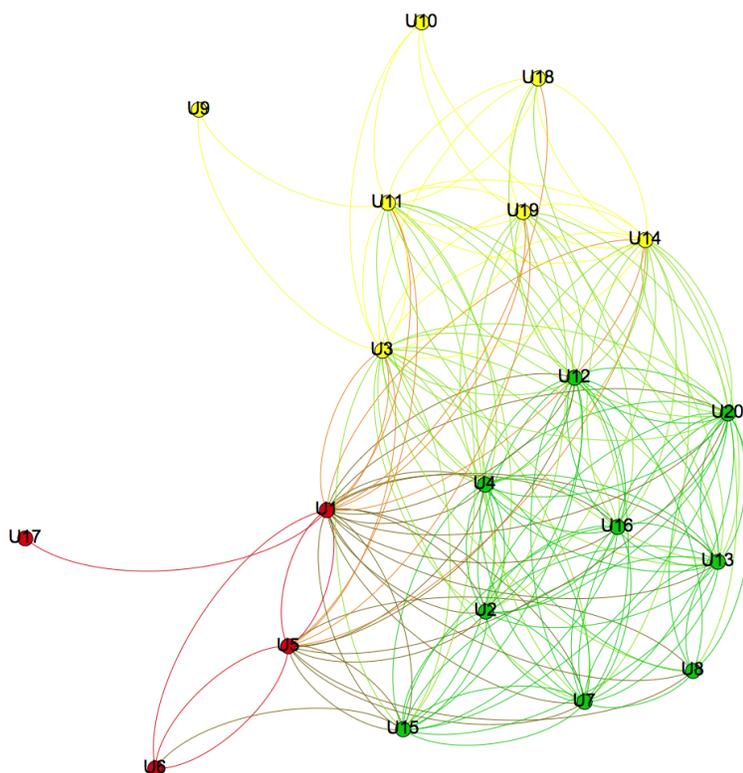


Figura 4.1: Grafo social de la Fundación Innovagen. Fuente propia

En el caso del grafo generado para la Fundación Innovagen, se aplicó un algoritmo de detección de comunidad que permitió determinar la cohesión del grupo; en total se detectaron 3 comunidades representadas por los colores rojo, amarillo y verde. Adicional a la detección de comunidad se calcularon los valores de grado promedio del grafo, distancia promedio, diámetro, densidad y coeficiente de agrupamiento, la tabla 4.5 contiene los resultados.

## Capítulo 4. Diseño e implementación del algoritmo de inferencia de confianza

Medida	Valor	Descripción
Grado promedio	17.5	El grado de un nodo representa el número de conexiones o aristas que este tiene con otros nodos. Un valor de 17.5 en una red de 20 nodos indica que, en promedio, cada usuario está conectado a 17 personas de la red
Distancia promedio	1.5	La distancia de un grafo representa la distancia promedio entre todos los pares de nodos. Entre más grande sea el valor, indica una red con grupos poco cohesionados; entre más cercano a 1 indica una red claramente cohesionada.
Diámetro	3	Representa la distancia máxima entre dos nodos en la red, un grafo con distancia cercana a $n - 1$ donde $n$ representa el número de nodos, indica un grafo disperso; si la distancia es igual a 1 indica que todos los nodos están conectados entre sí.
Densidad	0.589	La densidad mide qué tan cerca está un grafo de estar completo (con todos los nodos conectados entre sí), un valor de 1 representa un grafo completo. En este caso el 58 % de los nodos están conectados entre sí
Coefficiente de agrupamiento	0.705	El coeficiente de agrupamiento es una medida directa de la cohesión de un grupo, en este caso se tiene un grupo con 70.5 % de cohesión

Tabla 4.5: Estadísticas del grafo de confianza. Fuente Gephi.

En la sección 5.2 se presenta un piloto realizado en el marco del proyecto doctoral a cargo del PhD(c) Francisco Martínez denominado "SMARTA: Sistema de recomendaciones soportado en un esquema de cooperación Smart TV - Smartphone para entornos de publicidad ubicua" [128], un sistema diseñado para proporcionar recomendaciones grupales e individuales en un entorno de digital signage. Uno de los requerimientos para llevar a cabo el piloto, era tener un grupo cohesionado de usuarios donde la mayoría de personas tuvieran valores de confianza entre sí; el análisis del grafo realizado con Gephi permitió comprobar que los voluntarios de la Fundación Innovagen cumplían con ese requerimiento y además, mediante el análisis de comunidad, permitió la conformación de grupos de 3 personas (otro de los requerimientos para la experimentación) donde la mayoría de sus miembros tuvieran valores de confianza entre sí.

## 4.5. Algoritmo de inferencia de confianza

### 4.5. Algoritmo de inferencia de confianza

#### 4.5.1. Procesamiento de información de Facebook

La información obtenida de los usuarios mediante el Graph API Explorer debe ser procesada antes de ser incluida en el algoritmo de inferencia de confianza. Por cada usuario se tienen 4 archivos en formato JSON representando la siguiente información: publicaciones del muro y fotos subidas por el usuario (ver figura 4.2), inbox (ver figura 4.3) y fotos en las que aparece etiquetado (ver figura 4.4).

```
"data": [
  {
    "from": {
      "id": "576381956",
    },
    "type": "status",
    "id": "576381956_10152365158096957",
    "likes": {
      "data": [
        {
          "id": "1452362482"
        },
        {
          "id": "776853469"
        },
        {
          "id": "100004291026608"
        },
        {
          "id": "534938860"
        },
        {
          "id": "1078940048"
        }
      ]
    },
    "comments": {
      "data": [
        {
          "from": {
            "id": "1038213803"
          },
          "id": "10152365158096957_10152365224251957"
        },
        {
          "from": {
            "id": "100001284436074",
          },
          "id": "10152365158096957_10152365270076957"
        }
      ]
    }
  }
],
```

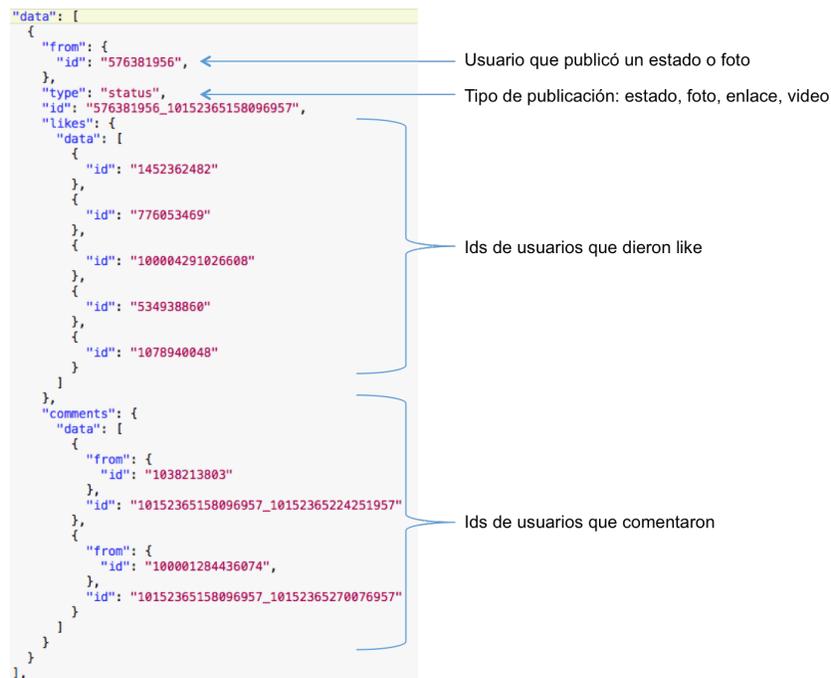


Figura 4.2: Representación en JSON de la información del muro y las fotos subidas por el usuario. Fuente propia

```
{
  "data": [
    {
      "message_count": 51080,
      "recipients": [
        "10203097935702790",
        "576381956"
      ]
    },
    { },
    { },
    { }
  ]
}
```

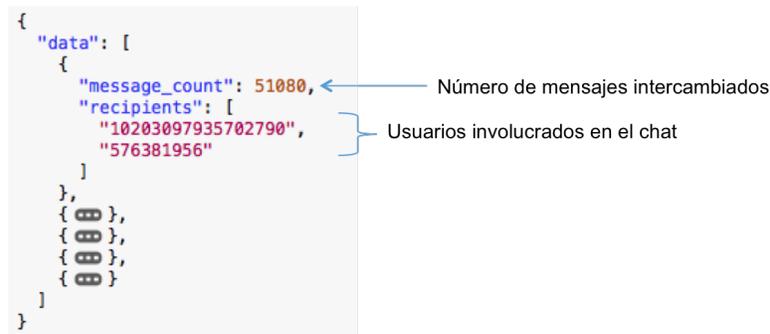


Figura 4.3: Representación en JSON de los mensajes intercambiados entre dos usuarios por inbox. Fuente propia

```

{
  "data": [
    {
      "from": {
        "id": "1163156837"
      },
      "tags": {
        "data": [
          {
            "id": "100003727964149"
          },
          {
            "id": "576381956"
          }
        ]
      }
    }
  ]
}

```

← Id de usuario que publicó la foto

Ids de usuarios etiquetados

Figura 4.4: Representación en JSON de las etiquetas en fotos. Fuente propia

Los archivos de cada usuario se almacenan en una carpeta identificada con el ID del usuario, a su vez, esta carpeta se almacena dentro de otra denominada JSON como se ilustra en la figura 4.5.

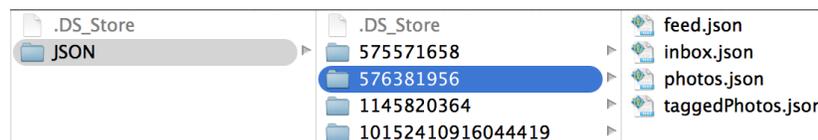


Figura 4.5: Organización de información JSON de la Fundación Innovagen. Fuente propia

En la figura 4.6 se presenta el flujo del algoritmo que genera el dataset de confianza tomando como base la información de la carpeta JSON. La matemática asociada al proceso resaltado en azul se presenta en la sección 4.5.2. Finalmente en la figura 4.7 se muestra una representación de los datos procesados.

### 4.5.2. Modelo matemático

Una vez procesada la información de los usuarios, se definió un modelo matemático. El objetivo principal es proponer una ecuación que combine los distintos parámetros para calcular un valor de confianza que luego pueda ser integrado en el sistema de recomendaciones propuesto en el capítulo 3. A partir del trabajo de Chen et al. [117] y de acuerdo a la teoría de utilidad basada en atributos múltiples (MAUT), se define que la confianza de un usuario A hacia un usuario B se calcula de la siguiente forma:

$$T_{A,B} = \alpha \frac{\sum IM}{totIM} + \beta \frac{\sum Likes}{totLikes} + \gamma \frac{\sum TagsTo}{totTagTo} + \delta \frac{\sum TagsFrom}{totTagsFrom} + \epsilon \frac{\sum Cotags}{totCotags} + \theta \frac{\sum C}{totC} + \phi \frac{\sum WP}{totWP} \quad (4.1)$$

donde  $\alpha + \beta + \gamma + \delta + \epsilon + \theta + \phi = 1$ , IM representa los mensajes intercambiados; C, los comentarios; WP, las publicaciones en el muro; TagsTo, las etiquetas de A donde menciona

## 4.5. Algoritmo de inferencia de confianza

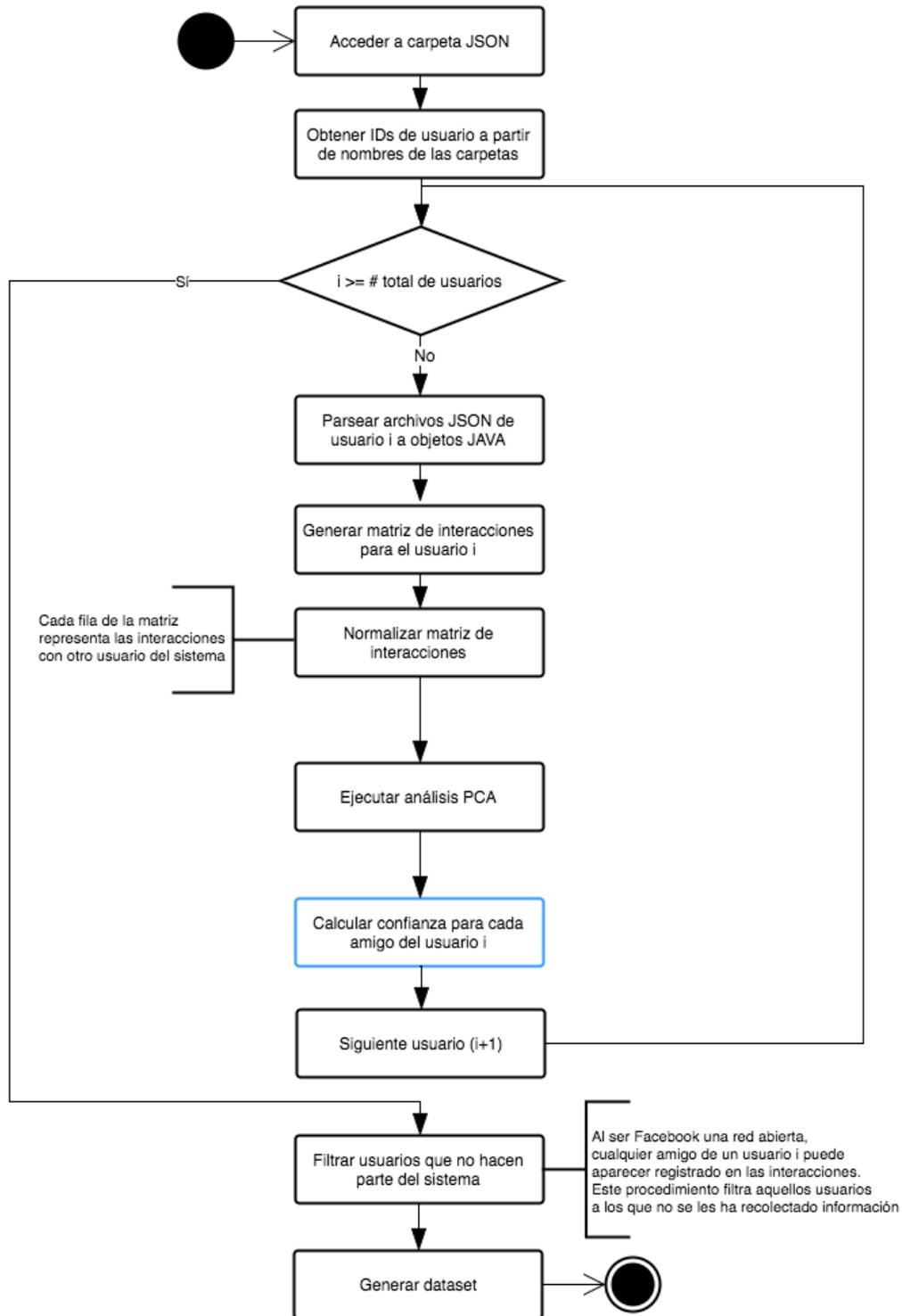


Figura 4.6: Flujo del algoritmo para generar el dataset de confianza. Fuente propia

USER A	USER B	TRUST
100005580760319	10152841976922974	1
100005580760319	10204369688818476	0.5388812
100005580760319	10152582692173285	0.5315525
100005580760319	721886537902337	0.5315525
100005580760319	10205349389877156	0.2873432
10152331830230869	10152563928748668	0.9676966
10152331830230869	10152833553273771	0.7200718
10152331830230869	10152582692173285	0.5278163

Figura 4.7: Dataset de confianza generado. Fuente propia

a B; TagsFrom, las etiquetas de B donde menciona a A; Cotags, las etiquetas en común, y tot es el prefijo que indica la cantidad total de interacciones para un parámetro particular, por ejemplo, totIM representa la cantidad total de mensajes intercambiados por un usuario con todos sus amigos. Esta normalización tiene sentido cuando consideramos la frecuencia de contacto entre dos individuos como un indicador claro de confianza [81], [129]. Un ejemplo se muestra en la figura 4.8. Daniel y Jhoana son amigos de Juan y se desea medir la confianza entre Juan y sus amigos. Las flechas representan 3 tipos de interacciones (tags, mensajes y comentarios) y los números denotan la cantidad de interacciones. Si tomamos en cuenta las interacciones como se muestran en la figura, se puede deducir que Juan confía más en Daniel que en Jhoana, sin embargo, eso puede deberse a que Daniel es un usuario más activo que Jhoana, por lo tanto se debe considerar el nivel de actividad de ambos usuarios. Tomando como referencia los mensajes intercambiados y suponiendo que Daniel tiene un  $totIM = 50$  y Jhoana un  $totIM = 8$ ; al normalizar se puede observar que, mientras para Daniel los mensajes intercambiados con Juan apenas representan el 20 % de sus mensajes, para Jhoana representan aproximadamente el 63 %, lo cual demuestra el impacto del usuario activo. En trabajos como el realizado por Li [85] se utiliza un enfoque similar para la normalización.

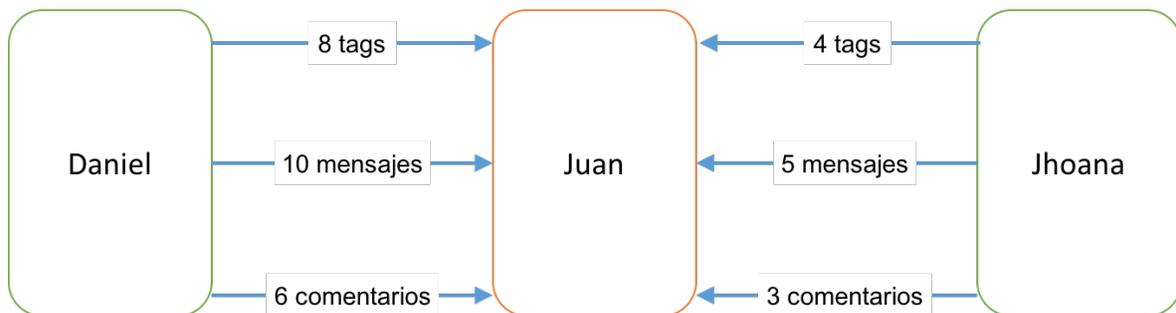


Figura 4.8: Interacciones entre usuarios. Fuente propia

Para la estimación de los valores  $\alpha, \beta, \gamma, \delta, \epsilon, \theta$  y  $\phi$  se consideraron dos enfoques: 1) se asigna un valor igual a cada peso y, mediante experimentación, se ajustan estos valores; 2) Análisis de componentes principales (PCA). El primero no cuenta con un sustento matemático.

## 4.6. Restricciones del algoritmo de inferencia de confianza

---

tico fuerte, además es muy subjetivo y no se puede garantizar que los pesos que aplican a una persona, apliquen a todo el conjunto de usuarios. PCA, por su parte, es una herramienta matemática utilizada para transformar un número de variables correlacionadas en un conjunto menor de variables sin correlación llamadas componentes principales (PCs) [130]. Cada PC es una combinación lineal de las variables originales; durante la transformación, la mayoría de la información original se conserva. Para el caso que concierne esta investigación, cada PC está compuesto de distintos tipos de interacciones cada una con un peso diferente, lo que permite una mayor universalidad de la técnica y por lo tanto es la seleccionada para esta investigación.

A continuación se presenta el proceso para obtener un valor de confianza utilizando PCA:

1. Cálculo de la matriz de interacciones: en la sección 4.5.1 se explicó el proceso para generar la matriz de interacciones para un usuario  $i$ . En términos matemáticos se define que para cada  $j \in \widehat{F}_i$ , donde  $F_i$  representa los amigos de  $i$ , se tiene un vector de interacciones  $\vec{s}_{i,j} = \{s_{i,j}^k\}$ ,  $k = 1, 2, \dots, 7$ . Si se considera cada vector  $\vec{s}_{i,j}$  como una fila de una matriz, existen  $|F_i|$  filas y 7 columnas (una por cada parámetro de interacción). Esta matriz se normaliza de acuerdo a lo expuesto anteriormente. Se utilizará  $Z$  para identificar esta matriz.
2. Calcular la matriz de covarianza: el propósito es encontrar las relaciones entre los conjuntos de parámetros. A partir del resultado obtenido, se escogen los  $m$  eigenvectores ( $u_1$  a  $u_m$ ) que correspondan al top  $m$  de eigenvalores ( $\lambda_1$  a  $\lambda_m$ ).
3. Calcular valor de confianza:

$$T_{i,j} = \sum_{i=1}^m \frac{\lambda_i}{\sum \lambda_m} * u_i * Z_{i,j} \quad (4.2)$$

Para esta investigación, el algoritmo implementado selecciona el top  $m$  eigenvectores que representen el 85 % de la matriz original de interacciones.

4. Normalizar valor de confianza: para facilitar la inclusión de este valor en el sistema de recomendaciones, se decidió normalizarlo en una escala  $[-1, 1]$

$$Trustscore = \frac{2(T_{i,j} - PCA_{min})}{PCA_{max} - PCA_{min}} - 1 \quad (4.3)$$

En la sección 5.2 se presentan los resultados de las pruebas de ground truth aplicadas a la versión del algoritmo de confianza que utiliza PCA (ver ecuación 4.3).

## 4.6. Restricciones del algoritmo de inferencia de confianza

Desde la versión 2.0 de la Graph API se limita el acceso a cierta información del usuario desde aplicaciones de terceros, sin antes pasar por aprobación de Facebook, aun cuando el

## **Capítulo 4. Diseño e implementación del algoritmo de inferencia de confianza**

---

usuario otorgue los permisos necesarios. Esta aprobación requiere tener la aplicación en producción y al usarla en modo de prueba únicamente permite su uso para los desarrolladores registrados en la plataforma. Debido a esto se suspendió el desarrollo de una aplicación propia y se utilizó el Graph API Explorer. Esta limitación en la recolección de información tiene un impacto significativo en la escalabilidad del sistema a un entorno real pues se depende de la aprobación de Facebook para utilizar la información, además, desde la versión 2.4, se quitó la posibilidad de acceder a la información de inbox para aplicaciones que no sean de la marca Facebook. Aquellas aplicaciones aprobadas por Facebook con permisos para leer el inbox y que realicen consultas utilizando la versión 2.3 del API, solo podrán funcionar hasta mediados del 2017, después de esa fecha deberán migrar a una nueva versión.

# Capítulo 5

## Experimentación

En los capítulos 3 y 4 fue presentado el diseño e implementación de un sistema de recomendaciones basado en confianza y de un algoritmo de inferencia de confianza, sin embargo, para evaluar que lo planteado satisface las hipótesis que rigen esta investigación, se deben realizar pruebas técnicas en entornos offline y online.

En este capítulo se presentan dos pilotos de prueba realizados en el marco del proyecto doctoral a cargo del PhD(c) Francisco Martínez denominado SMARTA, un sistema diseñado para proporcionar recomendaciones grupales e individuales en un entorno de digital signage. En el primer piloto se presenta un sistema de recomendaciones para entornos de digital signage soportado en un esquema de cooperación Smart TV - Smartphone. En el segundo piloto se presenta un sistema de recomendaciones basado en confianza para el mismo tipo de entorno. El framework de evaluación utilizado es aquel propuesto en [131], donde se establecen un conjunto de métricas a evaluar y se diseñan experimentos para un entorno offline (utilizando datasets y modelando el comportamiento de los usuarios) y para un entorno online (con interacción directa de los usuarios).

### 5.1. Primer piloto: sistema de recomendaciones para entornos de publicidad ubicua

El objetivo de este primer piloto fue probar los algoritmos de recomendaciones y alimentar el dataset de publicidad. En la sección 3.4.1.2 se presentaron unos resultados de la evaluación del algoritmo de similitud donde la correlación de Spearman dio mejores resultados que la correlación de Pearson o la distancia euclidiana y por lo tanto fue la seleccionada para el sistema. Aunque Spearman tiene un costo computacional mayor a los otros dos algoritmos, debido a que primero ordena los ratings y luego les otorga un nuevo valor de acuerdo a su posición en la lista, Mahout dispone de un mecanismo de caché que evita recalcular valores de similitud previamente calculados.

En la figura 5.1 se presenta la arquitectura de este primer piloto, cabe resaltar que para el contexto de esta investigación es de interés el módulo de recomendaciones. En [132], [133] se encuentran los detalles de la implementación y los resultados generales del sistema.

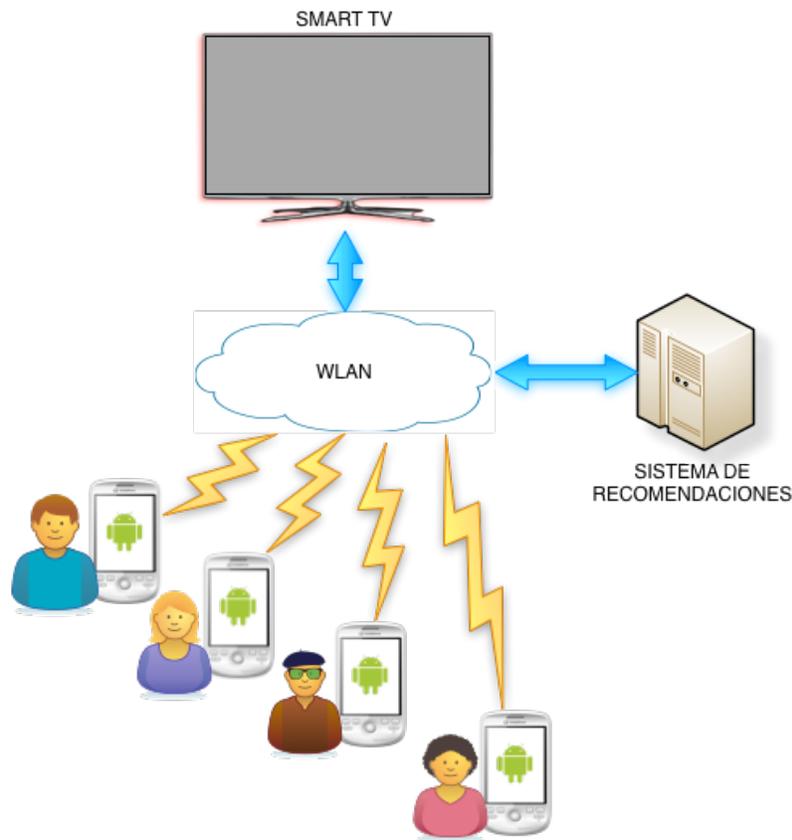


Figura 5.1: Arquitectura del primer piloto. Fuente [133]

Para este piloto, el sistema de recomendaciones es un filtro colaborativo clásico como el presentado en la sección 3.3.

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

Como se mencionó al inicio de este capítulo, el piloto se encuentra enmarcado en un proyecto denominado SMARTA, donde el TBRS es un componente fundamental pero no el único. SMARTA también está conformado por un middleware que soporta la interacción entre smartphones y un Smart TV, y por un módulo que permite generar recomendaciones grupales. La descripción detallada de estos componentes está por fuera del alcance de este trabajo, en [134] se encuentran más detalles al respecto.

A modo de resumen, las siguientes son las principales funcionalidades implementadas en el piloto:

1. Recomendación de anuncios para usuarios grupales que se encuentran observando la pantalla del televisor: el grupo está limitado a 4 usuarios.

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

2. Recomendación de anuncios para usuarios individuales: el sistema calcula una lista de anuncios recomendados para cada usuario que se encuentra en frente del televisor. Cada usuario desde su smartphone puede acceder a información de los anuncios y calificarlos.
3. Interacción básica entre un dispositivo móvil y el televisor: los usuarios pueden navegar entre los anuncios que se presentan en el televisor directamente desde su celular.
4. Ratings implícitos y explícitos: los usuarios pueden calificar los anuncios que se les presentan en el televisor o en el celular utilizando los botones de "me gusta" y "no me gusta". Además, gracias a las capacidades del middleware, es posible generar calificaciones implícitas de acuerdo a las acciones del usuario (ej. solicitar más información sobre un anuncio, ignorarlo, añadirlo a favoritos).
5. Publicar anuncios: los usuarios pueden publicar sus propios anuncios en el sistema utilizando la aplicación móvil.

La presente investigación se enfoca en la funcionalidad presentada en punto 2: la generación de recomendaciones para individuos utilizando información de confianza. Con respecto al propósito de la evaluación, el objetivo es validar las siguientes hipótesis: i) H1, es posible obtener un valor de confianza con suficiente grado de precisión a partir de la información de interacciones entre usuarios de Facebook; ii) H2, La inclusión de información de confianza en el algoritmo de filtrado colaborativo mejora el balance precisión/novedad durante el proceso de recomendación.

### 5.2.1. Tecnologías y herramientas

El resumen de las herramientas utilizadas para desarrollar el TBRS del piloto se encuentra en la tabla 5.1.

Herramienta	Descripción
Apache Mahout	Framework de recomendaciones que soporta el TBRS desarrollado. Más información en la sección 3.2.
Eclipse IDE for Java EE Developers	Entorno de desarrollo para el servidor de recomendaciones y la aplicación móvil
Android SDK	Conjunto de librerías que permiten el desarrollo de aplicaciones para el sistema operativo Android.
Graph API Explorer	Herramienta web de Facebook que permite realizar peticiones a Facebook sin implementar ninguna lógica adicional. Se utilizó para recolectar la información de interacciones de los participantes en el piloto. Más información en la sección 4.3.
Gephi	Herramienta de análisis de grafos utilizada para analizar la red de confianza de los participantes. Más información en la sección 4.4.3.
SPSS	Software estadístico utilizado para determinar la significancia de los resultados obtenidos

Tabla 5.1: Herramientas utilizadas en el desarrollo e implementación del segundo piloto. Fuente propia.

### 5.2.2. Arquitectura

En la figura 5.2 se presenta la arquitectura del segundo piloto, la cual es una evolución de la que se presentó en la sección 5.1. Sin embargo en este caso el esquema de conexión entre el Smart TV y los smartphones es distinto. Además se puede observar que el sistema de recomendaciones incluye el dominio de la confianza (ver sección 3.4).

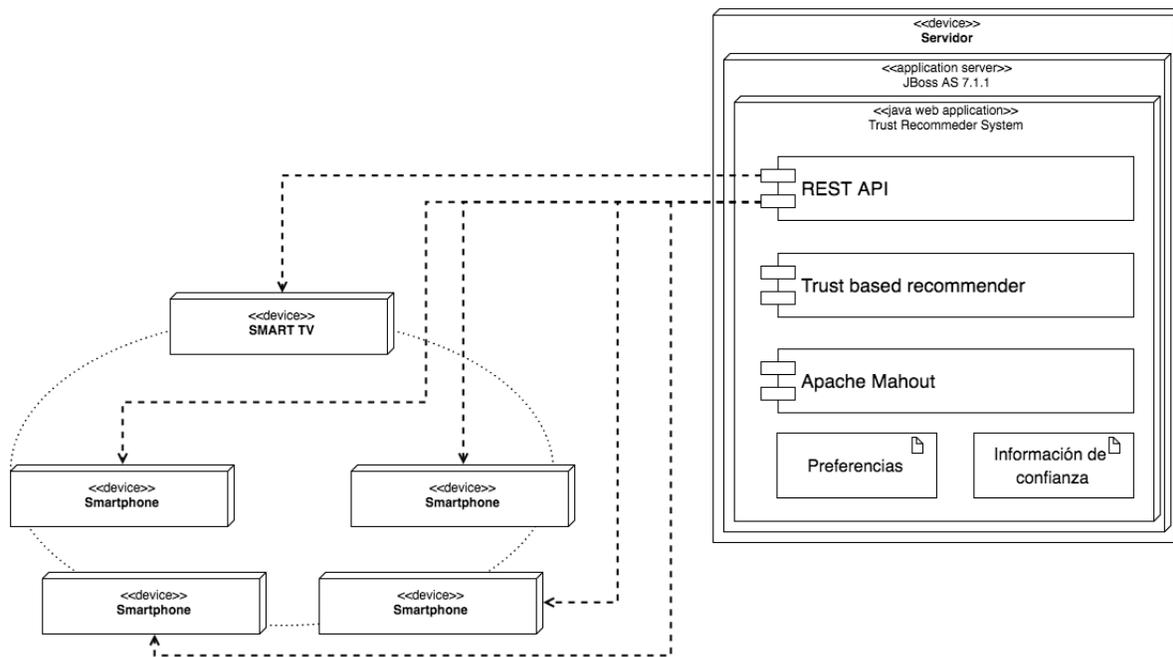


Figura 5.2: Arquitectura del segundo piloto. Fuente propia.

### 5.2.3. Pruebas offline

#### 5.2.3.1. Algoritmo de inferencia de confianza

Para la evaluación del algoritmo de confianza, se definió una serie de pruebas que permitieran determinar el conjunto idóneo de parámetros de interacciones en redes sociales (en este caso particular Facebook) para posteriormente evaluar el algoritmo en un experimento con voluntarios. El algoritmo se puso a prueba calculando el top 10 amigos de Facebook y posteriormente realizando una prueba de ground truth para verificar su precisión.

Inicialmente se diseñó una encuesta (ver figura 5.3) donde se preguntó qué parámetros representan mejor la confianza que tienen con otra persona en Facebook. Por simplicidad todos los parámetros relacionados con etiquetas se unieron en uno solo. La tabla 5.2 presenta los resultados obtenidos, en total participaron 57 personas para un total de 145 votos para las 5 variables.

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

Parámetro	Conteo	%	% Acumulado
Mensajes por Inbox	47	32 %	32 %
Comentarios	39	27 %	59 %
Publicaciones en el muro	31	21 %	81 %
Etiquetas	16	11 %	92 %
Likes	12	8 %	100 %
	145		

Tabla 5.2: Resultados de la encuesta de parámetros de confianza en Facebook. Fuente propia.

Confianza en Facebook

\*Obligatorio

¿De las siguientes opciones cuáles representan mejor la confianza que tiene con otra persona en Facebook? \*

- Likes
- Etiquetas
- Mensajes por inbox
- Posts en el muro
- Comentarios

Enviar

Con la tecnología de Google Forms

Google no creó ni aprobó este contenido.

[Denunciar abuso](#) - [Condiciones del servicio](#) - [Condiciones adicionales](#)

Figura 5.3: Encuesta de parámetros de confianza. Fuente propia.

Adicionalmente se definió un plan de pruebas que permitiera evaluar diferentes versiones del algoritmo de confianza con el objetivo de determinar cuál permitía tener mejores resultados tomando como base PCA.

**5.2.3.1.1. Pruebas iniciales** Las pruebas iniciales para el algoritmo de confianza se realizaron con los datos de una sola persona, el objetivo era determinar qué versiones del algoritmo mostraban mejores resultados. En total se realizaron 6 pruebas con diferentes variaciones del algoritmo de confianza, en la tabla 5.3 se ilustran las configuraciones para las distintas pruebas.

Prueba	Técnica	Variables a considerar
1	PCA	Mensajes, likes, comentarios y etiquetas
2	PCA	Mensajes, comentarios y etiquetas
3	PCA	Mensajes, etiquetas y publicaciones en el muro
4	PCA	Mensajes, comentarios y publicaciones en el muro
5	PCA	Mensajes, comentarios, publicaciones en el muro y etiquetas
6	PCA	Mensajes, comentarios, publicaciones en el muro y likes

Tabla 5.3: Plan de pruebas iniciales para el algoritmo de confianza. Fuente propia.

Para cada una de las pruebas de la tabla anterior se generó una lista ordenada de amigos de acuerdo a la confianza inferida por el algoritmo, posteriormente, se solicitó a un voluntario ordenar las listas de acuerdo a la confianza real con cada persona, finalmente, se compararon las dos listas para determinar la precisión. En las figuras 5.4, 5.5, 5.6, 5.7, 5.8, 5.9 se encuentran graficados el ranking generado por el algoritmo vs el ranking realizado por el usuario, entre más cercanas las curvas, más preciso el algoritmo. F1-F10 representan el top-10 de amigos del usuario.

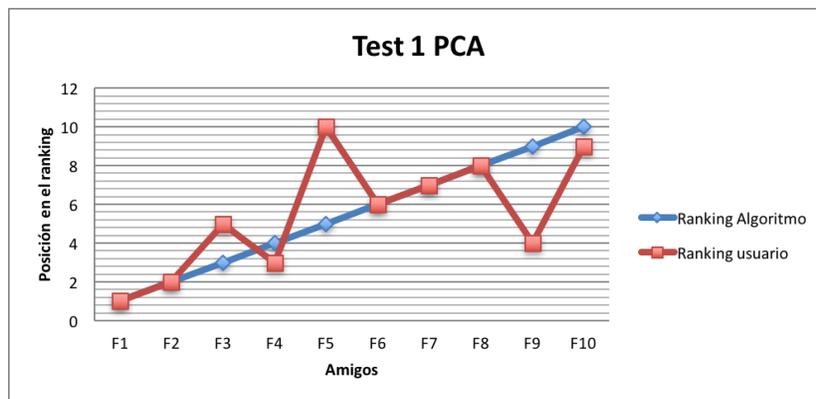


Figura 5.4: Prueba 1 PCA. Fuente propia.

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

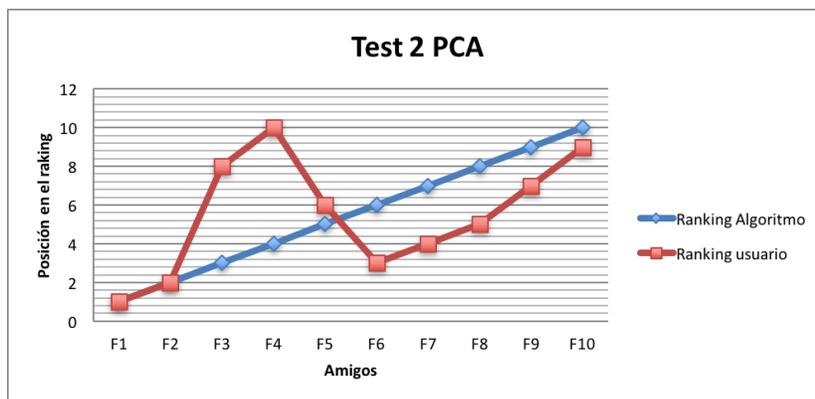
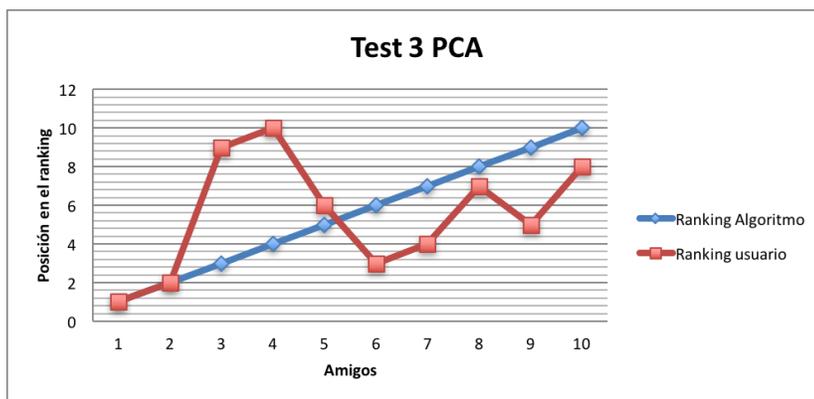


Figura 5.5: Prueba 2 PCA. Fuente propia.



Fuente propia.

Figura 5.6: Prueba 3 PCA. Fuente propia.

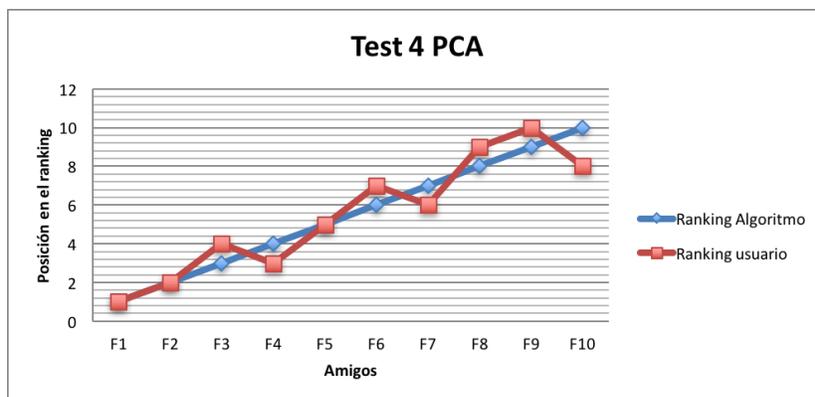


Figura 5.7: Prueba 4 PCA. Fuente propia.

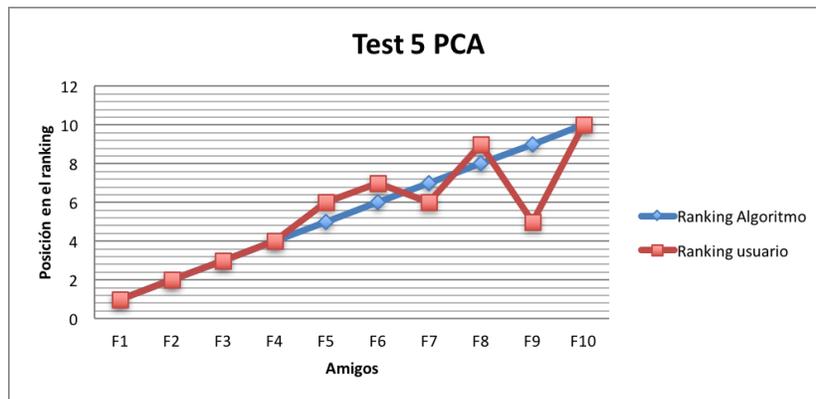


Figura 5.8: Prueba 5 PCA. Fuente propia.

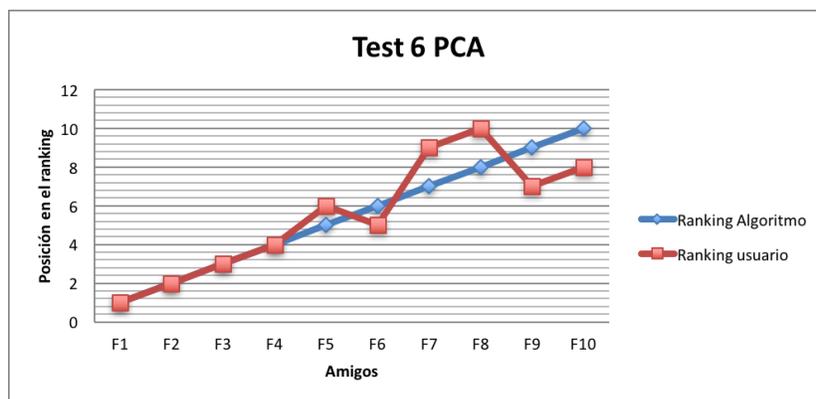


Figura 5.9: Prueba 6 PCA. Fuente propia.

Con base en los resultados de las pruebas anteriores, se pudo concluir lo siguiente:

- El parámetro más relevante para determinar la confianza entre dos personas en Facebook son los mensajes por Inbox
- Los likes no tienen un impacto real en la confianza
- Los mejores resultados para las pruebas iniciales se obtuvieron con las pruebas 4 (figura 5.7) y 5 (figura 5.8).

**5.2.3.1.2. Pruebas de ground truth** Una vez se realizaron las pruebas iniciales que permitieron filtrar los algoritmos con peores resultados, se procedió a diseñar y ejecutar pruebas de ground truth con un mayor volumen de usuarios,

Las pruebas de ground truth se dividen en 3 fases:

1. Recolección de datos de Facebook: cada uno de los participantes en el experimento extrae su información de Facebook (ver sección 4.3 y anexo A).

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

2. Análisis de la información recolectada: se analiza la información de los participantes con el objetivo de extraer su top de amigos. En total para cada participante se tienen 4 listas de amigos, una para cada variación del algoritmo.
3. Encuestas: se le pide a los usuarios que validen la información del algoritmo con el objetivo de determinar el más acertado.

Para generar la lista de amigos, se realiza una pequeña adaptación al algoritmo descrito en la sección 4.5.1 para que, en vez de generar el dataset de confianza, genere un top de amigos. En la figura 5.10 se presenta el proceso para generar el top-10 de amigos para un usuario.

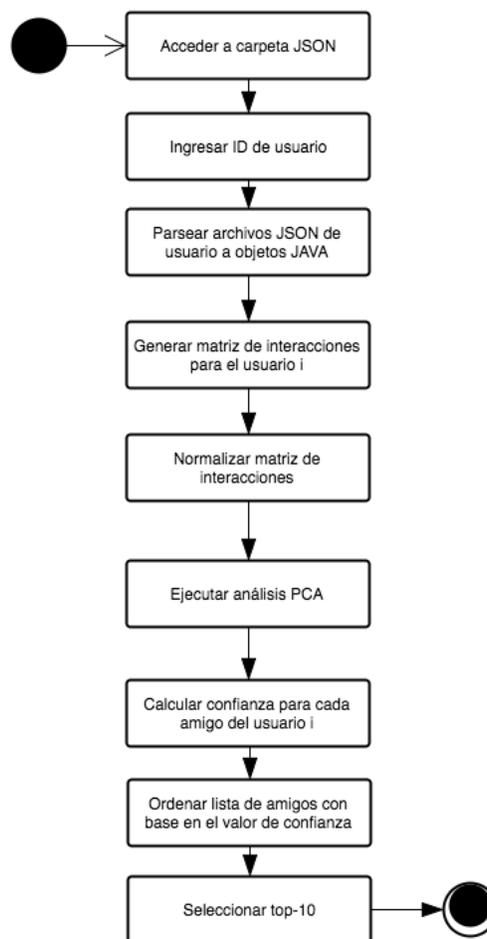


Figura 5.10: Flujo para generar el top-10 amigos de Facebook. Fuente propia.

**5.2.3.1.3. Resultados** Este proceso se aplicó a 5 voluntarios de la empresa Smartsoft Play quienes, una vez generada las listas, proporcionaron su propio ranking. En la tabla 5.4, se resumen los resultados obtenidos en la prueba de Ground Truth.

		Diferencia					Promedio diferencia	Porcentaje de acierto
		U1	U2	U3	U4	U5		
PCA	Test 1	26	14	16	28	8	18,4	63 %
	Test 2	20	20	16	24	8	17,6	65 %

Tabla 5.4: Resultados pruebas ground truth. Fuente propia.

La diferencia a la que hace referencia la tabla anterior, es resultado de la resta entre el ranking del algoritmo y el ranking del usuario, la ecuación 5.1 ilustra el proceso.

$$dif = \sum_{i=1}^{10} (|r_{ai} - r_{ui}|) \quad (5.1)$$

Donde  $r_{ai}$  representa el ranking otorgado por el algoritmo para el amigo  $i$ , y  $r_{ui}$  representa el ranking otorgado por el usuario para el amigo  $i$ . La ecuación 5.2, muestra la forma para calcular el porcentaje de acierto del algoritmo.

$$\% \text{ acierto} = 100 \% - \frac{\bar{dif}}{50} \quad (5.2)$$

Donde  $\bar{dif}$  representa el promedio de las diferencias de los usuarios para una versión del algoritmo y 50 representa la diferencia máxima, es decir, es el valor obtenido en caso de que los rankings sean totalmente opuestos.

Para determinar si las diferencias entre los algoritmos son significativas, se realizó un test t-student, para esto se plantean dos hipótesis.

- Ho: no hay diferencias significativas entre las variables
- H1: sí hay diferencias significativas entre las dos variables

Según el test realizado, el valor de significancia es mayor a 0.05, por lo tanto se acepta la hipótesis Ho. Los resultados del test se presentan en la tabla 5.5.

	Diferencias emparejadas			t	gl	sig. (p)
	Media	Desviación estándar	Media de error estándar			
PCA 1 & PCA 2	0,8000	4,604	2,059	0,389	4	0,717

Tabla 5.5: Test t-student sobre las pruebas de ground truth para dos versiones del algoritmo de inferencia de confianza. Fuente propia.

Con base en los resultados de las pruebas de ground truth, se pudo concluir lo siguiente:

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

- No existe diferencia significativa entre las dos versiones del algoritmo PCA.
- PCA presentó muy buenos resultados (porcentaje de acierto superior al 60 %) y, al ser una técnica estadística que se adapta a la información de cada usuario, es más fácil validar su universalidad.
- En la investigación realizada por Xiaoming Li [85], los resultados presentados muestran un porcentaje de acierto del algoritmo de 83 % utilizando PCA, pero considerando únicamente la información de un usuario y tomando en cuenta solamente el top-5 de amigos. En esas mismas condiciones, el algoritmo propuesto en este trabajo obtuvo un resultado máximo de acierto de 100 % para el test 1 con PCA y de 83 % para el test 2. Esto valida la hipótesis H1 planteada al inicio de la sección.

### 5.2.3.2. Sistema de recomendaciones

La segunda parte de las pruebas offline consistió en probar el sistema de recomendaciones con el objetivo de determinar el efecto de la confianza y ajustar el filtro colaborativo propuesto. Al igual que para la evaluación del algoritmo de confianza, se definió un conjunto de pruebas en las cuales se varían los valores  $\alpha$  y  $\beta$  de la ecuación 3.5. En la tabla 5.6 se ilustra el plan de pruebas para el sistema de recomendaciones.

Variación del algoritmo	
Test 1	$\alpha = 0, \beta = 1$
Test 2	$\alpha = 0.1, \beta = 0.9$
Test 3	$\alpha = 0.2, \beta = 0.8$
Test 4	$\alpha = 0.3, \beta = 0.7$
Test 5	$\alpha = 0.4, \beta = 0.6$
Test 6	$\alpha = 0.5, \beta = 0.5$
Test 7	$\alpha = 0.6, \beta = 0.4$
Test 8	$\alpha = 0.7, \beta = 0.3$
Test 9	$\alpha = 0.8, \beta = 0.2$
Test 10	$\alpha = 0.9, \beta = 0.1$
Test 11	$\alpha = 1, \beta = 0$

Tabla 5.6: Plan de pruebas para el sistema de recomendaciones. Este mismo plan se aplicó para las dos versiones del algoritmo descritas anteriormente. Fuente propia

**5.2.3.2.1. Métricas de novedad y precisión** El objetivo de estas pruebas fue determinar el efecto que tiene la confianza en un sistema de recomendaciones, específicamente en las métricas de novedad y precisión.

La precisión es una de las métricas más utilizadas para evaluar sistemas de recomendaciones y se define como la cantidad de ítems recomendados ( $N_r$ ) que son relevantes, está definida por la siguiente fórmula:

$$P = \frac{N_r}{N} \quad (5.3)$$

Donde N representa el conjunto total de ítems.

En cuanto a la segunda métrica de interés, [135] propone un framework para la definición de métricas de novedad y diversidad tomando en cuenta el ranking y la relevancia de los ítems. Estas propiedades son importantes para el contexto de la investigación puesto que en el dominio de la publicidad ubicua, el ranking y la relevancia toma en cuenta la forma como los usuarios interactúan con las recomendaciones (el top de ítems tiene más posibilidades de ser recomendado) y la subjetividad del usuario (qué tan relevante es el ítem para el usuario). De acuerdo con los autores del trabajo mencionado, la novedad de una pieza de información se refiere a "qué tan diferente es de lo que ha visto previamente un usuario" [135].

El framework de novedad propuesto por Vargas et al., se basa en tres relaciones fundamentales entre usuarios e ítems: i) descubrimiento, ej., un ítem es visto por un usuario; ii) selección, ej., un ítem es usado, escogido o consumido por un usuario ; y iii) relevancia, ej., el usuario disfruta o encuentra útil el ítem. En términos de distribución de probabilidad, estas relaciones se expresa de la siguiente forma:

$$p(choose) \sim p(seen)p(rel) \quad (5.4)$$

Dado un ranking de ítems R recomendado al usuario i, la novedad se puede expresar de la siguiente forma:

$$m(R | \Theta) = C \sum_{t \in R} p(choose | i, u, R)nov(i | \Theta) \quad (5.5)$$

Donde C es una constante de normalización y  $\Theta$  es una variable genérica de contexto para considerar diferentes perspectivas en la definición de novedad.  $p(choose | i, u, R)$  refleja el modelo de búsqueda asociado a la selección de ítems y  $nov(i | \Theta)$  es un modelo de novedad para ítems. Para el propósito de esta investigación, se utilizó un modelo de novedad de ítems basado en popularidad, donde valores altos de novedad corresponden a ítems con los que pocos usuarios han interactuado y valores bajos de novedad corresponden a los ítems más populares. De acuerdo a las características del dominio, este enfoque no solo tiene sentido para los usuarios sino también para los anunciantes: generalmente, los anunciantes van a querer promocionar nuevos productos que al principio serán poco conocidos en el sistema de recomendaciones, por lo tanto, una métrica de qué tan bien se comporta un sistema de recomendaciones en estos casos puede ser útil para ellos.

La métrica de novedad entonces queda definida por las siguientes ecuaciones:

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

$$nov(i | u) = C \sum_{i_k \in R} disc(k) p(rel | i_k, u) (1 - p(seen | i_k)) \quad (5.6)$$

En este caso, la métrica de novedad se puede entender como el número esperado de items relevantes recomendados desconocidos para el usuario. La ecuación incluye un componente de clasificación  $disc(k)$  que define una disminución logarítmica de acuerdo a la posición del ítem en la lista:

$$disc(k) = \frac{1}{\log_2(k + 1)} \quad (5.7)$$

También incluye un componente de relevancia  $p(rel | i_k, u)$  que puede ser modelado como un mapeo heurístico entre valores de ratings y la probabilidad de relevancia:

$$p(rel | i_k, u) \sim \frac{2^{g(u,i)} - 1}{2^{g_{max}}} \quad (5.8)$$

$$g(u, i) = \max(0, r(u, i) - \tau)$$

$g$  es una función de utilidad derivada de los ratings y  $\tau$  represente el valor de indiferencia de acuerdo con [136].

**5.2.3.2.2. Ejecución de las pruebas** Adicional a las variaciones establecidas en el plan de pruebas, se generó un dataset de confianza mediante código tomando como base la totalidad de usuarios del dataset de ratings, el cual se manipuló con el objetivo de determinar el efecto de la confianza en distintos escenarios. La tabla 5.7 muestra los 4 escenarios de prueba.

Prueba	Escenario
1	Todos los usuarios tienen relaciones de confianza con todos
2	Todos los usuarios tienen algunas relaciones de confianza
3	Todos los usuarios tienen relaciones de confianza con todos. La confianza es alta ( $> 0.75$ ) según la escala propuesta en [56].
4	Todos los usuarios tienen relaciones de confianza con todos. La confianza es baja ( $< 0.5$ ) según la escala propuesta en [56].

Tabla 5.7: Escenarios de prueba para el sistema de recomendaciones

**5.2.3.2.3. Resultados** En las figuras 5.11, 5.12, 5.13 y 5.14 se presentan los resultados de las pruebas de precisión. En cada gráfica el eje Y representa el valor de la precisión y el eje X las variaciones desde confianza 100 % ( $\alpha = 0, \beta = 1$ ) hasta similitud 100 % ( $\alpha = 1, \beta = 0$ ).

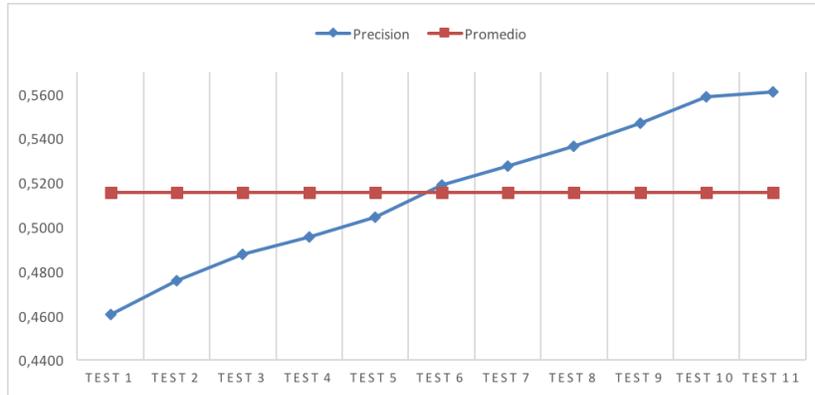


Figura 5.11: Test de precisión: Escenario 1. Fuente propia.

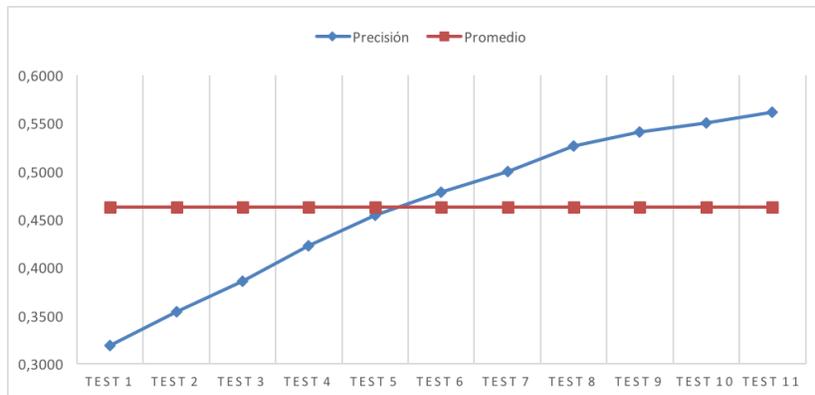


Figura 5.12: Test de precisión: Escenario 2. Fuente propia.

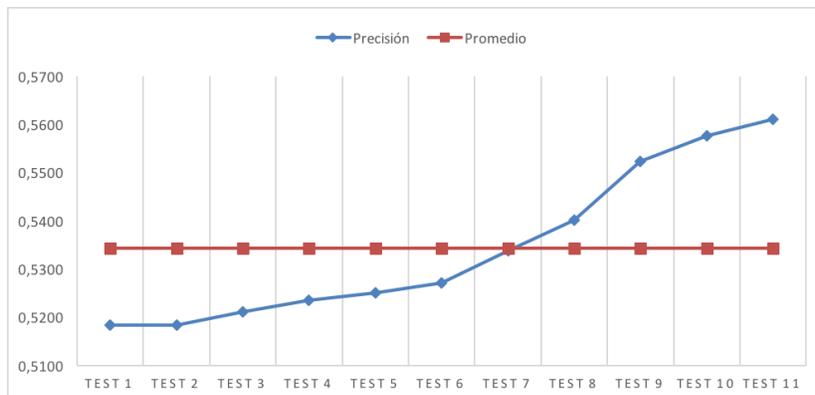


Figura 5.13: Test de precisión: Escenario 3. Fuente propia.

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

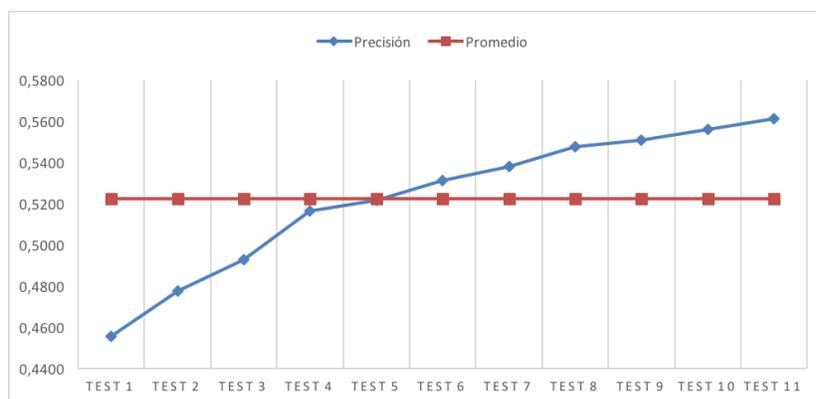


Figura 5.14: Test de precisión: Escenario 4. Fuente propia.

En base a los resultados de precisión se puede concluir lo siguiente:

- La confianza por sí misma no es suficiente para obtener valores adecuados de precisión, por lo tanto es necesario combinarla con la similitud. Esto corrobora resultados presentados en otras investigaciones.
- Los escenarios donde todos los usuarios tienen un valor de confianza con todos, presentan mejores resultados de precisión, siendo más notorio el caso donde existe confianza alta.
- En el escenario de prueba número 3 se muestra que la confianza por sí misma logra valores de precisión superiores al 51 %.

Como paso siguiente, se evaluó la novedad con el objetivo de probar la hipótesis H2 que indica que la inclusión de confianza mejora el balance precisión/novedad. La prueba consistió en tomar 50 usuarios del dataset de forma aleatoria y calcular la novedad para cada uno tanto para el sistema de recomendaciones tradicional ( $\alpha = 1, \beta = 0$ ), como para el TBRS propuesto ( $\alpha = 0.5, \beta = 0.5$ ). La tabla 5.8 muestra los resultados.

<b>Prueba</b>	<b>Media</b>	<b>Tamaño de la muestra</b>	<b>Desviación estándar</b>
Sistema de recomendaciones tradicional	0.18	50	0.058
TBRS	0.23	50	0.070

Tabla 5.8: Resultados de novedad incluyendo ranking y relevancia. Fuente propia

El valor de la media sugiere que el TBRS obtuvo mejores resultados desde una perspectiva de novedad, además al analizar los datos de novedad para cada usuario, se observa que el

TBRS presentó una mayor novedad en el 70 % de los casos. Para determinar si esta diferencia en la media es estadísticamente significativa, se realizó un t-test. Los resultados se resumen en la tabla 5.9.

	Diferencias emparejadas			t	gl	sig. (p)
	Media	Desviación estándar	Media de error estándar			
Novedad S100 & Novedad S50T50	-0,051	0,051	0,0072	-7,032	49	0,000

Tabla 5.9: Test t-student sobre los resultados de novedad. Fuente propia.

El  $p\text{-valor} = 0$ , indica que la diferencia entre las dos muestras es estadísticamente significativa, lo cual corrobora la hipótesis H2. Este hallazgo indica que la inclusión de confianza en el algoritmo de recomendación mejora el balance precisión/novedad. Sin embargo, es importante contrastar estos resultados con las percepciones de usuarios reales. En la sección 5.2.4 se describe este proceso.

## 5.2.4. Pruebas online

### 5.2.4.1. Marco experimental

Si bien en las pruebas offline se corroboró la hipótesis H2, donde se plantea que la inclusión de información de confianza en el algoritmo de filtrado colaborativo mejora el balance precisión/novedad durante el proceso de recomendación. Es necesario validar los resultados en un entorno real. A continuación se describe el marco experimental planteado para la evaluación del piloto.

#### 5.2.4.1.1. Objetivos

- Evaluar el impacto que tiene la confianza sobre la precisión
- Evaluar el impacto que tiene la confianza sobre la novedad

#### 5.2.4.1.2. Caracterización del dominio

- En el escenario real la tendencia marca la existencia de más usuarios que ítems.
- Tareas básicas en las cuales se enmarca el sistema de recomendaciones de SMARTA;
  - Encontrar algunos buenos ítems (contexto publicitario)
  - Maximizar la utilidad (el grado de novedad es muy importante en el contexto publicitario)

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

---

- Tipo de contenido: publicidad de productos y servicios.
- Granularidad de las preferencias de usuario: para el usuario que interactúa con el sistema, la escala será binaria ("me gusta" o "no me gusta"), sin embargo, las interacciones del usuario con el sistema se mapean a una escala de 1 a 5, que será tomada como rating para el sistema de recomendaciones.

**5.2.4.1.3. Dataset** En la sección 3.5 se describió el dataset de ratings utilizado en el sistema de recomendaciones y, en la sección 4.4.3, se describió el proceso para generar un dataset de confianza tomando en cuenta que uno de los requerimientos del piloto es tener un grupo cohesionado de usuarios donde la mayoría de personas tuvieran valores de confianza entre sí; para este caso particular los participantes fueron 20 voluntarios de la Fundación Innovagen.

**5.2.4.1.4. Configuración de la prueba** Inicialmente se dividieron los 20 voluntarios en 8 grupos de 3 personas (algunas personas forman parte de más de un grupo). Esta división se realizó analizando el grafo social de la Fundación Innovagen (ver sección 4.4.3, figura 4.1).

Cada grupo participó en dos sesiones; durante la primera sesión, el sistema de recomendaciones utilizó un filtrado colaborativo clásico ( $\alpha = 1$ ,  $\beta = 0$ ) mientras que en la segunda, se utilizó el TBRS propuesto ( $\alpha = 0,5$ ,  $\beta = 0,5$ ). En cada una de estas sesiones, los usuarios interactuaron con el sistema entre 5 y 10 minutos. Como parte de la interacción, se realizaron varias acciones como: navegar entre los anuncios (en el Smart TV y en el Smartphone), calificar anuncios (en el Smart TV y en el Smartphone), obtener información de un anuncio desplegado en el Smart TV para verlo en detalle en el Smartphone, añadir anuncio a la lista de favoritos y finalmente indicar si le gusta o no un anuncio. La tabla 5.10 muestra el rating asignado a cada una de estas interacciones .

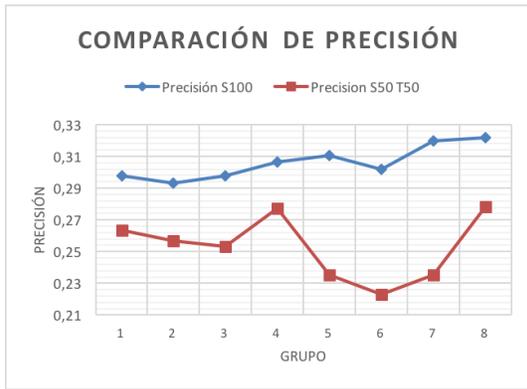
Interacción	Rating asignado
"Me gusta" explícito	5
Añadir a favoritos	4
Solicitar información del anuncio	3
Ignorar ítem en la pantalla	2
"No me gusta" explícito	1

Tabla 5.10: Representación de interacciones del usuario en una escala de 1 a 5

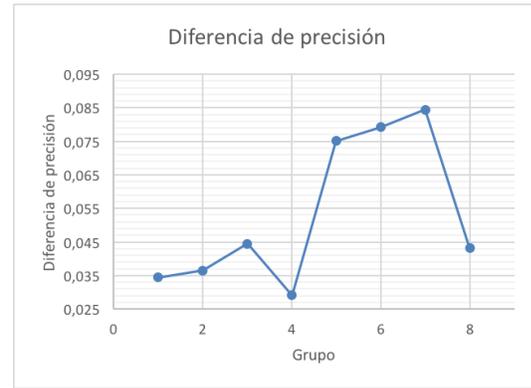
### 5.2.4.2. Resultados

**5.2.4.2.1. Influencia de la confianza sobre la precisión de las recomendaciones** La figura 5.15a muestra los resultados de precisión para el filtro colaborativo clásico y para el

TBRS, por su parte, la figura 5.15b muestra la diferencia de precisión entre ambos enfoques ( $P_{CF} - P_{TBRS}$ ), se puede observar que la gráfica siempre tiene valores positivos. Los resultados son coherentes con lo observado durante los experimentos offline.



(a) Precisión CF vs Precisión TBRS. Fuente propia.



(b) Diferencia de precisión  $P_{CF} - P_{TBRS}$ . Fuente propia.

Figura 5.15: Resultados de precisión para las pruebas online. Fuente propia

Para corroborar si la diferencia es significativa estadísticamente, se realiza un test t-student. La tabla 5.11 muestra que se obtiene un p-valor de 0, por lo tanto la diferencia es significativa. El siguiente paso fue determinar si, al igual que en las pruebas offline, con usuarios reales se obtenía una mejora en la novedad.

	Diferencias emparejadas			t	gl	sig. (p)
	Media	Desviación estándar	Media de error estándar			
Novedad S100 & Novedad S50T50	-0,051	0,051	0,0072	-7,032	49	0,000

Tabla 5.11: Test t-student sobre los resultados de precisión en las pruebas online

**5.2.4.2.2. Influencia de la confianza sobre la novedad de las recomendaciones** La tabla 5.12 muestra un análisis estadístico descriptivo del efecto que tiene sobre la novedad introducir confianza en el proceso de recomendación. La métrica de novedad toma en cuenta los componentes de ranking y relevancia y fue calculada para cada una de las sesiones.

## 5.2. Segundo piloto: sistema de recomendaciones basado en confianza

---

Sesión	Media	Desviación estándar	Error estándar
1. $\alpha = 1, \beta = 0$	0,6246	0,081	0,019
2. $\alpha = 0,5, \beta = 0,5$	0,707	0,079	0,018
		<b>p-valor</b>	0,011

Tabla 5.12: Análisis de impacto de la confianza sobre la novedad en las pruebas online

De acuerdo a estos resultados, la novedad se ve afectada de forma positiva cuando se incluye el componente de confianza en el algoritmo de recomendación, lo cual es coherente con los resultados observados en las pruebas offline y se comprueba la hipótesis H2. En el siguiente capítulo se resumen las conclusiones y aportes de esta investigación, además se presentan las limitaciones del trabajo y se señalan algunos puntos de partida para futuras investigaciones.



# Capítulo 6

## Conclusiones y trabajo futuro

### 6.1. Contribuciones del trabajo

#### 6.1.1. Implementación de referencia

En el capítulo 3 se detalla el proceso para incluir valores de confianza en un sistema de filtrado colaborativo, además se presenta una implementación de referencia para un sistema de recomendaciones basado en confianza construido sobre Mahout y la matemática asociada. Adicionalmente, en el capítulo 4 se definen los parámetros más relevantes para determinar la fortaleza de la relación entre dos personas y se presenta un algoritmo de inferencia de confianza que toma como base información de interacciones entre dos usuarios.

#### 6.1.2. Datasets

Esta investigación contribuyó a la creación de un dataset para el contexto de la publicidad ubicua y de un dataset de confianza basado en información de interacciones en Facebook.

#### 6.1.3. Prototipos funcionales

Se plantearon dos prototipos funcionales: en el primero se evaluaron distintas opciones de algoritmos de recomendación para un entorno de publicidad ubicua y se construyó un dataset pertinente para este dominio. En el segundo prototipo se implementó un TBRS que toma información de Facebook para calcular la confianza entre usuarios y posteriormente usar este valor en el proceso de recomendación.

#### 6.1.4. Publicaciones

Como resultado de esta investigación se tienen dos artículos publicados y uno actualmente aprobado para publicación. En el anexo C se encuentran más detalles.

## 6.2. Conclusiones

### 6.2.1. Conclusiones de la implementación del algoritmo de inferencia de confianza

- Si bien Facebook es una de las redes sociales más utilizadas y donde más datos se generan, las recientes preocupaciones con respecto a la privacidad de los usuarios ha limitado el acceso información relevante para el contexto de esta investigación. Inicialmente se planteó el desarrollo de una app móvil que, de forma automática, accediera a la información del usuario para posteriormente calcular la confianza con sus amigos y ofrecer las recomendaciones, sin embargo, debido a las limitaciones del API se optó por recolectar los datos de forma manual utilizando el Graph API Explorer. Estos datos fueron procesados y luego almacenados en un dataset de confianza que se utilizó posteriormente para generar las recomendaciones.
- La cantidad de interacciones entre dos personas permite determinar un valor de confianza entre ellas, no obstante, al considerar una red de personas y la confianza entre ellas, es necesario normalizar este valor de interacciones para evitar sesgos en el calculo de confianza para usuarios muy activos en la red.
- El parámetro más relevante para determinar la confianza entre dos personas en Facebook, son los mensajes intercambiados por Inbox.
- En el conjunto inicial de 7 parámetros de redes sociales se consideraron los likes, sin embargo, la experimentación probó que no tienen un impacto real en la confianza entre dos personas.
- PCA presentó muy buenos resultados (porcentaje de acierto superior al 60 %) y, al ser una técnica estadística que se adapta a la información de cada usuario, es más fácil validar su universalidad.
- En una investigación previa ([85]), los resultados presentados muestran un porcentaje de acierto del algoritmo de inferencia de confianza del 83 % utilizando PCA, pero considerando únicamente la información de un usuario y tomando en cuenta solamente el top-5 de amigos. En esas mismas condiciones, el algoritmo propuesto en este trabajo obtuvo un resultado máximo de acierto de 100 % para el test 1 con PCA y de 83 % para el test 2.

### 6.2.2. Conclusiones de la implementación del TBRS

- Si bien por las características propias del dataset (más usuarios que ítems), un enfoque item-item hubiese sido el más apropiado; para un TBRS las relaciones entre los usuarios son el nucleo del sistema, además, si se toma en cuenta el contexto de la publicidad ubicua, uno de los aspectos más importantes es la novedad. Por estas razones se escogió un enfoque user-user.

## 6.2. Conclusiones

---

- Los resultados de la comparación entre 3 diferentes enfoques para calcular similitud basada en usuarios mostraron a Spearman como la mejor opción. Esto se debe a que los datos no son lineales (Pearson funciona mejor cuando existe una relación lineal entre dos variables), sino que los datos se pueden representar como una función monótona.
- En la revisión de la literatura se encontraron diferentes alternativas para incluir confianza en un sistema de recomendaciones. Varios propusieron reemplazar el valor de la similitud por un valor de confianza calculado a partir del historial de ratings entre usuarios y la cantidad de buenas recomendaciones entre ellos, sin embargo, estos trabajos no utilizaban información de redes sociales sino que extendían el algoritmo tradicional de similitud. Trabajos más recientes optaron por incluir un valor de confianza calculado a partir de interacciones en redes sociales y combinar este valor con la similitud tradicional. Este último fue el enfoque que se utilizó en esta investigación.
- Existen datasets muy completos para los dominios de películas, música y opiniones. Sin embargo ninguno caracteriza de forma adecuada el dominio de la publicidad ubicua.
- La confianza por sí sola no es suficiente para obtener valores adecuados de precisión, por lo tanto es necesario combinarla con la similitud.
- Los escenarios donde el grafo de confianza tiene una densidad = 1 presentaron mejores resultados de precisión, siendo más notorio el caso donde existe confianza alta.
- En el escenario de prueba número 3 se muestra que la confianza por sí misma logra valores de precisión superiores al 51 %.
- Tanto en un entorno offline, como en pruebas con usuarios reales, la inclusión de confianza mejoró el balance precisión/novedad.
- El TBRS propuesto presentó mayor novedad que un filtro colaborativo clásico en más del 70 % de los casos.

### 6.2.3. Análisis final

En este trabajo se propuso un sistema de recomendaciones basado en confianza para un entorno de publicidad ubicua. Para el componente de confianza incluido en el proceso de recomendación se implementó un algoritmo de inferencia que toma como insumo la información de interacciones de Facebook. Inicialmente se plantearon 7 variables: mensajes intercambiados por inbox, likes, etiquetas de A a B, etiquetas de B a A, etiquetas en común, comentarios y publicaciones en el muro. Los resultados de la investigación demostraron que los mensajes intercambiados por inbox son el factor más relevante al momento de determinar la confianza entre dos personas, además, se demostró que con 4 variables (mensajes intercambiados por inbox, comentarios, etiquetas y publicaciones en el muro) se puede calcular con precisión la confianza entre dos personas.

En cuanto al sistema de recomendaciones, se propuso un enfoque que combina similitud y confianza de tal forma que se pudieran variar los pesos de cada uno y así determinar la mejor combinación del algoritmo. Se evaluó el sistema propuesto desde una perspectiva de precisión/novedad en entornos offline y online. La experimentación demostró que incluir el componente de confianza en la recomendación mejoró la novedad, sin embargo, el algoritmo donde solo se tomó en cuenta la confianza y no la similitud, afectó considerablemente la precisión. La combinación  $\alpha = 0.5$  y  $\beta = 0.5$ , donde  $\alpha$  representa el peso de la similitud y  $\beta$  el de la confianza, ofreció mejores resultados en cuanto al balance de precisión/novedad.

### 6.3. Trabajo futuro

- El dataset de anuncios generado es un buen punto de partida para futuras investigaciones en el área de la publicidad ubicua. Nuevos trabajos pueden enfocarse en ampliar este dataset para tener una base más robusta que permita probar sistemas de recomendaciones en el contexto de la publicidad ubicua.
- En la práctica, la operación del algoritmo propuesto tiene varias restricciones debido principalmente a la disponibilidad de información de interacciones. Para la experimentación, el enfoque utilizado en esta investigación fue obtener la información de forma manual utilizando el Graph API Explorer, sin embargo, las constantes actualizaciones de la plataforma de Facebook prohibirán el acceso a esta información a partir del 2017. Nuevos trabajos pueden enfocarse en inferir información de confianza en nuevas redes sociales, como Instagram por ejemplo, donde el acceso a la información no se encuentra tan restringido.
- El algoritmo de inferencia de confianza propuesto se probó en un grupo limitado de personas. Futuras investigaciones pueden realizar pruebas exhaustivas con grupos grandes de personas para validar la información aquí presentada
- Actualmente el TBRS propuesto considera información de confianza y similitud, entendiendo la confianza como una relación entre dos personas; sin embargo nuevos trabajos pueden evaluar el efecto incluir información de reputación en el proceso de recomendaciones, donde la reputación de A es un valor global dado por el promedio de confianza que tiene este usuario con todos los miembros de la red de confianza.

# Anexo A

## Anexo A. Guía de instalación de Mahout

Esta guía de instalación se trabajó en conjunto con el estudiante de Maestría en Telemática Víctor Garzón Marín.

### A.1. Preparación del entorno

Se recomienda cumplir con los siguientes requerimientos:

a Sistema operativo:

Preferiblemente entornos basados en Unix (Mac OS o Linux), para esta guía se utilizará Ubuntu 14.04.1. Por facilidad y para evitar tener que instalar un nuevo sistema operativo para completar la práctica, se instalará Ubuntu en una máquina virtual mediante VirtualBox.

**Nota:** si ya tiene instalado Ubuntu en su computador puede omitir este paso.

- I Para instalar VirtualBox se debe descargar desde la siguiente página <https://www.virtualbox.org/wiki/Downloads>.
- II Una vez instalado se debe crear la nueva máquina virtual con sistema operativo Ubuntu. Es recomendable crearla con los siguientes parámetros: 2GB de RAM y disco duro de 12GB.
- III Descargar Ubuntu de la siguiente página <http://www.ubuntu.com/download/desktop>.
- IV Iniciar la máquina virtual recién creada y cargar el ISO de Ubuntu cuando se lo solicite. Seguir los pasos presentados en pantalla hasta completar la instalación del nuevo sistema operativo. Al finalizar pedirá que reinicien la máquina virtual.
- V Para poder trabajar a pantalla completa, es necesario instalar los Guest Additions:

```
$ sudo apt-get install build-essential module-assistant
→ linux-headers-$(uname -r) dkms
$ sudo apt-get install virtualbox-guest-dkms
→ virtualbox-guest-utils virtualbox-guest-x11
```

```
$ sudo m-a prepare
$ sudo adduser YOUR_USERNAME vboxsf
```

### VI Reiniciar la máquina virtual.

#### b Instalar Java.

I Descargar Java en el siguiente enlace <http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html>. En el caso de esta guía se descarga el archivo `jdk-7u67-linux-i586.tar.gz` (antes de descargar se debe aceptar la licencia).

II Abrir un terminal y dirigirse a la carpeta donde se descargó Java.

```
$ cd /home/your_username/Downloads
```

III Copiar el paquete a `/usr/local/lib` y extraer su contenido.

```
$ sudo cp jdk-7u67-linux-i586.gz /usr/local/lib
$ sudo tar -xzvf jdk-7u67-linux-i586.gz
```

IV Añadir las variables de entorno de Java

```
$ cd /home/your_username
$ gedit .profile
```

Esto abre un archivo de texto donde se deben agregar las siguientes líneas al final:  
`export JAVA_HOME="/usr/local/lib/jdk1.7.0_67"`  
`PATH="$JAVA_HOME/bin:$PATH"`

V Reiniciar la máquina virtual

VI Verificar la correcta instalación

```
$ java -version
```

Se debería obtener la siguiente salida, en caso contrario se deben repetir los pasos de la instalación para verificar en dónde se encuentra el error.

```
java version "1.7.0_67"
Java(TM) SE Runtime Environment (build 1.7.0_67-b01)
Java HotSpot(TM) Cliente VM (build 24.65-b04, mixed
→ mode)
```

#### c IDE de desarrollo:

Para el desarrollo de esta práctica se utilizará Eclipse Indigo para Java EE el cual se puede descargar en el siguiente enlace: <http://www.eclipse.org/downloads/packages/eclipse-ide-java-ee-developers/indigosr2>. Se recomienda descomprimir eclipse en la ruta `/home/your_username/Mahout`.

#### d Maven 2.2:

Se puede instalar Maven directamente desde Aptitude

## A.2. Convertir el proyecto Maven a Eclipse

---

```
$ sudo apt-get install maven
```

Se puede verificar la correcta instalación con el siguiente comando:

```
$ mvn --version
```

Este último comando debería retornar la versión que acabamos de instalar de Maven, en caso contrario es necesario repetir los pasos de la instalación para verificar en dónde se encuentra el error.

e Código fuente:

El código fuente de Mahout se puede descargar desde el siguiente enlace. Actualmente se encuentra en la versión 0.9 <http://mirror.cogentco.com/pub/apache/mahout/0.9/>. Se recomienda descomprimir el código fuente en la ruta `/home/your_username/Mahout/workspace`

## A.2. Convertir el proyecto Maven a Eclipse

Al descargar el código fuentes, se obtendrá un paquete cuyo contenido es el proyecto Maven (para este caso supondremos que la ruta del directorio al descomprimir el paquete es `/home/your_username/Mahout/workspace/mahout-distribution-0.9`), para convertirlo a un proyecto que es posible usar en Eclipse, debe realizarse un proceso de transformación

```
$ cd /home/your_username/Mahout/workspace/mahout-distribution-0.9/  
$ mvn eclipse:eclipse
```

## A.3. Configurar variables del sistema

Es necesario fijar una variable al classpath.

a En un terminal escribir:

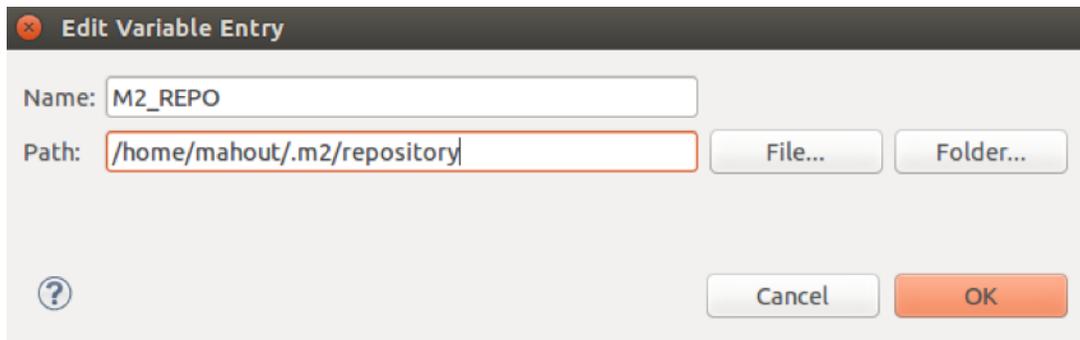
```
$ mvn -Declipse.workspace=eclipse:add-maven-repo
```

**Nota:** este paso puede tomar varios minutos en realizarse

b Desde eclipse en el menú

I Window > Preferences > Java > Build Path > Classpath Variables

II Seleccionamos New y agregamos la variable M2\_REPO con ruta `/home/your_username/.m2/repository`



## A.4. Importar el proyecto a Eclipse

Finalmente es necesario importar el proyecto a Eclipse desde donde ya se podrá empezar a usar mahout como motor de recomendaciones.

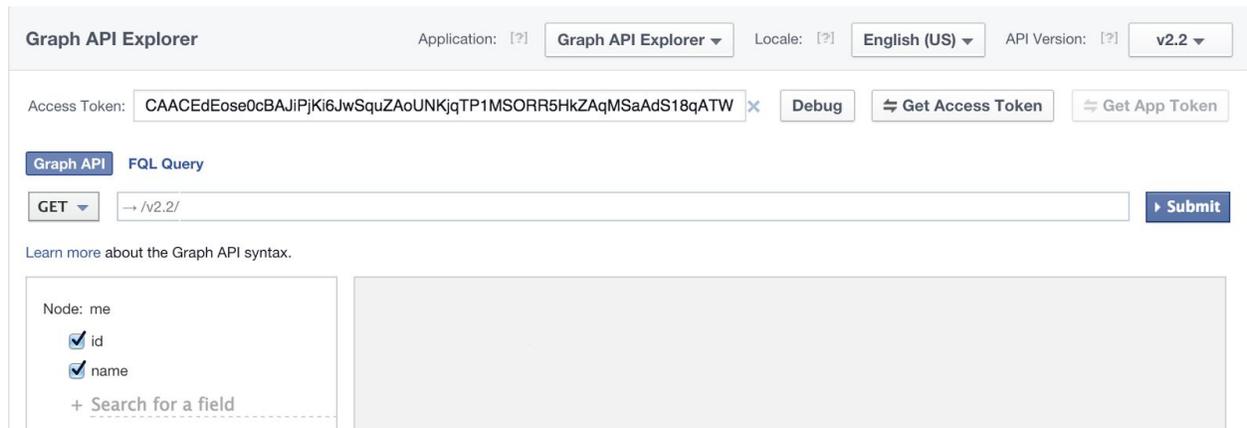
- a Desde el menú de Eclipse: File > Import > General Existing Projects into Workspace
- b Escoger la ruta del proyecto y finalizar

## **Anexo B**

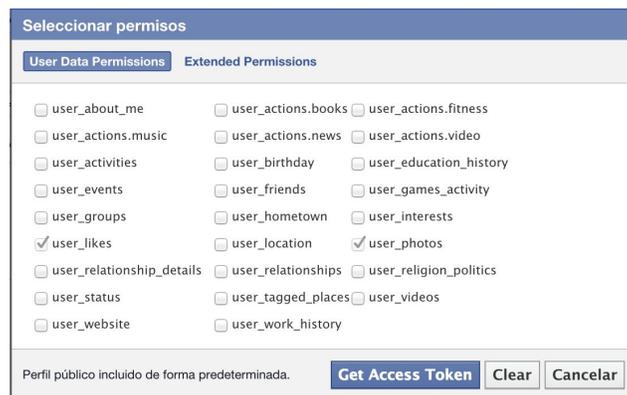
### **Anexo B. Recolección de Datos - Piloto Smarta**

## Recolección de Datos - Piloto Smarta

1. Entrar a <https://developers.facebook.com/tools/explorer>
2. Iniciar sesión con Facebook
3. Una vez iniciada la sesión, aparecerá la siguiente pantalla



4. Lo primero que se debe realizar es pedir un access token a Facebook para darle los permisos necesarios para acceder a los datos. Los permisos que se deben dar son user\_likes, user\_photos (como se ve en la imagen)



Luego se da click en la pestaña que dice Extended Permissions y se exigen read\_stream, read\_mailbox. Una vez seleccionados, seleccionar la opción Get Access Token.

Seleccionar permisos

User Data Permissions
Extended Permissions

ads\_management

ads\_read

email

manage\_notifications

manage\_pages

publish\_actions

read\_friendlists

read\_insights

read\_mailbox

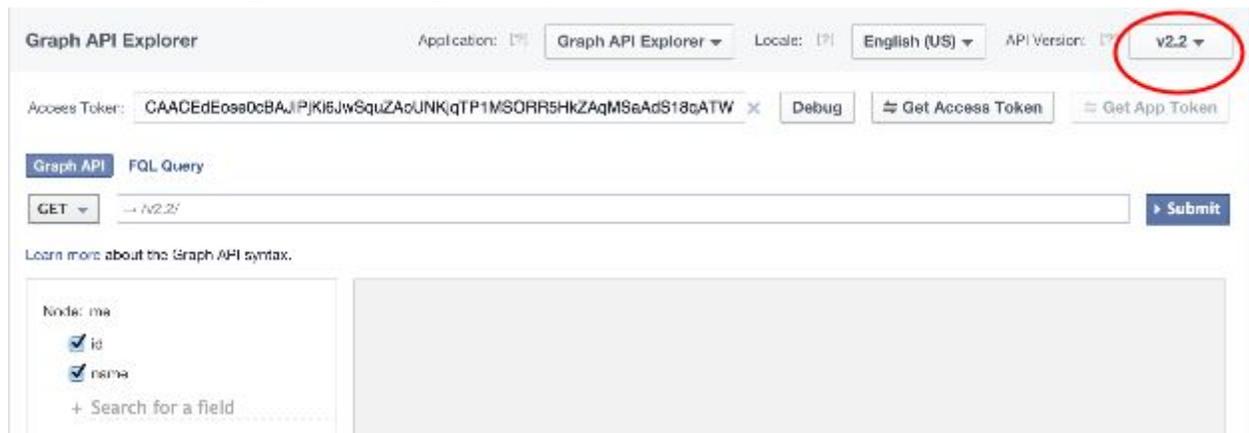
read\_page\_mailboxes

read\_stream

rsvp\_event

Perfil público incluido de forma predeterminada.
Get Access Token
Clear
Cancelar

5. Una vez se garantizan los permisos, se debe cambiar la versión del API a 2.0



6. En el campo de texto ubicado al lado del botón Submit, ingresar la siguiente sentencia y presionar submit

`me?fields=id,name`

Esto nos retornará una respuesta similar a esta:

```
{
  "id": "576381956",
  "name": "Juan Ospina"
}
```

7. Crear una carpeta con el número del id del usuario (en mi caso es 576381956)

8. Ingresar la siguiente sentencia y presionar submit:

`me/feed?fields=from,type,id,created_time,likes.limit(500){id,name},comments.limit(500){id,from}`

## Anexo B. Anexo B. Recolección de Datos - Piloto Smarta

Copiar el texto resultante en un archivo de texto y guardarlo como feed.json en la carpeta creada en el paso anterior

9. Ingresar la siguiente sentencia y presionar submit:

```
me/photos?fields=likes.limit(500){id,name},comments.limit(500){from},tags.limit(100){id,name},from
```

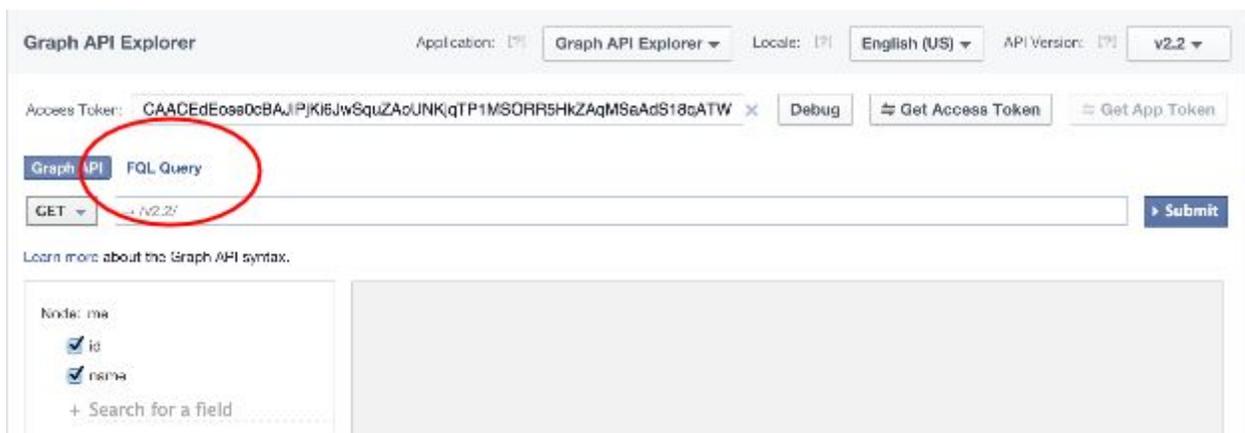
Copiar el texto resultante en un archivo de texto y guardarlo como taggedPhotos.json en la carpeta creada en el paso 7

10. Ingresar la siguiente sentencia y presionar submit:

```
me/photos/uploaded?fields=likes.limit(500){id,name},comments.limit(500){from},tags.limit(100){id,name}
```

Copiar el texto resultante en un archivo de texto y guardarlo como photos.json en la carpeta creada en el paso 7

11. Cambiar a la pestaña FQL



12. Ingresar la siguiente sentencia y presionar submit:

```
SELECT message_count, recipients FROM thread WHERE folder_id=0 LIMIT 100
```

Copiar el texto resultante en un archivo de texto y guardarlo como inbox.json en la carpeta creada en el paso 7

# Anexo C

## Publicaciones

<b>Ficha técnica de publicaciones</b>	
<b>Tipo de publicación</b>	Artículo de conferencia
<b>Título</b>	Enriching public displays ads recommendations using an individual - group cooperation model
<b>Autores</b>	Martínez Pabón, F. O., Ospina Quintero J. C., Garzón Marín V., Astaiza A. C., Muñoz Organero M., & Ramírez González G. A.
<b>Abstract</b>	<p>Ads recommendations delivery on digital signage environments must consider group and individual profiles. Although the ads recommendation precision is relevant and frequently explored, a good serendipity/precision balance must be a concern on Digital Signage context. This paper introduces a public display recommender system approach based on an individual-group cooperation model implemented throughout a Smart TV – Smartphone interaction scheme.</p>
<b>Nombre de la conferencia</b>	3rd Intenational Symposium on pervasive displays
<b>Año de publicación</b>	2014
<b>Editor</b>	ACM New York
<b>Sitio de la conferencia</b>	Copenhaguen, Denmark
<b>URL</b>	<a href="http://dl.acm.org/citation.cfm?id=2617196">http://dl.acm.org/citation.cfm?id=2617196</a>
<b>DOI</b>	10.1145/2611009.2617196

<b>Ficha técnica de publicaciones</b>	
<b>Tipo de publicación</b>	Artículo de revista
<b>Título</b>	Sistemas de recomendaciones para entornos de Digital Signage soportado en un esquema de cooperación Smart TV - Smartphone
<b>Autores</b>	Martínez Pabón, F. O., Garzón Marín V., Ibarra Samboní J. J., Ospina Quintero J. C., & Caicedo Guerrero J.
<b>Abstract</b>	La publicidad en entornos de Digital Signage demanda el enriquecimiento de la aproximación clásica de recomendación orientada a individuos, a través de la entrega de anuncios para un grupo de personas que observa una pantalla pública. Aunque la precisión que se pueda lograr en las recomendaciones es importante, lo es aún más el grado de novedad que puedan percibir los usuarios. Este balance puede obtenerse no solo a través de la mejora de los algoritmos de recomendación, sino también a través de un adecuado esquema de cooperación multi-pantalla. En este sentido, el presente artículo propone una aproximación para la construcción de un sistema de recomendaciones para entornos de Digital Signage soportado en un esquema de cooperación Smart TV-Smartphone.
<b>Nombre de la revista</b>	Revista Ingenierías Universidad De Medellín
<b>Año de publicación</b>	2014
<b>URL</b>	<a href="http://revistas.udem.edu.co/index.php/ingenierias/article/view/924/1046">http://revistas.udem.edu.co/index.php/ingenierias/article/view/924/1046</a>
<b>Clasificación de la revista</b>	A2

<b>Ficha técnica de publicaciones</b>	
<b>Tipo de publicación</b>	Artículo de revista
<b>Título</b>	Recommending ads from trustworthy relationships in pervasive environments
<b>Autores</b>	Francisco Martinez-Pabon, Juan Camilo Ospina-Quintero, Gustavo Ramirez-Gonzalez and Mario Munoz-Organero.
<b>Abstract</b>	<p>The use of pervasive computing technologies for advertising purposes is an interesting emergent field for large, medium and small companies. Although recommender systems have been a traditional solution to decrease users' cognitive effort to find good and personalized items, the classic collaborative filtering needs to include contextual information to be more effective. The inclusion of users' social context information in the recommendation algorithm, specifically trust in other users, may be a mechanism for obtaining ads' influence from other users in their closest social circle. There is no consensus about the variables to use during the trust inference process, and its integration into a classic collaborative filtering recommender system deserves a deeper research. On the other hand, the pervasive advertising domain demands a recommender system evaluation from a novelty/precision perspective. The improvement of the precision/novelty balance is a matter not only related to the recommendation algorithm itself, but also a better recommendations display strategy.</p> <p>In this article, we propose a novel approach for a collaborative filtering recommender system based on trust. We designed an algorithm to calculate a trust score using an implicit trust inference mechanism from Facebook interaction information; then, we included this score in the collaborative filtering recommendation algorithm. Additionally, we developed a digital signage prototype using a multi-screen scheme for recommendations delivery to evaluate our proposal using a novelty/precision approach.</p>
<b>Nombre de la revista</b>	Mobile Information Systems
<b>Año de publicación</b>	Aprobado, pendiente de publicación
<b>Clasificación de la revista</b>	A2



# Bibliografía

- [1] K. Partridge y B. Begole, «Activity-based advertising», en *Pervasive Advertising*, Springer London, 2011, págs. 83-101, ISBN: 9780857293510. DOI: 10.1007/978-0-85729-352-7.
- [2] J. Müller, F. Alt y D. Michelis, «Pervasive advertising», en *Pervasive Advertising*, Springer London, 2011, págs. 1-29, ISBN: 9780857293510. DOI: 10.1007/978-0-85729-352-7.
- [3] P. Santos, F. R. Ribeiro y J. Metrôlho, «Using pervasive computing technologies to deliver personal and public ads in public spaces», en *Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on*, IEEE, 2012, págs. 1-6.
- [4] J. Zhou, E. Gilman, M. Ylianttila y J. Riekki, «Pervasive service computing: visions and challenges», *Proceedings - 10th IEEE International Conference on Computer and Information Technology, CIT-2010, 7th IEEE International Conference on Embedded Software and Systems, ICES-2010, ScalCom-2010*, n.º Cit, págs. 1335-1339, 2010. DOI: 10.1109/CIT.2010.239.
- [5] A. Ranganathan y R. H. Campbell, «Advertising in a pervasive computing environment», en *Proceedings of the 2Nd International Workshop on Mobile Commerce*, ép. WMC '02, New York, NY, USA: ACM, 2002, págs. 10-14, ISBN: 1-58113-600-5. DOI: 10.1145/570705.570708.
- [6] B. Schilit y N. Adams, «Context-aware computing applications», en *Proceedings of Workshop on Mobile Computing Systems and Applications, 1994*, 1994, págs. 85-90. DOI: 10.1.1.29.5833.
- [7] «American lifestyles - us - april 2015», Mintel Group Ltd, inf. téc., 2015. dirección: <http://store.mintel.com/american-lifestyles-2015-the-connected-consumer-seeking-validation-from-the-online-collective-us-april-2015>.
- [8] P. Victor, C. Cornelis y M. DeCock, «Trust networks for recommender systems», *Trust Networks for Recommender Systems*, vol. 4, n.º June, págs. 1-202, 2011, ISSN: 1875-7650. DOI: 10.2991/978-94-91216-08-4.
- [9] E. D. Spiegler, C. Hildebrand y F. Michahelles, «Social networks in pervasive advertising and shopping», *Pervasive Advertising*, vol. 2010, n.º Compete 2010, págs. 207-225, 2011. DOI: 10.1007/978-0-85729-352-7.

- [10] B. De Carolis, «Pervasive advertising», en, J. Müller, F. Alt y D. Michelis, eds. London: Springer London, 2011, cap. Adapting N, págs. 227-246, ISBN: 978-0-85729-352-7. DOI: 10.1007/978-0-85729-352-7\_11.
- [11] G. Adomavicius y A. Tuzhilin, «Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions», *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, n.º 6, págs. 734-749, 2005.
- [12] N. Lathia, S. Hailes y L. Capra, «Trust-based collaborative filtering», en *Trust Management II*, Y. Karabulut, J. Mitchell, P. Herrmann y C. Jensen, eds., Springer US, 2008, págs. 119-134.
- [13] A. Guide, *A Guide To The Project Management Body Of Knowledge (PMBOK® GUIDE)*. Project Management Institute, 2004, ISBN: 9781933890517.
- [14] P. Kruchten, *The Rational Unified Process: An Introduction*, 3.ª ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2003, ISBN: 0321197704.
- [15] N. Tintarev y J. Masthoff, *Recommender Systems Handbook*. 2011, vol. 54, págs. 479-510, ISBN: 978-0-387-85819-7. DOI: 10.1007/978-0-387-85820-3.
- [16] W. B. Ashton y R. A. Klavans, *Keeping Abreast of Science and Technology: Technical Intelligence in Business*. Battelle Press, 1997, ISBN: 9781574770186.
- [17] G. Linden, B. Smith y J. York, «Amazon. com recommendations: item-to-item collaborative filtering», *Internet Computing, IEEE*, vol. 7, n.º 1, págs. 76-80, 2003.
- [18] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan y J. T. Riedl, «Movielens unplugged: experiences with a recommender system on four mobile devices», en *People and Computers XVII — Designing for Society*, Springer London, 2004, págs. 263-279, ISBN: 978-1-85233-766-7. DOI: 10.1007/978-1-4471-3754-2.
- [19] D. Billsus, C. A. Brunk, C. Evans, B. Gladish y M. Pazzani, «Adaptive interfaces for ubiquitous web access», *Commun. ACM*, vol. 45, n.º 5, págs. 34-38, mayo de 2002, ISSN: 0001-0782. DOI: 10.1145/506218.506240.
- [20] C. C. Peddy y D. Armentrout, *Building Solutions with Microsoft Commerce Server 2002*, 1.ª ed. Redmond, WA, USA: Microsoft Press, 2003, ISBN: 0735618542.
- [21] G. Dror, Y. Koren, Y. Maarek e I. Szpektor, «I want to answer; who has a question?: yahoo! answers recommender system», en *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ép. KDD '11, New York, NY, USA: ACM, 2011, págs. 1109-1117, ISBN: 978-1-4503-0813-7. DOI: 10.1145/2020408.2020582.
- [22] R. Burke, «Hybrid recommender systems: survey and experiments», *User modeling and user-adapted interaction*, vol. 12, n.º 4, págs. 331-370, 2002.
- [23] M. J. Pazzani, «A framework for collaborative, content-based and demographic filtering», *Artificial Intelligence Review*, vol. 13, n.º 5-6, págs. 393-408, 1999.

## BIBLIOGRAFÍA

---

- [24] R. E. Petty, D. T. Wegener y L. R. Fabrigar, «Attitudes and attitude change», *Annual review of psychology*, vol. 48, n.º 1, págs. 609-647, 1997.
- [25] R. H. Guttman, A. G. Moukas y P. Maes, «Agent-mediated electronic commerce: a survey», *The Knowledge Engineering Review*, vol. 13, n.º 02, págs. 147-159, 1998.
- [26] M. Balabanović e Y. Shoham, «Fab: content-based, collaborative recommendation», *Communications of the ACM*, vol. 40, n.º 3, págs. 66-72, 1997.
- [27] C. Cornelis, J. Lu, X. Guo y G. Zhang, «One-and-only item recommendation with fuzzy logic techniques», *Information Sciences*, vol. 177, n.º 22, págs. 4906-4921, 2007.
- [28] G. Lekakos y P. Caravelas, «A hybrid approach for movie recommendation», *Multimedia tools and applications*, vol. 36, n.º 1-2, págs. 55-70, 2008.
- [29] R. J. Mooney y L. Roy, «Content-based book recommending using learning for text categorization», en *Proceedings of the fifth ACM conference on Digital libraries*, ACM, 2000, págs. 195-204.
- [30] M. J. Pazzani y D. Billsus, «Content-based recommendation systems», en *The adaptive web*, Springer, 2007, págs. 325-341.
- [31] I. Guy, «Recommender systems handbook», en F. Ricci, L. Rokach y B. Shapira, eds. Boston, MA: Springer US, 2015, cap. Social Rec, págs. 511-543, ISBN: 978-1-4899-7637-6. DOI: 10.1007/978-1-4899-7637-6\_15.
- [32] F. Ricci, L. Rokach y B. Shapira, «Introduction to recommender systems handbook», en *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Saphira y P. B. Kantor, eds., Springer US, 2011, págs. 1-35.
- [33] P. van Waart, I. Mulder y C. Bont, «Meaningful advertising», en *Pervasive Advertising*, J. Müller, F. Alt y D. Michelis, eds., Springer London, 2011, págs. 57-81.
- [34] R. Shannon, M. Stabeler, A. Quigley y P. Nixon, «Profiling and targeting opportunities in pervasive advertising», en *1st Workshop on Pervasive Advertising@ Pervasive*, 2009.
- [35] M. Deutsch, *The Resolution of Conflict: Constructive and Destructive Processes*, ép. Carl Hovland memorial lectures. Yale University Press, 1973, ISBN: 9780300016833.
- [36] P. Shapiro, «The social control of impersonal trust», 1987.
- [37] M. Deutsch, «Cooperation and trust: some theoretical notes», en *Nebraska Symposium on Motivation*, Oxford, England: Univer. Nebraska Press, 1962, págs. 275-320.
- [38] D. H. McKnight y N. L. Chervany, «Trust and distrust definitions: one bite at a time», en *Proceedings of the Workshop on Deception, Fraud, and Trust in Agent Societies Held During the Autonomous Agents Conference: Trust in Cyber-societies, Integrating the Human and Artificial Perspectives*, London, UK, UK: Springer-Verlag, 2001, págs. 27-54, ISBN: 3-540-43069-5.

- [39] F. Fukuyama, *Trust: the social virtues and the creation of prosperity*, ép. Free press paperback. Free Press, 1995, ISBN: 9780684825250.
- [40] K. Cook, *Trust in society*. Russell Sage Foundation, 2001.
- [41] E. M. Uslaner, *The moral foundations of trust*. Cambridge University Press, 2002.
- [42] N. Luhmann, *Trust and Power: Two Works*, ép. UMI Books on Demand partes 1-2. Wiley, 1979, ISBN: 9780471997580.
- [43] L. Mui, M. Mohtashemi y A. Halberstadt, «A computational model of trust and reputation», en *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, IEEE, 2002, págs. 2431-2439.
- [44] T. Grandison y M. Sloman, «A survey of trust in internet applications», *Communications Surveys Tutorials, IEEE*, vol. 3, n.º 4, págs. 2-16, 2000, ISSN: 1553-877X. DOI: 10.1109/COMST.2000.5340804.
- [45] D. Olmedilla, O. F. Rana, B. Matthews y W. Nejdl, «Security and trust issues in semantic grids», *System*, págs. 1-11, 2006, ISSN: 1862-4405.
- [46] E. Amoroso, T. Nguyen, J. Weiss, J. Watson, P. Lapiska y T. Starr, «Toward an approach to measuring software trust», en *Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on*, mayo de 1991, págs. 198-218. DOI: 10.1109/RISP.1991.130788.
- [47] S. J. Knapskog, «A metric for trusted systems», en *Proceedings of the 21st National Security Conference*, 1998, págs. 16-29.
- [48] D. W. Manchala, «Trust metrics, models and protocols for electronic commerce transactions», en *Distributed Computing Systems, 1998. Proceedings. 18th International Conference on*, IEEE, 1998, págs. 312-321.
- [49] J. Golbeck, «Computing and applying trust in web-based social networks», *Annals of Physics*, vol. 54, pág. 199, 2005, ISSN: 00223956. DOI: 10.1016/j.jpsychemes.2008.02.003.
- [50] B. Bhargava, L. Lilien, A. Rosenthal, M. Winslett, M. Sloman, T. S. Dillon, E.-J. Chang, F. K. Hussain, W. Nejdl, D. Olmedilla y col., «The pudding of trust [intelligent systems]», *Intelligent Systems, IEEE*, vol. 19, n.º 5, págs. 74-88, 2004.
- [51] S. P. Marsh, «Formalising trust as a computational concept», *Computing*, vol. Doctor of, n.º April, pág. 184, 1994. DOI: 10.2165/00128413-199409230-00010.
- [52] D. Gambetta, «Can we trust trust?», en *Trust: Making and Breaking Cooperative Relations*, D. Gambetta, ed., Blackwell, 1988, págs. 213-237.
- [53] S. Marsh y F. La, «Trust and reliance in multi-agent systems: a preliminary report», en *In Proceedings of the 4th European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, 1992.

## BIBLIOGRAFÍA

---

- [54] M. Richardson, R. Agrawal y P. Domingos, «The semantic web - iswc 2003: second international semantic web conference, sanibel island, fl, usa, october 20-23, 2003. proceedings», en, D. Fensel, K. Sycara y J. Mylopoulos, eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, cap. Trust Mana, págs. 351-368, ISBN: 978-3-540-39718-2. DOI: 10.1007/978-3-540-39718-2\_23.
- [55] D. Artz e Y. Gil, «A survey of trust in computer science and the semantic web», 2007.
- [56] B. Kosko, *Fuzzy Thinking: The New Science of Fuzzy Logic*. Hyperion, 1993, ISBN: 9781562828394.
- [57] J. Sabater y C. Sierra, «Regret: reputation in gregarious societies», en *Proceedings of the Fifth International Conference on Autonomous Agents*, ép. AGENTS '01, New York, NY, USA: ACM, 2001, págs. 194-195, ISBN: 1-58113-326-X. DOI: 10.1145/375735.376110.
- [58] ———, «Review on computational trust and reputation models», *Artif. Intell. Rev.*, vol. 24, n.º 1, págs. 33-60, 2005, ISSN: 0269-2821. DOI: 10.1007/s10462-004-0041-5.
- [59] T. Beth, M. Borchering y B. Klein, «Valuation of trust in open networks», en *Computer Security — ESORICS 94*, D. Gollmann, ed., Springer Berlin Heidelberg, 1994, págs. 1-18.
- [60] S. Xiao e I. Benbasat, «The formation of trust and distrust in recommendation agents in repeated interactions: a process-tracing analysis», en *Proceedings of the 5th International Conference on Electronic Commerce*, ép. ICEC '03, New York, NY, USA: ACM, 2003, págs. 287-293, ISBN: 1-58113-788-5. DOI: 10.1145/948005.948043.
- [61] J. O'Donovan y B. Smyth, «Trust in recommender systems», en *Proceedings of the 10th international conference on Intelligent user interfaces*, ACM, 2005, págs. 167-174.
- [62] J. Golbeck y J. Hendler, «Inferring reputation on the semantic web», en *Proceedings of the 13th International World Wide Web Conference*, vol. 316, 2004.
- [63] J. Golbeck y J. a. Hendler, «Accuracy of metrics for inferring trust and reputation in semantic web-based social networks», *Engineering Knowledge in the Age of the SemanticWeb*, vol. 3257, págs. 116-131, 2004, ISSN: 16113349. DOI: 10.1007/978-3-540-30202-5\_8.
- [64] S. D. Ramchurn, N. R. Jennings, C. Sierra y L. Godo, «Devising a trust model for multi-agent interactions using confidence and reputation», *Applied Artificial Intelligence*, vol. 18, n.º 9-10, págs. 833-852, 2004.
- [65] J. S. Coleman, «Social capital in the creation of human capital», *Frontier Issues in Economic Thought*, vol. 3, págs. 213-215, 1997.
- [66] J. S. Coleman y J. S. Coleman, *Foundations of social theory*. Harvard university press, 1994.

- [67] M. Granovetter, «Economic action and social structure: the problem of embeddedness», *American journal of sociology*, págs. 481-510, 1985.
- [68] N. Nohria y R. G. Eccles, «Networks and organizations: structure, form, and action», 1992.
- [69] B. Uzzi, «The sources and consequences of embeddedness for the economic performance of organizations: the network effect», *American sociological review*, págs. 674-698, 1996.
- [70] J. S. Olson y G. M. Olson, «I2i trust in e-commerce», *Communications of the ACM*, vol. 43, n.º 12, págs. 41-44, 2000.
- [71] D. L. Paul y R. R. McDaniel Jr, «A field study of the effect of interpersonal trust on virtual collaborative relationship performance», *MIS quarterly*, págs. 183-227, 2004.
- [72] P. Massa y P. Avesani, «Controversial users demand local trust metrics: an experimental study on epinions.com community», en *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 1*, ép. AAAI'05, AAAI Press, 2005, págs. 121-126, ISBN: 1-57735-236-x.
- [73] G. Lausen y C.-N. Ziegler, «Propagation models for trust and distrust in social networks», *Information Systems Frontiers*, vol. 7, págs. 337-358, 2005, ISSN: 13873326. DOI: 10.1007/s10796-005-4807-3.
- [74] S. D. Kamvar, M. T. Schlosser y H. Garcia-Molina, «The eigentrust algorithm for reputation management in p2p networks», en *Proceedings of the 12th International Conference on World Wide Web*, ép. WWW '03, New York, NY, USA: ACM, 2003, págs. 640-651, ISBN: 1-58113-680-3. DOI: 10.1145/775152.775242.
- [75] R. Guha, «Open rating systems», Stanford Knowledge Systems Laboratory, inf. téc., 2003.
- [76] J. Berger, «Interpersonal online trust in new online social networks», n.º November, 2011.
- [77] M. S. Granovetter, «The strength of weak ties», *American journal of sociology*, págs. 1360-1380, 1973.
- [78] K. Panovich, R. Miller y D. Karger, «Tie strength in question {&} answer on social network sites», *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, págs. 1057-1066, 2012. DOI: 10.1145/2145204.2145361.
- [79] E. Gilbert y K. Karahalios, «Predicting tie strength with social media», *ACM Conference on Human Factors in Computing Systems*, págs. 211-220, 2009, ISSN: 01672738. DOI: 10.1145/1518701.1518736.
- [80] R. Xiang, J. Neville y M. Rogati, «Modeling relationship strength in online social networks», *Proceedings of the 19th International Conference on World Wide Web*, págs. 981-990, 2010, ISSN: 15564029. DOI: 10.1145/1772690.1772790.

## BIBLIOGRAFÍA

---

- [81] V. Arnaboldi, A. Guazzini y A. Passarella, «Egocentric online social networks: analysis of key features and prediction of tie strength in facebook», *Computer Communications*, vol. 36, n.º 10-11, págs. 1130-1144, 2013, ISSN: 01403664. DOI: 10.1016/j.comcom.2013.03.003.
- [82] S. Grabner-Kräuter y S. Bitter, «Trust in online social networks: a multifaceted perspective», en *Forum for social economics*, Taylor & Francis, vol. 44, 2015, págs. 48-68.
- [83] I. Kahanda y J. Neville, «Using transactional information to predict link strength in online social networks», *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, págs. 74-81, 2009.
- [84] J. J. Jones, J. E. Settle, R. M. Bond, C. J. Fariss, C. Marlow y J. H. Fowler, «Inferring tie strength from online directed behavior», *PLoS ONE*, vol. 8, págs. 2-7, 2013, ISSN: 19326203. DOI: 10.1371/journal.pone.0052168.
- [85] X. Li, «Interpersonal trust measurements from social interactions in facebook», Tesis doct., Montana State University, 2014.
- [86] J. Golbeck, «Trust management: 4th international conference, itrust 2006, pisa, italy, may 16-19, 2006. proceedings», en K. Stølen, W. H. Winsborough, F. Martinelli y F. Massacci, eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, cap. Generating Predictive Movie Recommendations from Trust in Social Networks, págs. 93-104, ISBN: 978-3-540-34297-7. DOI: 10.1007/11755593\_8.
- [87] C.-N. Ziegler y G. Lausen, «Analyzing correlation between trust and user similarity in online communities», *Trust Management*, págs. 251-265, 2004.
- [88] U. Shardanand y P. Maes, «Social information filtering: algorithms for automating 'word of mouth'», *Proceedings of the ACM Conference on Human Factors in Computing Systems*, vol. 1, págs. 210-217, 1995. DOI: 10.1145/223904.223931.
- [89] C. Lueg, «Social filtering and social reality», en *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, 1997.
- [90] P. Bonhard y P. Bonhard, «Improving recommender systems with social networking», *In Vivo*, págs. 9-10, 2004.
- [91] J. C. Wang y C. C. Chiu, «Recommending trusted online auction sellers using social network analysis», *Expert Systems with Applications*, vol. 34, págs. 1666-1679, 2008, ISSN: 09574174. DOI: 10.1016/j.eswa.2007.01.045.
- [92] R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni y M. Tennenholtz, «Trust-based recommendation systems: an axiomatic approach», en *Proceedings of the 17th International Conference on World Wide Web*, ép. WWW '08, New York, NY, USA: ACM, 2008, págs. 199-208, ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367525.

- [93] M. I. Martín-Vicente, A. Gil-Solla, M. Ramos-Cabrer, Y. Blanco-Fernández y M. López-Nores, «Semantic inference of user's reputation and expertise to improve collaborative recommendations», *Expert Systems with Applications*, vol. 39, págs. 8248-8258, 2012, ISSN: 09574174. DOI: 10.1016/j.eswa.2012.01.162.
- [94] Y. Wang, G. Yin, Z. Cai, Y. Dong y H. Dong, «A trust-based probabilistic recommendation model for social networks», *Journal of Network and Computer Applications*, vol. 55, págs. 59-67, 2015.
- [95] X. Shen, H. Long y C. Ma, «Incorporating trust relationships in collaborative filtering recommender system», en *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015 16th IEEE/ACIS International Conference on*, IEEE, 2015, págs. 1-8.
- [96] J. Wu, L. Chen, Q. Yu, P. Han y Z. Wu, «Trust-aware media recommendation in heterogeneous social networks», *World Wide Web*, vol. 18, n.º 1, págs. 139-157, 2015.
- [97] L. Quijano-Sanchez, J. A. Recio-Garcia, B. Diaz-Agudo y G. Jimenez-Diaz, «Social factors in group recommender systems», *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, n.º 1, pág. 8, 2013.
- [98] P. Kazienko y K. Musial, «Recommendation framework for online social networks», págs. 1-10, 2006.
- [99] R. Burke, M. P. O'Mahony y N. J. Hurley, «Robust collaborative recommendation», en *Recommender Systems Handbook*, Springer, 2015, págs. 961-995.
- [100] C. N. Ziegler y G. Lausen, «Spreading activation models for trust propagation», en *E-Technology, e-Commerce and e-Service, 2004. EEE '04. 2004 IEEE International Conference on*, 2004, págs. 83-97. DOI: 10.1109/EEE.2004.1287293.
- [101] K. Andersen y T. Clevenger Jr, «A summary of experimental research in ethos», *Communications Monographs*, vol. 30, n.º 2, págs. 59-78, 1963.
- [102] M. C. Gilly, J. L. Graham, M. F. Wolfenbarger y L. J. Yale, «A dyadic study of interpersonal information search», *Journal of the Academy of Marketing Science*, vol. 26, n.º 2, págs. 83-100, 1998.
- [103] R. R. Harmon y K. A. Coney, «The persuasive effects of source credibility in buy and lease situations», *Journal of Marketing Research*, págs. 255-260, 1982.
- [104] D.-N. Lascu, W. O. Bearden y R. L. Rose, «Norm extremity and interpersonal influences on consumer conformity», *Journal of Business Research*, vol. 32, n.º 3, págs. 201-212, 1995.
- [105] W. J. McGuire, «The nature of attitudes and attitude change», *The handbook of social psychology*, vol. 3, n.º 2, págs. 136-314, 1969.
- [106] S. Senecal y J. Nantel, «Online influence of relevant others: a framework», *Quarterly Journal of Electronic Commerce*, 2002.

## BIBLIOGRAFÍA

---

- [107] S. Senecal y J. Nantel, «The influence of online product recommendations on consumers' online choices», *Journal of retailing*, vol. 80, n.º 2, págs. 159-169, 2004.
- [108] D. J. O'keefe, *Persuasion: Theory and research*. Sage Publications, 2015.
- [109] J. P. Dillard y M. Pfau, *The persuasion handbook: Developments in theory and practice*. Sage Publications, 2002.
- [110] J. Noel, S. Sanner, K.-N. Tran, P. Christen, L. Xie, E. V. Bonilla, E. Abbasnejad y N. Della Penna, «New objective functions for social collaborative filtering», en *Proceedings of the 21st international conference on World Wide Web*, ACM, 2012, págs. 859-868.
- [111] X. Yang, H. Steck, Y. Guo e Y. Liu, «On top-k recommendation using social networks», en *Proceedings of the sixth ACM conference on Recommender systems*, ACM, 2012, págs. 67-74.
- [112] R. Sinha y K. Swearingen, «Comparing recommendations made by online systems and friends», *DELOS Workshop on Personalisation and Recommender Systems in Digital Libraries*, 2001.
- [113] P. Victor, M. De Cock y C. Cornelis, «Trust and recommendations», en *Recommender systems handbook*, Springer, 2011, págs. 645-675.
- [114] X. Zhou, Y. Xu, Y. Li, A. Josang y C. Cox, «The state-of-the-art in personalized recommender systems for social networking», *Artificial Intelligence Review*, vol. 37, n.º 2, págs. 119-132, 2012.
- [115] C. C. Aggarwal, *Recommender Systems: The Textbook*. Springer, 2016, págs. 345-384.
- [116] P. Massa y P. Avesani, «Trust-aware recommender systems», en *Proceedings of the 2007 ACM conference on Recommender systems*, ACM, 2007, págs. 17-24.
- [117] W. C. W. Chen y S. Fong, «Social network collaborative filtering framework and online trust factors: a case study on facebook», *Digital Information Management (ICDIM), 2010 Fifth International Conference on*, págs. 266-273, 2010. DOI: 10.1109/ICDIM.2010.5664676.
- [118] R. Anil, S. Owen, T. Dunning y E. Friedman, *Mahout in Action*. Manning Publications Co. Sound View Ct. 3B Greenwich, CT 06830, 2010, pág. 350.
- [119] J. B. Schafer, J. A. Konstan y J. Riedl, «E-commerce recommendation applications», en *Applications of Data Mining to Electronic Commerce*, Springer, 2001, págs. 115-153.
- [120] P. Clements, D. Garlan, L. Bass, J. Stafford, R. Nord, J. Ivers y R. Little, «Documenting software architectures: Views and beyond», 2002.
- [121] M. D. Ekstrand, «Collaborative filtering recommender systems», *Foundations and Trends{®} in Human-Computer Interaction*, vol. 4, n.º 2, págs. 81-173, 2010, ISSN: 1551-3955. DOI: 10.1561/11000000009.

- [122] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom y J. Riedl, «Grouplens: an open architecture for collaborative filtering of netnews», en *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, ACM, 1994, págs. 175-186.
- [123] C. Than y S. Han, «Improving recommender systems by incorporating similarity, trust and reputation», *Journal of Internet Services and Information Security (JISIS)*, vol. 4, n.º 1, págs. 64-76, 2014.
- [124] K. V. L. Rhoads y R. B. Cialdini, «The business of influence: principles that lead to success in commercial settings», *The persuasion handbook*, págs. 513-542, 2002.
- [125] J. Feng, J. Lazar y J. Preece, «Empathy and online interpersonal trust: a fragile relationship», *Behaviour & Information Technology*, vol. 23, n.º 2, págs. 97-106, 2004.
- [126] R. J. Trudeau, *Introduction to graph theory*. Courier Corporation, 2013.
- [127] P. Fletcher, H. B. Hoyle y C. W. Patty, *Foundations of discrete mathematics*. Brooks/Cole Publishing Co., 1990.
- [128] F. Martínez Pabón, «Smarta: sistema de recomendaciones soportado en un esquema de cooperación smart tv - smartphone para entornos de publicidad ubicua», Tesis doct., Universidad del Cauca, 2016.
- [129] D. Easley y J. Kleinberg, «Networks , crowds , and markets : reasoning about a highly connected world», *Science*, vol. 81, pág. 744, 2010, ISSN: 19467567. DOI: 10.1017/CBO9780511761942.
- [130] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.
- [131] G. Shani y A. Gunawardana, «Evaluating recommendation systems», *Recommender systems handbook*, págs. 257-298, 2011, ISSN: 08909369. DOI: 10.1007/978-0-387-85820-3\_8.
- [132] F. Martínez-Pabón, J. C. Ospina-Quintero, V. Garzón-Marín, Á. Chantre-Astaiza, M. Muñoz-Organero y G. Ramírez-González, «Enriching public displays ads recommendations using an individual: Group cooperation model», en *Proceedings of The International Symposium on Pervasive Displays*, ép. PerDis '14, Copenhagen, Denmark: ACM, 2014, 186:186-186:187, ISBN: 978-1-4503-2952-1. DOI: 10.1145/2611009.2617196. dirección: <http://doi.acm.org/10.1145/2611009.2617196>.
- [133] F. M. Pabón, V. G. Marín, J. C. O. Quintero, J. J. I. Samboní, J. C. Guerrero, Á. C. Astaiza y G. R. González, «Sistema de recomendaciones para entornos de digital signage soportado en un esquema de cooperación smart tv-smartphone», *Revista Ingenierías Universidad de Medellín*, vol. 13, n.º 24, 2014.
- [134] F. Martínez-Pabón, J. Caicedo-Guerrero, J. J. Ibarra-Samboní, G. Ramírez-González y D. Hernández-Leo, «Smart tv-smartphone multiscreen interactive middleware for public displays», *The Scientific World Journal*, vol. 2015, 2015.

## BIBLIOGRAFÍA

---

- [135] S. Vargas y P. Castells, «Rank and relevance in novelty and diversity metrics for recommender systems», en *Proceedings of the fifth ACM conference on Recommender systems*, ACM, 2011, págs. 109-116.
- [136] J. S. Breese, D. Heckerman y C. Kadie, «Empirical analysis of predictive algorithms for collaborative filtering», en *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1998, págs. 43-52.