

Pontificia Universidad Javeriana Cali  
Facultad de Ingeniería.  
Ingeniería de Sistemas y Computación.  
Anteproyecto de Grado.

## Sistemas de recomendación basados en la inferencia de redes sociales

Santiago Uribe Pastás

Director: Dr. Jorge Finke Ortiz  
Co-director: Dr. Carlos Ernesto Ramírez Ovalle

Marzo 31 de 2021





Santiago de Cali, Marzo 31 de 2021.

Señores

**Pontificia Universidad Javeriana Cali.**

Dr. Gerardo Mauricio Sarria Montemiranda

Director Carrera de Ingeniería de Sistemas y Computación.

Cali.

Cordial Saludo.

Por medio de la presente me permito informarle que el estudiante de Ingeniería de Sistemas y Computación Santiago Uribe Pastás (cod: 8925546) trabaja bajo nuestra dirección en el proyecto de grado titulado “Sistemas de recomendación basados en la inferencia de redes sociales”.

Atentamente,

---

Dr. Jorge Finke Ortiz

---

Dr. Carlos Ernesto Ramírez Ovalle

Santiago de Cali, Marzo 31 de 2021.

Señores

**Pontificia Universidad Javeriana Cali.**

Dr. Gerardo Mauricio Sarria Montemiranda

Director Carrera de Ingeniería de Sistemas y Computación.

Cali.

Cordial Saludo.

Me permito presentar a su consideración el anteproyecto de grado titulado “Sistemas de recomendación basados en la inferencia de redes sociales” con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el proyecto de grado y posteriormente optar al título de Ingeniero de Sistemas y Computación.

Al firmar aquí, doy fe que entiendo y conozco las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería aprobadas el 26 de Noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del anteproyecto y del trabajo de grado.

Atentamente,

---

Santiago Uribe Pastás

Código: 8925546

# Resumen

Los sistemas de recomendación son una parte vital para empresas con una activa participación en la web. Dichas empresas requieren de estrategias que les permitan apalancarse en las calificaciones de los usuarios en sus productos, para poder brindar futuras recomendaciones a otros usuarios. En la última década se han desarrollado diversos algoritmos para la recomendación de películas, uno de los más populares es *Iterative Matrix Factorization*. El enfoque de esta tesis es evaluar el desempeño de este modelo de recomendación, infiriendo redes sociales a partir del conjunto de datos de *MovieLens*, el cual consta de 100836 valoraciones de 610 usuarios en 9742 películas, durante el 29 de marzo de 1996 y el 24 de septiembre de 2018. La tesis considera tanto algoritmos para la inferencia de redes sociales como diferentes métricas sobre estas redes inferidas. Nuestra mayor motivación es entender cómo el uso de la información de las redes sociales podría mejorar potencialmente la precisión del algoritmo *Iterative Matrix Factorization*.

**Palabras Clave:** Sistemas de Recomendación, Redes Sociales Inferidas, Recomendación de Películas, *Iterative Matrix Factorization*.



# Índice general

<b>1. Descripción del Problema</b>	<b>11</b>
1.1. Planteamiento del Problema . . . . .	11
1.1.1. Formulación . . . . .	12
1.1.2. Sistematización . . . . .	12
1.2. Objetivos . . . . .	12
1.2.1. Objetivo General . . . . .	12
1.2.2. Objetivos Específicos . . . . .	12
1.3. Justificación . . . . .	12
1.4. Delimitaciones y Alcances . . . . .	13
1.4.1. Entregables . . . . .	13
<b>2. Desarrollo del Proyecto</b>	<b>15</b>
2.1. Marco de Referencia . . . . .	15
2.1.1. Áreas Temáticas . . . . .	15
2.1.2. Marco Teórico . . . . .	15
2.1.3. Trabajos Relacionados . . . . .	15
2.2. Metodología . . . . .	17
2.2.1. Tipo de Estudio . . . . .	17
2.2.2. Actividades . . . . .	17
2.3. Resultados Esperados . . . . .	18
2.4. Cronograma . . . . .	18
2.5. Recursos . . . . .	18
2.5.1. Humanos . . . . .	18
2.5.2. Técnicos . . . . .	18
2.5.3. Presupuesto . . . . .	19
<b>Bibliografía</b>	<b>21</b>





# Introducción

Con el aumento cada vez mayor del volumen de la información en línea, los sistemas de recomendación (SR) han sido una herramienta utilizada para facilitarle al usuario a encontrar información que satisfaga sus necesidades e intereses, para así incrementar sus niveles de satisfacción. Plataformas actuales de *streaming* (como Netflix, Amazon Prime Video y HBO) aprovechan una variedad de conjuntos de datos diferentes y técnicas de *Machine Learning* para recomendar material a sus usuarios [1].

En los últimos años, se han generado diferentes algoritmos para la recomendación de películas, demostrando su efectividad cuando se aplican [2]; el filtrado colaborativo es uno de los enfoques más utilizados. Aunque los modelos generan buenos resultados, estos no explotan patrones ocultos en los conjuntos de datos que permitan agregar valor a la empresa.

En cuanto a las redes sociales inferidas, estas permiten ver como se encuentran conectados los usuarios de cierto conjunto de datos; esto es útil para determinar como se difunde la información de cierto producto entre los usuarios de dicha red, lo que permitiría recomendar nuevos o diferentes productos a los usuarios.

Por lo anterior, el propósito de esta investigación es evaluar la mejora en el desempeño del algoritmo de *Iterative Matrix Factorization* para la recomendación de películas, teniendo en cuenta modelos de redes sociales inferidas y sus métricas.



# Descripción del Problema

---

## 1.1. Planteamiento del Problema

Actualmente la información en internet crece exponencialmente, diversas paginas de ventas o servicios contienen diferente material para mostrar a una persona que ingrese a sus sitios web. Para ayudar a estas empresas a centrar el foco de sugerencias de productos en solo una parte de la información, existen los sistemas de recomendación. Estos sistemas son algoritmos que generan recomendaciones a partir de los meta-datos del producto o del usuario, para ayudar a las personas en procesos de toma de decisiones.

En 2012 Gomez *et al.* [3] propusieron que detrás de los conjuntos de datos existe alguna red estática desconocida subyacente, sobre la cual se propaga la información. Al propagarse sobre la red se genera una *cascada* y gracias a estas, junto con otros datos, se puede inferir redes sociales. El algoritmo desarrollado por Gomez *et al.* se llama *NetInf* [4].

En la Universidad de Stanford en 2011 Fan *et al.* plantearon fusionar el trabajo de Gomez *et al.* junto con datos de clasificación de películas, para evaluar el algoritmo de inferencia en función de las recomendaciones [5]. Los autores a futuro sugirieron evaluar el algoritmo de Gomez *et al.* en un conjunto de datos sin red social, como *MovieLens*, para comparar la precisión de las recomendaciones.

En 2017 Kose *et al.* compararon los algoritmos de *User-Based Collaborative Filtering*, *Iterative Matrix Factorization* y *Yehuda Koren's Integrated model*, para predecir las calificaciones que usuarios potenciales darían a películas [2]. Esta investigación dio como resultado un desempeño similar para los tres algoritmos, con respecto a su raíz del error cuadrático medio, o *RMSE* por sus siglas en inglés. Cabe resaltar que el algoritmo de *Iterative Matrix Factorization* destaca gracias a sus buenos resultados y a su sencilla implementación.

En general, los sistemas de recomendación actuales no toman en cuenta las estructuras internas que existen detrás de los conjuntos de datos, las cuales permiten saber como se encuentran conectados los usuarios. En consecuencia, surge la pregunta ¿De que manera mejora el desempeño de algoritmos de recomendación, haciendo uso de las métricas de redes sociales inferidas?

### 1.1.1. Formulación

¿Como se ve afectado el desempeño del algoritmo *Iterative Matrix Factorization* para la recomendación de películas, al añadir métricas de redes sociales inferidas?

### 1.1.2. Sistematización

- ¿Qué algoritmos o métodos existen para inferir redes sociales?
- ¿Qué modelos de redes sociales se pueden inferir a partir de conjunto de datos?
- ¿Cuáles métricas de redes sociales sirven para realizar mejores predicciones?
- ¿Como se ve afectado la precisión en las predicciones del algoritmo *Iterative Matrix Factorization* agregando métricas de redes sociales inferidas?

## 1.2. Objetivos

### 1.2.1. Objetivo General

El objetivo de esta tesis es hacer uso de diferentes métodos para inferir redes sociales en base al conjunto de datos de *MovieLens* [6], para evaluar el beneficio de inferir una red social a partir de las valoraciones y utilizar las métricas de esta red, a fin de realizar recomendaciones utilizando el algoritmo *Iterative Matrix Factorization*.

### 1.2.2. Objetivos Específicos

- **O1:** Investigar diferentes métodos o algoritmos para la inferencia de redes sociales.
- **O2:** Explorar los diferentes modelos de redes sociales que se pueden inferir.
- **O3:** Determinar que métricas de redes sociales generan mejores resultados en las predicciones.
- **O4:** Analizar las métricas de evaluación para el algoritmo *Iterative Matrix Factorization*, con el fin de determinar los beneficios de añadir propiedades de redes sociales inferidas.

## 1.3. Justificación

El centro de esta investigación se fundamenta en encontrar diferentes modelos de redes sociales en el conjunto de datos de *MovieLens*; para evaluar que tan beneficioso es el inferir una red social a partir de valoraciones y utilizar las métricas de esta red, en un algoritmo de recomendación de películas (*Iterative Matrix Factorization*).

La finalidad de este trabajo es que cualquier organización de servicio o consumo pueda usar los beneficios de esta investigación, para mejorar notablemente sus sistemas de recomendación. Esto

se podría reflejar en recomendaciones por estructuras de comunidad y no en un único usuario.

Mejorar los sistemas de recomendación implicaría una optimización de recursos, ampliar la cobertura y la accesibilidad a sus clientes; lo cual se reflejaría en mayores beneficios financieros para las organizaciones que lo adopten.

## 1.4. Delimitaciones y Alcances

Teniendo en cuenta los conocimientos en Teoría de Redes, Algoritmia y *Machine Learning* adquiridos en las carreras de Matemáticas Aplicadas e Ingeniería de Sistemas y Computación se pretende extraer diferentes modelos de redes sociales del conjunto de datos *MovieLens* con diferentes técnicas, para agregar las propiedades de estas redes al algoritmo de recomendación *Iterative Matrix Factorization*. Y comparar las métricas de evaluación de este algoritmo, para determinar los cambios en su desempeño.

De esta forma lograr el objetivo de evaluar que tan beneficioso es el inferir una red social a partir de las valoraciones y utilizar las métricas de esta red para realizar recomendaciones de películas. Así se optimizarían recursos, se ampliaría la cobertura y la accesibilidad a los clientes de organizaciones no solo de películas, sino de servicios y consumo.

### 1.4.1. Entregables

Una librería de Python que se pueda descargar e importar; la cual permite inferir redes sociales de diferentes maneras a partir de conjuntos de datos; para poder realizar recomendaciones.



# Desarrollo del Proyecto

---

## 2.1. Marco de Referencia

### 2.1.1. Áreas Temáticas

A continuación se presentan las categorías relacionadas con este proyecto:

- Information systems → Information systems applications → Data mining → Collaborative filtering.
- Mathematics of computing → Discrete mathematics → Graph theory.
- Networks → Social Networks → Network Properties.
- Computing methodologies → Machine learning → Machine learning approaches → Factorization methods → Non-negative matrix factorization.

### 2.1.2. Marco Teórico

Para el desarrollo de la presente propuesta se deben tener en cuenta conocimientos tanto de Ingeniería de Sistemas como de Matemáticas. A continuación se presentan los conceptos y las bases teóricas sobre las que se trabajará en el desarrollo del proyecto.

#### 2.1.2.1. Filtrado Colaborativo (FC)

#### 2.1.2.2. Factorización de Matrices (FM)

#### 2.1.2.3. Redes Sociales Inferidas

#### 2.1.2.4. Métricas de Redes Sociales

### 2.1.3. Trabajos Relacionados

- Gomez *et al.* [3] propusieron que detrás de ciertos conjuntos de datos existe alguna red estática desconocida subyacente, sobre la cual se propaga la información. Al propagarse sobre la red se genera una *cascada* y gracias a estas, junto con otros datos, se puede inferir redes sociales. Como resultado, basándose en un modelo probabilístico generativo, desarrollaron un algoritmo escalable para la inferencia de redes sociales llamado *NetInf*.

- Fan *et al.* [5] plantearon fusionar el trabajo de Gomez *et al.* junto con datos de clasificación de películas. Propusieron el algoritmo *MOVINF*, para predecir una red social mediante cascadas de clasificación de películas. Este algoritmo mejora la complejidad y la precisión del algoritmo *NetInf* bajo ciertas condiciones.
- Koren *et al.* [7] presentan diferentes técnicas de factorización matricial para sistemas de recomendación, como la descomposición en valores singulares (SVD por sus siglas en inglés), el descenso de gradiente estocástico y mínimos cuadrados alternos. Debido a su primer puesto en *Netflix Prize Competition* (competencia que buscaba el mejor algoritmo de filtrado colaborativo para predecir las valoraciones de los usuarios sobre películas) en 2007 y 2008, Koren *et al.* indican que estos son los métodos más exitosos y dominantes dentro de los recomendadores de filtrado colaborativo.
- Ma *et al.* [8] basándose en que la red social de un usuario afectará los comportamientos de dicho usuario, presentan un marco de recomendación social que fusiona una matriz de calificación de elementos de usuario con la red social del usuario utilizando factorización de matriz probabilística.
- He *et al.* [9] proponen un sistemas de recomendación basado en redes sociales. Este sistema utiliza la información de las redes sociales, incluidas las preferencias de los usuarios, la aceptación general de los artículos y la influencia de los amigos sociales. A partir de esa información desarrollan un modelo probabilístico que les permite hacer recomendaciones personalizadas.
- Liu *et al.* [10] proponen un método de recomendación basado en la red local de confianza en el ámbito de la recomendación de películas. Para esto infieren la red social y las múltiples fuentes de opiniones de los usuarios para generar una red de confianza de los usuarios. Posteriormente identifica a los usuarios con valoraciones comparables de forma fiable y realiza una predicción basada en la red local de confianza.
- Wang *et al.* [11] proponen ver el problema de recomendaciones como un problema de predicción de enlaces en redes bipartitas. Para dar solución a esto, propusieron el algoritmo *SRNMF*, que toma explícitamente las características latentes de los nodos junto con su estructura topológica intrínseca y codifica la información geométrica de la red mediante la construcción de una matriz basada en similitudes. En comparación con otros 17 métodos de predicción de enlaces, el método *SRNMF* es significativamente superior en términos de precisión y estabilidad.
- Li *et al.* [12] plantean que la factorización matricial (FM) estándar no captura las correlaciones estructurales jerárquicas, por ende, proponen una técnica de Factorización Matricial Jerárquica Oculta (*HHMF*), que aprende la estructura jerárquica oculta a partir de los registros de usuario-artículo. Esta técnica no requiere el conocimiento previo de la estructura jerárquica; por lo tanto, puede aplicarse cuando esta información sea explícita o implícita. *HHMF* supera a los métodos tradicionales de FM, a los métodos de FM jerárquica y a los métodos basados en redes neuronales.



- Gasparetti *et al.* [13] revisan la técnica de recomendación social basada en la detección de comunidades. Hacen énfasis en que grupos de usuarios que tienen características sociales más similares puede proporcionar evidencia valiosa y adicional para el proceso de recomendación. Sin embargo, debido a que se requieren varios pasos para las técnicas de detección de comunidades en sistemas de recomendación, esto hace que sea un objetivo mas que todo teórico y prácticamente desafiante.

## 2.2. Metodología

### 2.2.1. Tipo de Estudio

Esta investigación es de tipo teórico-experimental. Al entregar una librería de Python es necesario crear un modelo computacional, el cual incluye la inferencia de redes sociales, las métricas de estas redes y el algoritmo *Iterative Matrix Factorization* para la recomendación. Una vez el modelo esta planteado, este se prueba con el conjunto de datos *MovieLens*.

### 2.2.2. Actividades

Para alcanzar los objetivos específicos de esta investigación, se realizarán las siguientes actividades:

- **A1: Revisión sistemática de la literatura.** Revisar los estudios previos que se reportan en artículos, libros, conferencias, etc relacionados con sistemas de recomendación e inferencia de redes sociales.
- **A2: Análisis de los datos y almacenamiento en la base de datos.** Hacer un primer acercamiento al conjunto de datos *MovieLens*. Posteriormente, almacenar esta información en Neo4j para poder visualizarla.
- **A3: Implementación de algoritmos.** Implementación y validación de los diversos algoritmos o métodos para la inferencia de redes sociales.
- **A4: Exploración de modelos en redes sociales.** A partir de los algoritmos implementados, explorar los diferentes modelos de redes sociales que se pueden generar.
- **A5: Elección de métricas a usar.** Determinar las mejores métricas de redes sociales para realizar predicciones.
- **A6: Extracción de métricas a las redes sociales.** Hacer uso de algoritmos para la extracción de las métricas escogidas.
- **A7: Evaluación del trabajo propuesto.** Probar el modelo obtenido con el conjunto de datos *MovieLens*, con el fin de analizar las métricas de evaluación (*accuracy*, *precision*, *recall*, *f1-score*) del algoritmo *Iterative Matrix Factorization*.
- **A8: Documentación.** Presentación de los resultados obtenidos y conclusiones.

## 2.3. Resultados Esperados

- Evidenciar la aplicación de los conocimientos adquiridos en las carreras de Matemáticas Aplicadas e Ingeniería de Sistemas y Computación.
- Una librería de Python que permita inferir redes sociales de diferentes maneras a partir de conjuntos de datos; para poder realizar recomendaciones.
- Obtener buenos resultados en cuanto a las predicciones, evidenciándose en las métricas de evaluación de *Iterative Matrix Factorization*.
- Entender como el uso de las redes sociales junto con sus propiedades, podría mejorar potencialmente la precisión de los sistemas de recomendación.
- Contribuir positivamente en el ámbito de los sistemas de recomendación y la inferencia de redes sociales.
- Publicar un artículo donde se evidencie el proceso y los resultados de este trabajo.

## 2.4. Cronograma

Cronograma por semana incluyendo las actividades descritas en la Sección de actividades

## 2.5. Recursos

### 2.5.1. Humanos

Jorge Finke Ortiz, docente del Departamento de Ingeniería Eléctrica y Ciencias de la Computación, miembro del grupo de investigación GAR, y miembro activo en la organización IEEE, es el director de esta propuesta.

Carlos Ernesto Ramírez Ovalle, docente del Departamento de Ciencias Naturales y Matemáticas y miembro del grupo de investigación EMAP, es el codirector de esta propuesta.

Santiago Uribe Pastás, estudiante de Ingeniería de Sistemas y Computación - Matemáticas Aplicadas, es el responsable de realizar la investigación propuesta.

### 2.5.2. Técnicos

Se hará uso del lenguaje de programación Python (junto con algunas de sus librerías) para construir algoritmos que transformen, procesen y analicen los datos correspondientes.

Suscripciones a los diferentes motores de búsqueda académicos (IEEE, ACM, Scopus y SpringerLink), para adquirir los artículos completos necesarios para esta investigación.

También se hará uso de la base de datos orientada a grafos Neo4j, para poder cargar y visualizar los datos de la manera necesaria para el enfoque propuesto.

### 2.5.3. Presupuesto

En la Tabla 2.1 se presentan los valores de todos los recursos a usar en el proyecto.

Recurso	Cantidad	Valor
Recursos Humanos	3	12,000,000
Software	3	0
Computador	1	2,500,000
<b>Total</b>	-	14,500,000

Tabla 2.1: Tabla de Presupuesto



# Bibliografía

- [1] Netflix, “How Netflix’s Recommendations System Works.” <https://help.netflix.com/en/node/100639/us>. Online. [Accessed: 27- Mar- 2021].
- [2] A. Kose, C. Kanbak, and N. Evirgen, “Performance comparison of algorithms for movie rating estimation,” in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, pp. 955–959, Institute of Electrical and Electronics Engineers Inc., 2017.
- [3] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, “Inferring networks of diffusion and influence,” in *ACM Transactions on Knowledge Discovery from Data*, vol. 5, (New York, New York, USA), pp. 1–37, ACM PUB27, feb 2012.
- [4] J. Leskovec and R. Sosič, “SNAP: A general-purpose network analysis and graph-mining library,” *ACM Transactions on Intelligent Systems and Technology*, vol. 8, pp. 1–20, jul 2016.
- [5] C. Fan and L. Yu, “Inferring Social Networks Based on Movie Rating Data,” tech. rep., Stanford University, CA, 2011.
- [6] F. M. Harper and J. A. Konstan, “The movielens datasets: History and context,” *ACM Transactions on Interactive Intelligent Systems*, vol. 5, pp. 1–19, dec 2015.
- [7] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [8] H. Ma, H. Yang, M. R. Lyu, and I. King, “SoRec: Social recommendation using probabilistic matrix factorization,” in *International Conference on Information and Knowledge Management, Proceedings*, (New York, New York, USA), pp. 931–940, ACM Press, 2008.
- [9] J. He and W. Chu, “A social network-based recommender system (snrs),” in *Data Mining for Social Network Data*, vol. 12, pp. 47–74, jun 2010.
- [10] B. Liu and Z. Yuan, “Incorporating Social Networks and User Opinions for Collaborative Recommendation: Local Trust Network based Method,” in *Proceedings of the Workshop on Context-Aware Movie Recommendation - CAMRa ’10*, (New York, New York, USA), ACM Press, 2010.
- [11] W. Wang, X. Chen, P. Jiao, and D. Jin, “Similarity-based Regularized Latent Feature Model for Link Prediction in Bipartite Networks,” *Scientific Reports*, vol. 7, dec 2017.
- [12] H. Li, Y. Liu, Y. Qian, N. Mamoulis, W. Tu, and D. W. Cheung, “HHMF: hidden hierarchical matrix factorization for recommender systems,” *Data Mining and Knowledge Discovery*, vol. 33, pp. 1548–1582, nov 2019.

- [13] F. Gasparetti, G. Sansonetti, and A. Micarelli, “Community detection in social recommender systems: a survey,” *Applied Intelligence*, pp. 1–21, nov 2020.