

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА
АКАДЕМІЯ»

Кафедра математики факультету інформатики

Кваліфікаційна робота

освітній ступінь – бакалавр

на тему: **«СТВОРЕННЯ СОЦІАЛЬНОЇ МЕРЕЖІ
ВИКОРИСТОВУЮЧИ ТЕОРІЮ ГРАФІВ»**

Виконав студент

4-го року навчання спеціальності

113 "Прикладна математика"

Коваль Олег Русланович

Керівник: Борозенний С.О.

старший викладач

Рецензент: Афонін А.О.

Кваліфікаційна робота захищена

з оцінкою _____

Секретар ЕК _____

(підпис)

«_____» _____ 2022р.

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА
АКАДЕМІЯ»

Кафедра математики факультету інформатики

ЗАТВЕРДЖУЮ

Зав.кафедри математики,
проф., д.ф-м.н., Б. В. Олійник

_____ (підпис)

" ____ " _____ 2021 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

на кваліфікаційну роботу

студенту 4-го курсу, факультету інформатики Ковалю Олегу Руслановичу

Тема: Створити соціальну мережу з використанням теорії графів

Вихідні дані: Досвід у створенні сайтів та знання теорії графів

Зміст ТЧ до бакалаврської роботи:

Зміст

Анотація

Вступ

1. Соціальний граф
2. Колаборативна фільтрація
3. Створення соціальної мережі

Висновки

Список літератури

Додатки

Дата видачі „ ____ ” _____ 2021 р. Керівник _____ (підпис)
(підпис)

Завдання отримав _____
(підпис)

Тема: Створити соціальну мережу з використанням теорії графів

Календарний план виконання роботи:

№	Назва етапу дипломного проєкту (роботи)	Термін виконання етапу	Примітка
1.	Отримання завдання на дипломну роботу.	21.10.2021	
2.	Огляд технічної літератури за темою роботи.	15.03.2022	
3.	Аналіз соціального графа	29.03.2022	
4.	Створення інтерфейсу для соціальної мережі.	15.04.2022	
5.	Аналіз алгоритмів створення рекомендаційної системи.	01.05.2022	
6.	Написання теоретичної частини роботи.	1.06.2022	
7.	Аналіз отриманих результатів з керівником та передзахист.	14.06.2022	
8.	Корегування роботи за результатами попереднього захисту.	24.06.2022	
9.	Остаточне оформлення роботи.	27.06.2022	
10.	Захист кваліфікаційної роботи.	04.07.2022	

Студент _____

Керівник _____

" _____ "

Зміст

Анотація	5
Вступ	6
1 Соціальний граф	7
1.1 Теорія мереж	7
Визначення	7
Властивості мереж	7
1.2 Метрики	10
1.3 Сегментація	12
1.4 Пошук в соціальній мережі	12
2 Колаборативна фільтрація	13
2.1 Визначення	13
2.2 Проблеми колаборативної фільтрації	13
2.3 Існуючі алгоритми	14
Item-Base	14
User-Base	16
Алгоритм колаборативної фільтрації Probabilistic matrix factori- zation (PMF)	18
Алгоритм колаборативної фільтрації Bayesian probabilistic matrix factorization (BPMF)	19
Алгоритм колаборативної фільтрації Singular value decomposi- tion (SVD)	19
3 Створення соціальної мережі	21
3.1 Створення сайту для демонстрації можливостей алгоритму	21
3.2 Реалізація алгоритму	24
3.3 Результат	26
4 Висновок	28

5 Джерела**28**

Анотація

В роботі досліджуються можливості використання соціального графу при створенні соціальної мережі, його властивості та метрики на цьому графі. Розглядаються також можливості використання різних алгоритмів колаборативної фільтрації, таких як Probabilistic matrix factorization, Bayesian probabilistic matrix factorization, Singular value decomposition. В практичній частині створюється соціальна мережа, з використанням Python та мови розмітки.

Ключові слова: соціальний граф, теорія графів, рекомендаційні системи.

Вступ

Соціум зараз - це те, без чого люди не можуть уявити свого життя. Кожного дня ми вимушені сотні разів взаємодіяти з оточенням для отримання того чи іншого результату. Так було й раніше, але тепер це легше ніж будь-коли, бо з'явився інтернет, а разом з ним і можливість швидко обмінюватись інформацією. Згодом вигадали соціальні мережі, однак нинішній їх вигляд не можна уявити без тих функцій, які вони надають. Легкий пошук знайомих людей та спільнот, спілкування з людьми з усього світу та багато іншого. Ці можливості надають користувачам величезний вибір співрозмовників, тобто можна спілкуватися з будь-якою людиною із будь-якої іншої країни. Однак на практиці ми бачимо трохи іншу картину: користувачі надають перевагу спілкуванню з людьми які знаходяться близько до них та мають схожі інтереси. Це легко можна перевірити, створивши та проаналізувавши граф, вершини якого створені з користувачів соціальної мережі. Однак аналіз може не тільки допомогти зробити висновки щодо користувачів створеної мережі, але ще й створити систему рекомендацій друзів для зручності користувачів. Основним алгоритмом для пошуку потенційних друзів у соціальних мережах є колаборативна фільтрація, у якої є велика кількість реалізацій. Алгоритм колаборативної фільтрації має як переваги, так і недоліки, однак є найбільш розповсюдженим алгоритмом і саме тому він є дуже актуальним. Аналіз колаборативної фільтрації та соціального графу, створеного на основі користувачів у соціальній мережі, і є цілком даної роботи.

1 Соціальний граф

1.1 Теорія мереж

Визначення

- Граф — це множина об'єктів та зв'язків між ними. Об'єкти є вершинами (вузлами) графу, а зв'язки — ребра. Вузли (в соціальному графі) — це об'єкти (зазвичай) в мережі та мають два види властивостей: власні та мережеві. До власних може відноситись вік, ім'я, локація тощо, а до мережевих — кількість сусідів (ступінь) та компонента зв'язності.
- Ребра (в соціальному графі) — це зв'язки між вузлами в соціальному графі, які у деяких випадках також містять властивості. Наприклад, вага, яка визначає силу з'єднання, напрямок і т.д.
- Соціальний граф (англ. Social graph) — це граф, в якому вершини представлені вузлами з особливостями користувачів (або об'єктів соціальної мережі), а зв'язки — ребрами в такому графі.

Властивості мереж

Соціальна мережа, побудована на основі соціального графу, має атрибути, які можуть бути обчислені для аналізу властивостей та характеристик такої мережі.

- **Розмір**

Розмір мережі зазвичай визначається числом вузлів N або числом ребер E . При цьому, кількість ребер може варіюватись від $N-1$ у деревах до E_{\max} у повних графах. Є різні варіанти обрахування розміру графу за вершинами:

- Якщо у графі між парою вершин може бути виключно одне неорієнтоване ребро і не містить петлі, то його формула буде виглядати так: $E_{\max} = \binom{N}{2} = N(N-1)/2$.

- Якщо цей граф орієнтований, але не має петель, то його формула буде виглядати так: $E_{\max} = N(N - 1)$.
- Якщо це орієнтований граф з петлями, то його формула буде виглядати так: $E_{\max} = N^2$.
- Якщо в графі також можливі кратні ребра, то його формула буде виглядати так: $E_{\max} = \infty$.

• Щільність

Щільність мережі D з N вузлами визначається як відношення кількості ребер E до кількості усіх можливих ребер у мережі і, якщо це простий граф, задається біноміальним коефіцієнтом $\binom{N}{2}$, який можна описати двома варіантами:

- $D = \frac{E - (N - 1)}{E_{\max} - (N - 1)} = \frac{2(E - N + 1)}{N(N - 3) + 2}$
- $D = \frac{T - 2N + 2}{N(N - 3) + 2}$, де зв'язки T є неорієнтованими.

• Планарна мережева щільність

Планарна мережева щільність мережі D , визначається так як і її щільність, однак виключає перетини ребер. $E_{\max} = 3N - 6$, тому формула змінюється на таку $D = \frac{E - N + 1}{2N - 5}$.

• Середній степінь вузла

Степінь вузла k — це кількість ребер, пов'язаних з ним.

- $\langle k \rangle = \frac{2E}{N}$ для неорієнтовних графів
- $\langle k \rangle = \frac{E}{N}$ для орієнтовних графів

• Середня довжина найкоротшого шляху

Для того, щоб обрахувати таку довжину потрібно знайти найкоротший шлях між усіма парами вузлів та обчислити середню довжину

за всіма знайденими шляхами. Довжиною називається кількість ребер, що містяться в шляху, тобто відстань $d_{u,v}$ між двома вершинами u, v в графі.

За цією характеристикою можна зробити висновок щодо того, чи відображає мережа ефект тісного світу. Якщо її можна описати формулою $O(\ln N)$, то вона є мережею малого світу. Якщо ж це значення не більше за логарифмічне, то така модель є моделлю "тісного світу". Дана модель також відома за "теорією шести рукошлягань".

- **Діаметр**

Діаметр - це довжина $\max_{u,v} d(u, v)$ "найдовшого найкоротшого шляху" між будь-якими двома вершинами графа (u, v) , де $d(u, v)$ — відстань графіка. Тобто, це найбільша кількість вершин, яку необхідно подолати для переходу з однієї вершини до іншої, йдучи найкоротшим шляхом.

- **Коефіцієнт кластеризації**

Коефіцієнт кластеризації вузла дорівнює відношенню зв'язків, які існують між сусідами вузла. Коефіцієнт кластеризації мережі дорівнює середньому коефіцієнту кластеризації всіх вузлів. Високий коефіцієнт кластеризації є характерною ознакою мережі "тісного світу".

Коефіцієнт кластеризації i -го вузла дорівнює

$C_i = \frac{2e_i}{k_i(k_i - 1)}$, де $k_i k_i$ — число сусідів даного вузла, а $e_i e_i$ — число зв'язків між ними. Максимальна кількість можливих зв'язків дорівнює

$$\binom{k}{2} = \frac{k(k-1)}{2}$$

- **Зв'язність**

Мережі поділяються на чотири категорії, в залежності від наявної зв'язності:

- Повний граф — повністю зв’язна мережа, кожна вершина якої має зв’язок з будь-якою іншою вершиною в мережі. Ці мережі є симетричними.
- Гігантська компонента — це група вершин, яка містить більшість вершин з мережі.
- Слабко-зв’язна компонента — це група вершин, в якій з кожної вершини є шлях в будь-яку іншу вершину. При цьому всі ребра є неорієнтованими.
- Сильно-зв’язна компонента — це група вершин, в якій з кожної вершини є шлях в будь-яку іншу вершину. При цьому всі ребра є орієнтованими..

• Центральність вузла

Показники центральності обраховуються для того, щоб встановити найважливіші вузли в моделі мережі. В залежності від цілі можна виділити кілька варіантів оцінки центральності:

- Центральність за посередництвом - вершина є важливою, якщо вона утворює багато мостів між іншими вершинами.
- Центральність за впливовістю - вершина є важливою, якщо з ним пов’язано багато інших важливих вузлів
- Центральність за степенем вузла мережі — це число зв’язків (вершин), інцидентних вузлу.
- Центральність за близькістю визначає, наскільки «близький» вузол мережі іншим вузлам обчисленням суми найкоротших відстаней між цим вузлом і іншими вузлами мережі.

1.2 Метрики

Для покращення аналізу соціального графу, вживаються метрики, які в числовій формі можуть відобразити характеристики вершин, сегментів або

груп вершин графу.

Наступні метрики показують характер зв'язків соціальних об'єктів:

* Міст — в теорії графів, міст це вершина, видалення якої збільшує кількість компонент зв'язності. В соціальному ж графі - це користувач, який поєднує два сегменти. * Густина — частка прямих зв'язків у мережі по відношенню до загального числа можливих. * Відстань — мінімальна кількість зв'язків, які потрібні, щоб встановити існування шляху між двома користувачами. * Структурні діри — відсутність зв'язків між двома частинами мережі.

Те, наскільки зв'язок є сильним, визначається трьома параметрами: - Лінійна комбінація часу - Близькість - Взаємність Сильним зв'язкам притаманна транзитивність, гомофілія та сусідство, в той час як слабким зв'язкам - мости.

Також варто надати визначення для метрик, які допомагають визначати силу зв'язку:

- Гомофілія — це рівень схильності користувача до утворення зв'язків з іншими користувачами зі схожими властивостями. Тобто, з користувачами схожого віку, статтю, хоббі і т.п. - Множинність — це число зв'язків, які має користувач. Тобто, якщо два користувачі додали один одного в друзі, їх рівень множинності рівняється 2. - Взаємність — це рівень схильності користувачів до взаємодії один з одним. Такий рівень залежить від того, чи відповідають користувачі взаємністю на дії один одного. - Мережева закритість — міра яка показує чи знаходиться користувач в мережі, в якій його друзі також є друзями один одному. Якщо користувач знаходиться в мережевій закритості, це може бути названо транзитивністю. - Сусідство — тенденція користувачів мати велику кількість зв'язків з географічно близькими користувачами.

1.3 Сегментація

Далі наведені визначення сегментів, які можуть бути присутні в соціальному графі.

Кліка — група, користувачі якої з'єднані один з одним напряму. Соціальне коло — група, користувачі якої не обов'язково з'єднані між собою. Коефіцієнт кластеризації — ступінь ймовірності того, що два різних користувача, пов'язані з конкретним користувачем. Чим вище цей коефіцієнт, тим вище замкненість груп, що значить, що така група може бути клікою. Згуртованість — міра, яка показує наскільки користувачі пов'язані між собою одним зв'язком. Якщо група є структурно згуртованою, це значить, видалення малої кількості користувачів з цієї мережі може призвести до розриву групи.

1.4 Пошук в соціальній мережі

Рекомендація друзів — користувачі зазвичай не ділять своїх друзів на соціальні групи, однак в залежності від їх поведінки в мережі, можна зробити висновки щодо їх приналежності.

Існує традиційний підхід при створенні рекомендаційної системи:

Колаборативна фільтрація — цей метод полягає у формуванні списку рекомендованих об'єктів на основі думок користувачів, провідних себе схожим чином та іншими схожими методами. Фільтрація вмісту — базується на характеристиках предмета та інформації про нього.

Колаборативна фільтрація - хороший інструмент, який має багато варіацій для кожного конкретного випадку. Деякі методи будуть наведені у наступному розділі.

2 Колаборативна фільтрація

2.1 Визначення

Системи рекомендацій часто працюють за допомогою методу, відомого як спільна фільтрація. Загалом колаборативна фільтрація стосується вибору інформації або шаблонів за допомогою кількох агентів, точок зору, джерел даних тощо. Спільна фільтрація часто використовується з надзвичайно великими наборами даних.

2.2 Проблеми колаборативної фільтрації

- Розрідженість даних - ця проблема існує з тієї причини, що користувачі не схильні залишати оцінки товарам, тому часто системі доводиться тільки здогадуватися чи сподобався товар користувачу, що, зрештою, заважає правильному функціонуванню.
- Масштабованість - чим більше користувачів, тим складніше зробити вибірку з усіх користувачів. Тобто, маючи 10 мільйонів покупців $O(M)O(M)$ і мільйон предметів $O(N)O(N)$, алгоритм колаборативної фільтрації зі складністю рівній $O(MN)O(MN)$ вже занадто складний для розрахунків
- Проблема холодного старту - нові товари та користувачі є дуже складною задачею для алгоритму з тієї причини, що у нього немає жодної кореляції з іншими користувачами/товарами.
- Шахрайство - оскільки колаборативна фільтрація часто використовує оцінки користувачів в якості даних, на основі яких потрібно робити висновки, таку систему легко обдурити, поставивши багато поганих (або хороших оцінок) з багатьох різних порожніх аккаунтів.
- Різноманітність - хоча рекомендаційні системи і створені для того, щоб відкривати користувачам нові продукти з безлічі предметів, ча-

сто так трапляється, що алгоритми можуть дати збій і рекомендувати тільки ті речі, які схожі на обрані користувачем предмети, що негативно впливає на різноманіття рекомендованих продуктів.

2.3 Існуючі алгоритми

Item-Base

Даний метод колаборативної фільтрації базується на тому, що користувачу сподобаються ті товари, які є схожими на ті, які він вже обирає.

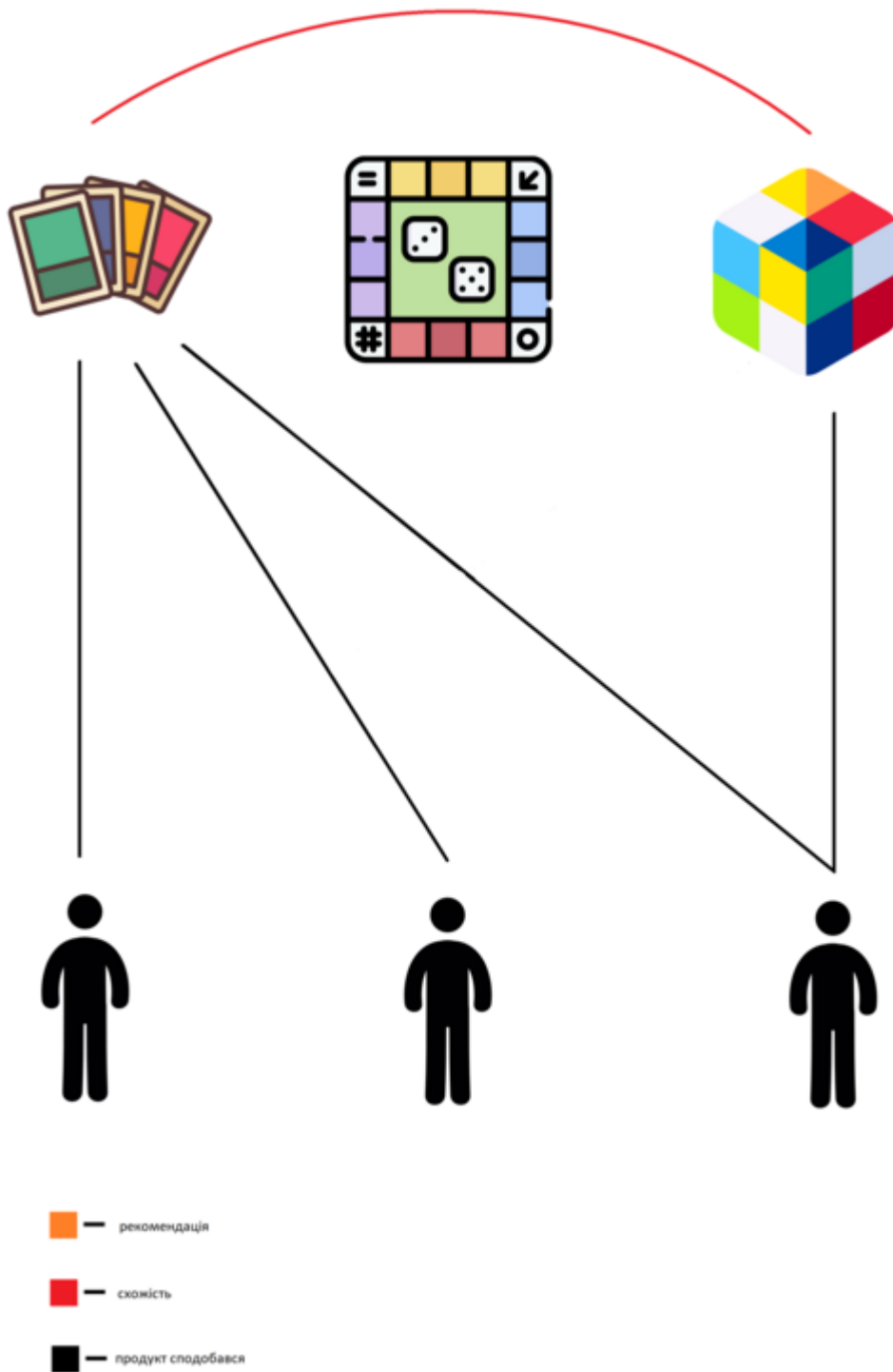


Рис 2.1

Умовний користувач А може бути описаним об'єктами, які він переглянув або оцінив. Для кожного оціненого об'єкта визначається m об'єктів-сусідів. Для цього потрібно знайти m найбільш схожих об'єктів з оцінених користувачами. Значення m може бути різним, в залежності від потреб. У випадку соціальної мережі, побудованої мною, $m=5$. Всі об'єкти-сусіди

об'єднуються в колекцію, з якої виключаються об'єкти, переглянуті або оцінені користувачем А. А з останків колекції будується топ n рекомендацій. Таким чином, при такому підході у виборі рекомендацій беруть участь всі користувачі, які оцінили об'єкти.

$$\bar{r}_i + \frac{1}{\sum_{i' \in R(u)} |sim(i, i')|} \sum_{i' \in R(u)} sim(i, i') (r_{u, i'} - \bar{r}_i')$$

Найпопулярнішими метриками схожості ($sim(u, u')$) на даний момент є кореляція Пірсона та косинуса відстань матриці.

User-Base

Даний метод колаборативної фільтрації базується на тому, що користувачу сподобаються ті товари, які є обирали схожі на нього інші користувачі.

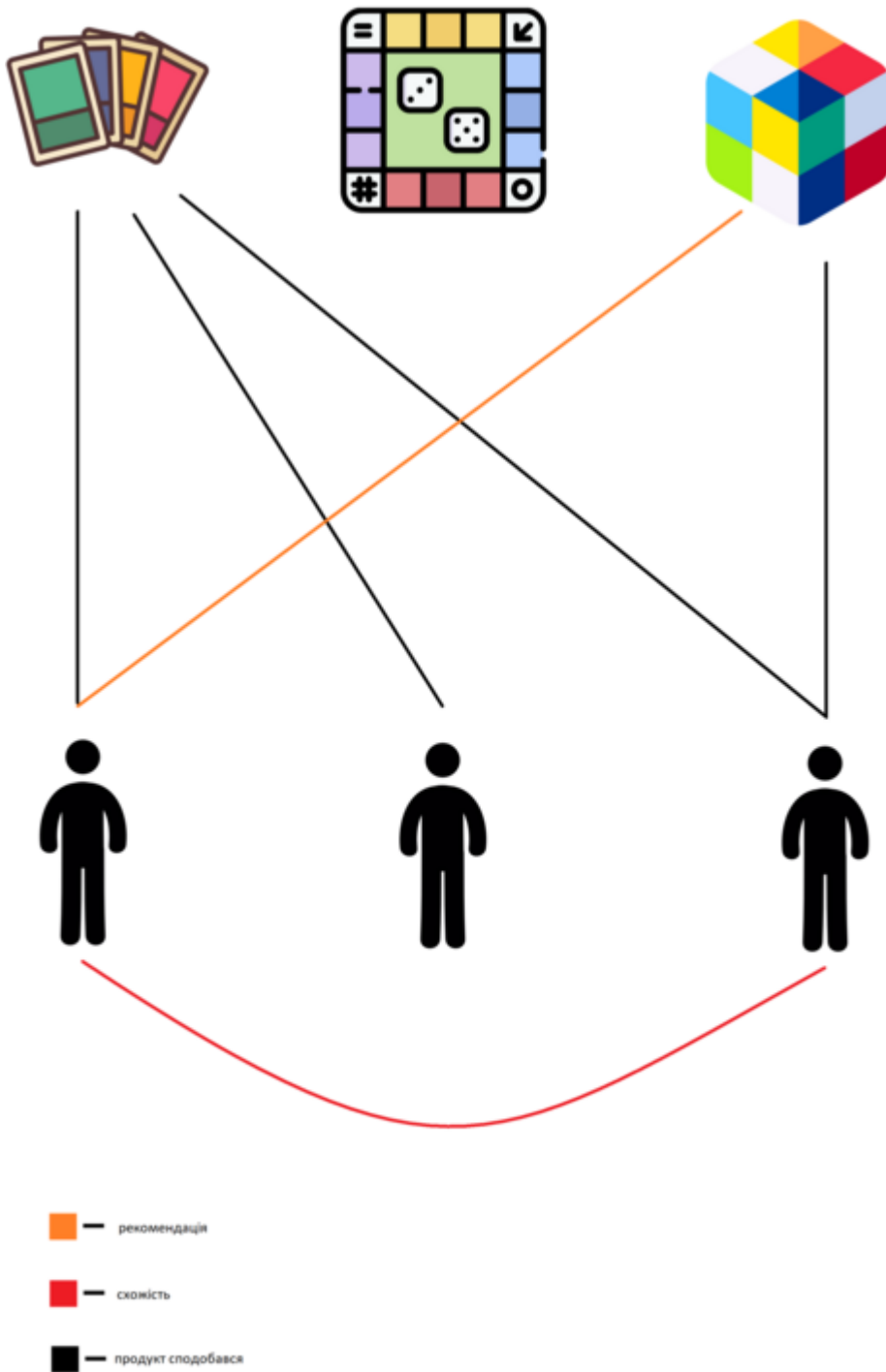


Рис 2.2

$$sim(u, u') = \frac{\sum_{i \in R(u) \wedge R(u')} (r_{u,i} - \bar{r}_u)(r_{u',i} - \bar{r}_{u'})}{\sum_{i \in R(u) \wedge R(u')} (r_{u,i} - \bar{r}_u)^2 (r_{u',i} - \bar{r}_{u'})^2}$$

$$sim(u, u') = \frac{\sum_{i \in R(u) \wedge R(u')} (r_{u,i} r_{u',i})}{\sqrt{\sum_{i \in R(u)} r_{u,i}^2 \sum_{i \in R(u')} r_{u',i}^2}}$$

Алгоритм колаборативної фільтрації Probabilistic matrix factorization (PMF)

Ймовірнісна матрична факторизація (PMF) - імовірнісна лінійна модель з гаусовим спостережним шумом. Припустимо, у нас є N користувачів і M фільми. Нехай R_{ij} стане рейтинговим значенням для користувача i для фільму j , U_i та V_j представляють D -розмірні специфічні для користувача і для фільму вектори латентної функції, відповідно. Умовний розподіл за спостереженими оцінками $R \in^{N \times M}$ (термін правдоподібності) та попередні розподіли за $U \in^{D \times N}$ і $V \in^{D \times M}$ і визначаються за допомогою:

$$p(R|U, V, \alpha) = \prod_{i=1}^N \prod_{j=1}^M [N(R_{ij}|U_i^T V_j, \alpha^{-1})]^{I_{ij}}$$

$$p(U|\alpha_U) = \prod_{i=1}^N N(U_i|0, \alpha_U^{-1}I)$$

$$p(V|\alpha_V) = \prod_{j=1}^M N(V_j|0, \alpha_V^{-1}I)$$

Де $N(x|\mu, \alpha^{-1})$ позначає гауссовський розподіл з середнім μ і точністю α , а I_{ij} - змінна-індикатор, має значення 1, якщо користувач оцінив ij -тий фільм та 0 - в іншому випадку. Навчання в цій моделі виконується шляхом максимізації лог-попереду над фільмом та функціями користувача з фіксованими гіперпараметрами (тобто дисперсією шумів спостережень та попередніми відхиленнями):

$$\ln p(U, V | R, \alpha, \alpha_U, \alpha_V) = \ln p(R | U, V, \alpha) + \ln p(U | \alpha_U) + \ln p(V | \alpha_V) + C$$

де C - це константа, яка не залежить від параметрів. Максимізація цього поперечного розподілу по відношенню до U та V еквівалентно мінімізації функції помилок суми квадратів з членами квадратичної регуляризації

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - U_i^T V_j)^2 + \frac{\lambda_U}{2} \sum_{i=1}^N \|U_i\|_{\text{Fro}}^2 + \frac{\lambda_V}{2} \sum_{j=1}^M \|V_j\|_{\text{Fro}}^2$$

де $\lambda_U = \alpha_U/\alpha$, $\lambda_V = \alpha_V/\alpha$, $\|\cdot\|_{\text{Fro}}^2$ позначає норму Фробеніуса.

Алгоритм колаборативної фільтрації Bayesian probabilistic matrix factorization (BPMF)

Попередні розподіли над векторами функцій користувача та фільму вважаються гауссовськими: $p(U | \mu_U, \Lambda_U) = \prod^N \mathcal{N}(U_i | \mu_U, \Lambda_U^{-1})$ $p(V | \mu_V, \Lambda_V) = \prod_{i=1}^M \mathcal{N}(V_i | \mu_V, \Lambda_V^{-1})$

Надалі ми розміщуємо прихильників Gaussian-Wishart на користувацькому та гіперпараметрах фільму

$$\Theta_U = \{\mu_U, \Lambda_U\} \text{ і } \Theta_V = \{\mu_V, \Lambda_V\} :$$

$$p(\Theta_U | \Theta_0) = p(\mu_U | \Lambda_U) p(\Lambda_U) = \mathcal{N}(\mu_U | \mu_0, (\beta_0 \Lambda_U)^{-1}) \mathcal{W}(\Lambda_U | W_0, \nu_0)$$

$$p(\Theta_V | \Theta_0) = p(\mu_V | \Lambda_V) p(\Lambda_V) = \mathcal{N}(\mu_V | \mu_0, (\beta_0 \Lambda_V)^{-1}) \mathcal{W}(\Lambda_V | W_0, \nu_0)$$

Тут W - розподіл Wishart з ν_0 ступенями свободи і матриця масштабу $D \times D$ W_0 :

$$\mathcal{W}(\Lambda | W_0, \nu_0) = \frac{1}{C} |\Lambda|^{(\nu_0 - D - 1)/2} \exp\left(-\frac{1}{2} \text{Tr}(W_0^{-1} \Lambda)\right)$$

де C - нормалізаційна константа. Для зручності ми також визначаємо $\theta_0 = \mu_0, \nu_0, W_0$. У наших експериментах ми також встановили $\nu_0 = D$ і W_0 на матрицю тотожності як для гіперпараметрів користувача, так і для фільмів, і вибираємо $\mu_0 = 0$ за симетрією.

Алгоритм колаборативної фільтрації Singular value decomposition (SVD)

Майже всі алгоритми колаборативної фільтрації мають такі недоліки, як холодний старт, тривіальність результатів рекомендацій тощо. SVD алгоритм є новішою варіацією колаборативної фільтрації, який дозволяє покращити звичні алгоритми.

SVD - це методика факторизації матриць, яка зазвичай використовується для зменшення кількості функцій набору даних шляхом зменшення розмірів простору від N до K , де $K < N$. Однак для систем рекомендацій нас цікавить лише матрична факторизація, де частини зберігають таку ж розмірність. Матрична факторизація виконується на матриці рейтингів користувацьких позицій. Матрична факторизація може розглядатися як

пошук 2 матриць.

3 Створення соціальної мережі

3.1 Створення сайту для демонстрації можливостей алгоритму

Для демонстрації можливостей алгоритму було створено сайт, який може робити мінімальну, але необхідну кількість дій для роботи алгоритму. Найголовніший створений функціонал, який потрібен для реалізації колаборативної фільтрації, це можливість зареєструватись, зайти до себе в аккаунт та знайти і додати собі нового друга. Виглядає цей сайт ось так:

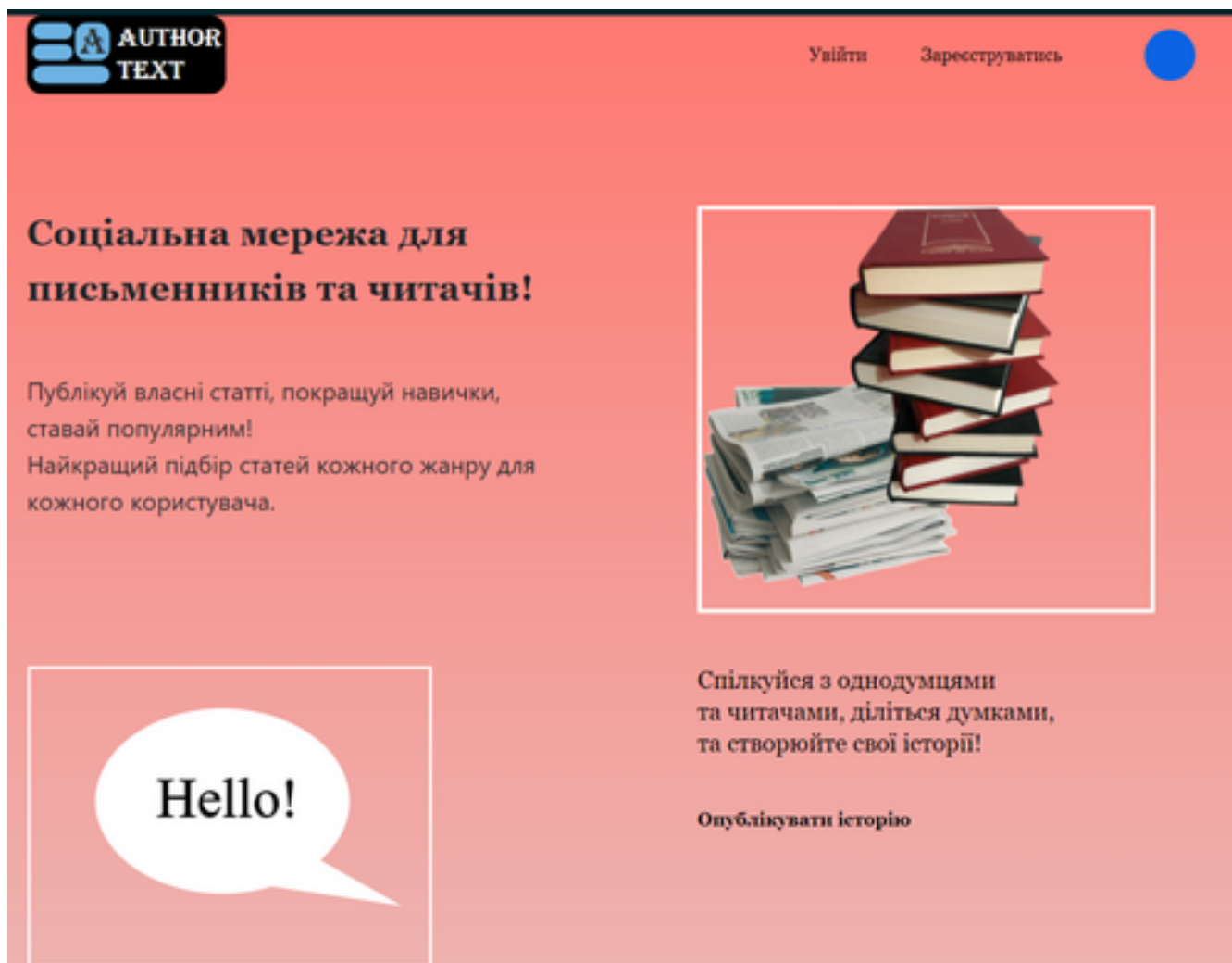


Рис 3.1.1 (головна сторінка)

The image shows a registration form titled "Реєстрація" (Registration) on a dark red background. The form is centered and contains the following elements:

- Logo:** "AUTHOR TEXT" with a stylized 'A' icon in the top left corner.
- Language:** "Українська" (Ukrainian) in the top right corner.
- Title:** "Реєстрація" (Registration) in large white text.
- Form Fields:**
 - Ім'я:** (Name) with the value "first_test" entered.
 - Пароль:** (Password) with masked characters "*****".
- Buttons:**
 - A large blue button labeled "Зареєструватися" (Register).
 - A smaller blue link below it labeled "Вже маєте акаунт?" (Already have an account?).

Рис 3.1.2 (реєстрація)

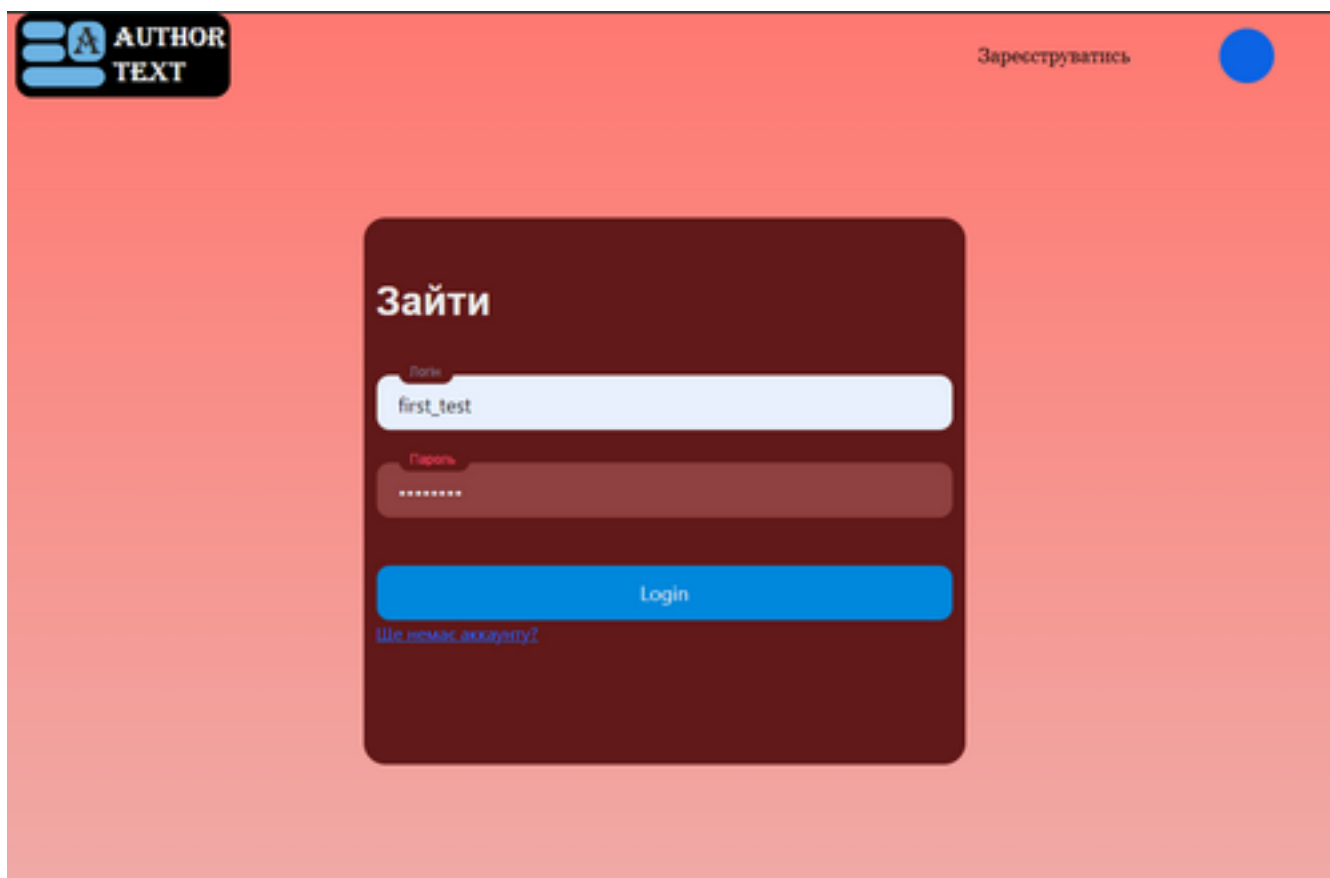


Рис 3.1.3 (вхід)



Рис 3.1.4 (новий профіль)



Рис 3.1.5 (профіль з доданими друзями)

3.2 Реалізація алгоритму

Оскільки інформації про користувачів недостатньо, бо створений сайт не збирає додаткову інформацію про людей, було використано алгоритм, який базується на User-Base алгоритмі для фільтрації. При запусканні серверу, на основі відомих даних генерується граф, вершинами якого є користувачі, а зв'язками - чи є два користувачі друзями, чи ні. До цього графу і застосовується алгоритм для пошуку рекомендацій.

На даний момент граф користувачів виглядає так:

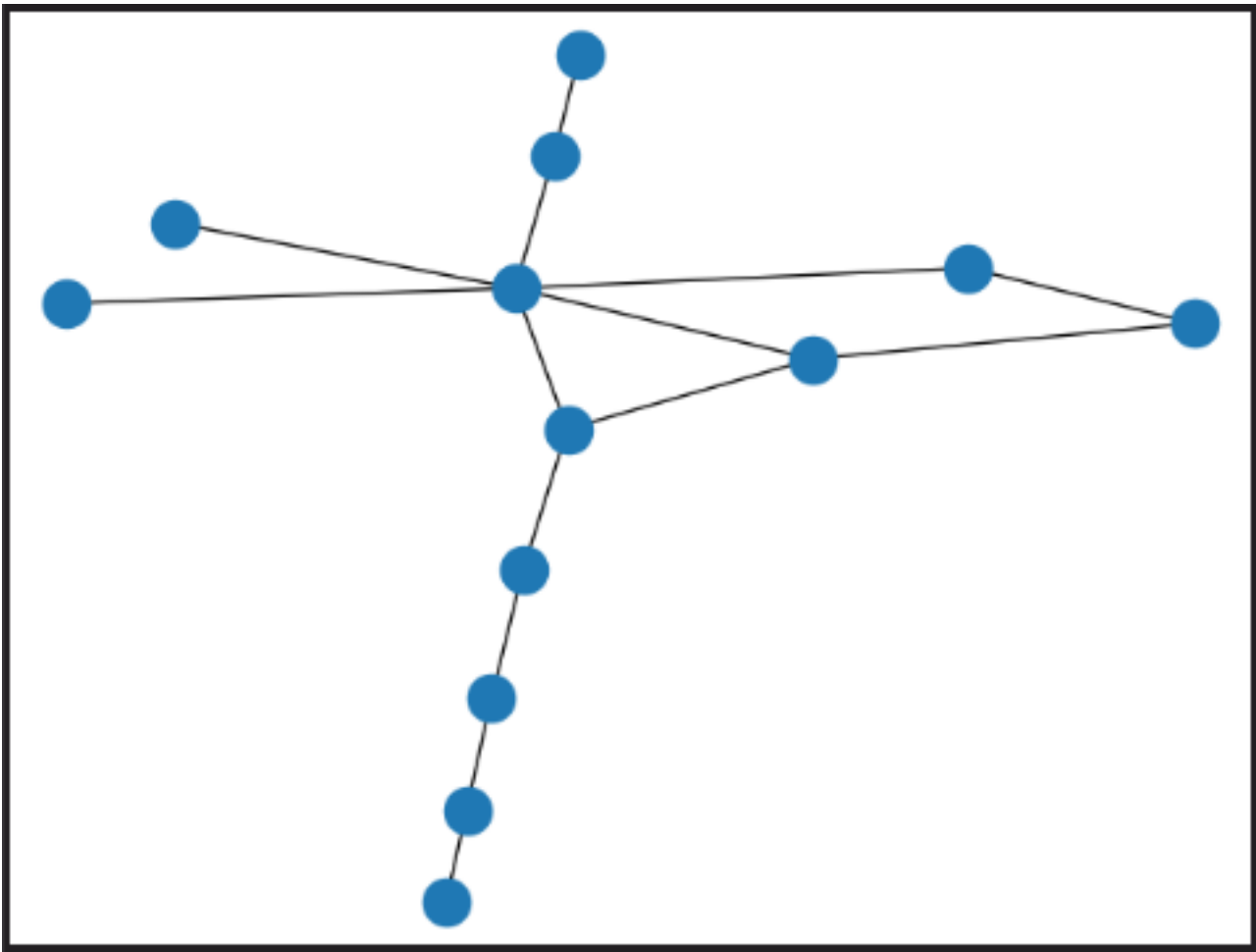


Рис 3.2.1

Наведений нижче код створений для того, щоб генерувати рекомендації для конкретного користувача:

```

def get_user_friends_recommendations(self, id):
    recomm = []
    queue = []
    visited = []
    queue.append(id)
    visited.append(id)
    visited.extend(list(self.G.neighbors(id)))
    while len(recomm) < 5:
        if queue == []:
            break
        curr = queue.pop(0)
        if curr in visited:
            neigh = list(self.G.neighbors(id))
            for i in neigh:
                if not i in visited:
                    queue.append(i)
        else:
            visited.append(curr)
            recomm.append(curr)

    if len(recomm) < 5:
        while len(self.G.nodes) != len(visited) and len(recomm) != len(visited) and len(recomm) < 5:
            nodes = self.G.nodes
            uniq_nodes = [i for i in nodes if not i in visited]
            ch = random.choice(uniq_nodes)
            recomm.append(ch)
            visited.append(ch)

    recommend_users = []
    print(recomm)
    for i in recomm:
        recommend_users.append(self.find_user(i))
    return recommend_users

```

Рис 3.2.2

Робота знаходиться за посиланням: <https://github.com/kuzneliOf/SocialMedia>

3.3 Результат

Результатом виконаної роботи став соціальний граф, вершинами якого є користувачі, а ребрами стали зв'язки між ними. В базі даних моєї роботи зареєстровано 13 тестових користувачів та їх зв'язки між собою. Завдяки зручному інтерфейсу можна легко зареєструватись та перейти на сторінку пошуку друзів.

Після реєстрації нового користувача він додається до бази даних з користувачами, а також до графу. Проблема алгоритму колаборативної фільтрації в тому, що для користувачів, які тільки зареєструвались, дуже важко знайти рекомендації друзів. Великі компанії вирішують цю проблему за

допомогою, наприклад, імпортуванням контактів, знаходженням друзів поблизу (тобто за допомогою залежності від локації) та іншими способами. В моїй роботі я зберігаю в базі даних лише пароль та логін користувача, тому важко знайти хорошу рекомендацію друзів для нового користувача. Однак рекомендації необхідні, тому вони показуються випадково, але лише до тих пір, поки користувач не додасть одного друга. Як тільки це зробити, запрацює рекомендаційна система, яка показує друзів друзів, використовуючи алгоритми теорії графів.

4 Висновок

Під час написання даної роботи було розглянуто 2 теми:

- Соціальний граф та його властивості
- Рекомендаційна система на основі колаборативної фільтрації

Обидві теми були коротко описані для того, щоб створити та проаналізувати створену соціальну мережу для створення в подальшому рекомендаційної системи.

Практична ж частина роботи представляє собою створення демонстративного сайту з реалізованим алгоритмом для перевірки того, як працюють методи колаборативної фільтрації та які в них є проблеми.

В заключення, алгоритми колаборативної фільтрації є ефективним рішенням для пошуку рекомендацій в мережах, однак не кожен алгоритм може бути застосований для рекомендаційної системи в мережі, де є величезна кількість користувачів, тому що деякі алгоритми можуть занадто довго працювати.

5 Джерела

<https://builtin.com/data-science/recommender-systems>

<https://towardsdatascience.com/recommendation-systems-explained-a42fc60591ed>

<https://www.int-res.com/abstracts/cr/v30/n1/p79-82/>

<https://medium.com/nuances-of-programming/анализ-социальных-сетей-от-теории-графов-до-приложений-на-python-93220a17c3d3>

https://ela.kpi.ua/bitstream/123456789/26898/1/Prasolov_magistr.pdf

<https://uk.wikipedia.org/wiki/Наука_{ii}>

<https://mathworld.wolfram.com/GraphDiameter.html>