

Домашнее задание 1

Глубинное обучение в анализе графовых данных.

Владимир Лузин

21 октября 2024 г.

1 Задание 1. Анализ графа.

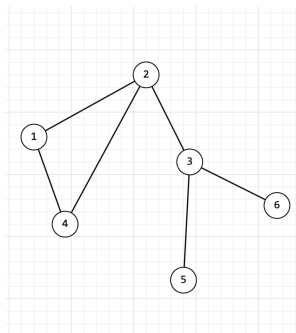


Рис. 1: Граф 1

1. Матрица смежности

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

2. Степени всех вершин

вершина	степень
1	2
2	3
3	3
4	2
5	1
6	1

Таблица 1: Степени вершин

3. Меры центральности и коэффициент кластеризации для вершин 1, 3, 4

- Мера близости

$$c_v = \frac{1}{\sum_{u \neq v} \text{длина кратчайшего пути между } u \text{ и } v}$$

$$c_1 = \frac{1}{1 + 2 + 1 + 3 + 3} = \frac{1}{10} = 0.1$$

$$c_3 = \frac{1}{2+1+2+2+1} = \frac{1}{7} \approx 0.14$$

$$c_4 = \frac{1}{1+1+2+3+3} = \frac{1}{10} = 0.1$$

Возможно в задании требовалось вычислить нормализованную форму, поэтому добавлю нормализацию

$$c_v = \frac{N-1}{\sum_{u!=v} \text{длина кратчайшего пути между } u \text{ и } v}$$

$$c_1 = \frac{5}{1+2+1+3+3} = \frac{5}{10} = 0.5$$

$$c_3 = \frac{5}{2+1+2+2+1} = \frac{5}{7} \approx 0.71$$

$$c_4 = \frac{5}{1+1+2+3+3} = \frac{5}{10} = 0.5$$

- Мера соседства

$$c_v = \sum_{s!=v!=t} \frac{\text{количество кратчайших путей между } s \text{ и } t \text{ содержащих } v}{\text{Количество кратчайших путей между } s \text{ и } t}$$

$$c_1 = \frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} = \frac{0}{4} = 0$$

$$c_3 = \frac{2}{4} + \frac{2}{4} + \frac{2}{4} + \frac{4}{4} + \frac{4}{4} = \frac{14}{4} = 3.5$$

$$c_4 = \frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} = \frac{0}{4} = 0$$

Возможно в задании требовалось вычислить нормализованную форму, поэтому добавлю нормализацию

$$c_v = \frac{2}{(n-1)(n-2)} \sum_{s!=v!=t} \frac{\text{количество кратчайших путей между } s \text{ и } t \text{ содержащих } v}{\text{Количество кратчайших путей между } s \text{ и } t}$$

$$c_1 = \frac{2}{5*4} \left(\frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} \right) = \frac{2}{20} * \frac{0}{4} = \frac{0}{80} = 0$$

$$c_3 = \frac{2}{5*4} \left(\frac{2}{4} + \frac{2}{4} + \frac{2}{4} + \frac{4}{4} + \frac{4}{4} \right) = \frac{2}{20} * \frac{14}{4} = \frac{38}{80} = 0.475$$

$$c_4 = \frac{2}{5*4} \left(\frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} + \frac{0}{4} \right) = \frac{2}{20} * \frac{0}{4} = \frac{0}{80} = 0$$

- Коэффициент кластеризации

$$c_u = \frac{|(v_1, v_2) \in E : v_1, v_2 \in N(u)|}{\binom{d_u}{2}}$$

$$c_1 = \frac{1}{\frac{2!}{2!(2-2)!}} = \frac{1}{\frac{2}{2}} = 1$$

$$c_3 = \frac{0}{\frac{3!}{2!(3-2)!}} = 0$$

$$c_4 = \frac{1}{\frac{2!}{2!(2-2)!}} = \frac{1}{\frac{2}{2}} = 1$$

4. статистики локального пересечения

- по общим соседям

$$S[u, v] = |\mathcal{N}(u) \cap \mathcal{N}(v)|$$

$$S[1, 3] = |(2, 4) \cap (2, 5, 6)| = |(2)| = 1$$

$$S[5, 6] = |(3) \cap (3)| = |(3)| = 1$$

$$S[1, 4] = |(2, 4) \cap (1, 2)| = |(2)| = 1$$

- коэффициент Жаккара

$$S_{\text{Jaccard}}[u, v] = \frac{|\mathcal{N}(u) \cap \mathcal{N}(v)|}{|\mathcal{N}(u) \cup \mathcal{N}(v)|}$$

$$S_{\text{Jaccard}}[1, 3] = \frac{|(2, 4) \cap (2, 5, 6)|}{|(2, 4) \cup (2, 5, 6)|} = \frac{|(2)|}{|(2, 4, 5, 6)|} = \frac{1}{4} = 0.25$$

$$S_{\text{Jaccard}}[5, 6] = \frac{|(3) \cap (3)|}{|(3) \cup (3)|} = \frac{1}{1} = 1$$

$$S_{\text{Jaccard}}[1, 4] = \frac{|(2, 4) \cap (1, 2)|}{|(2, 4) \cup (1, 2)|} = \frac{|(2)|}{|(1, 2, 4)|} = \frac{1}{3} = 0.(3)$$

- индекс Адамика-Адара

Буду использовать логарифм по основанию 10

$$S_{\text{AA}}[v_1, v_2] = \sum_{u \in \mathcal{N}(v_1) \cap \mathcal{N}(v_2)} \frac{1}{\log(d_u)}$$

$$S_{\text{AA}}[1, 3] = \sum_{u \in (2, 4) \cap (2, 5, 6)} \frac{1}{\log(d_u)} = \sum_{u \in (2)} \frac{1}{\log(d_u)} = \frac{1}{\log(3)} \approx \frac{1}{0,477} \approx 2,096$$

$$S_{\text{AA}}[5, 6] = \sum_{u \in (3) \cap (3)} \frac{1}{\log(d_u)} = \sum_{u \in (3)} \frac{1}{\log(d_u)} = \frac{1}{\log(3)} \approx \frac{1}{0,477} \approx 2,096$$

$$S_{\text{AA}}[1, 4] = \sum_{u \in (2, 4) \cap (1, 2)} \frac{1}{\log(d_u)} = \sum_{u \in (2)} \frac{1}{\log(d_u)} = \frac{1}{\log(3)} \approx \frac{1}{0,477} \approx 2,096$$

5. GDV

Буду считать вектор графлетов строго размера 3 для вершин 2, 4 и 5. Для этого рассчитаю для каждой вершины вектор размера 3, где первым элементов будет "развернутый" графлет из 3 вершин с боковым корнем ("развёрнутым" графлетом я назову полный граф размера 3 без одного ребра), вторым элементов будет "развёрнутый" графлет с центральным корнем и третьим элементов будет полный графлет размера 3 (треугольник). Все графлеты я буду обозначать через "-" и разделять одинаковые графлеты с помощью "|".

$$\vec{c}_2 = \langle 2-3-5 | 2-3-6, 1-2-3 | 4-2-3, 2-4-1-2 \rangle = \langle 2, 2, 1 \rangle$$

$$\vec{c}_4 = \langle 4-2-3, 0, 4-1-2-4 \rangle = \langle 1, 0, 1 \rangle$$

$$\vec{c}_5 = \langle 5-3-6 | 5-3-2, 0, 0 \rangle = \langle 1, 0, 0 \rangle$$

2 Задание 2. Подсчет ядер.

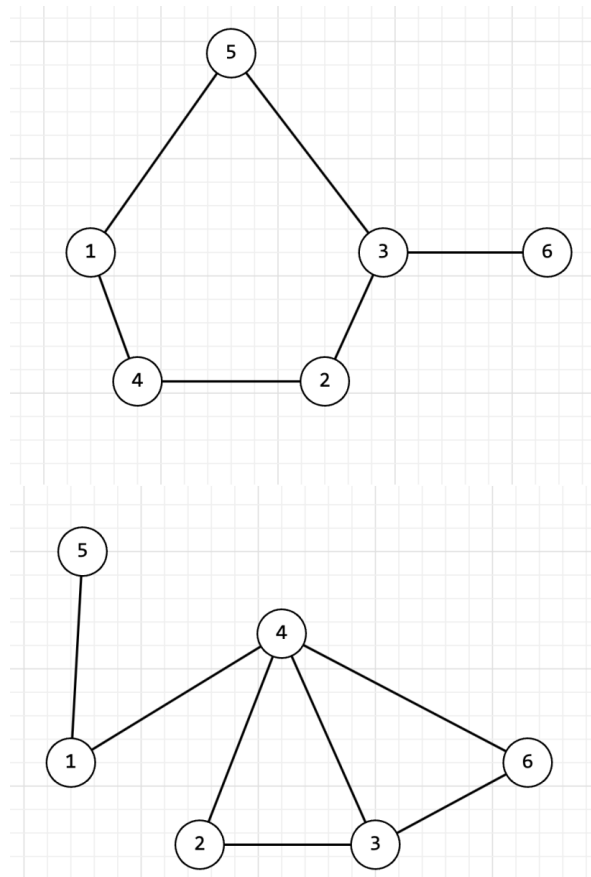


Рис. 2: Граф 2-3

2.1 графлетовое ядро

Буду считать графлетовое ядро размера 3 вручную, без программ. Для начала найду количество графлетов размера 3 в каждом графе (вершины графлета могут быть не соединены)

$$C_6^3 = \frac{6!}{3!(6-3)!} = 20$$

Значит вектор мешка вершин будет размера 20 для обоих графов. Найду этот вектор с помощью простого перебора всех вариантов вершин. Набор вершин сначала буду записывать в таблицу для удобства проверки.

- первое значение - Это треугольный, полносвязный, графлет
- второе значение - Это треугольный, полносвязный, графлет без одного ребра
- третье значение - это треугольный, полносвязный, графлет без двух ребер
- четвёртое значение - это три несвязанных вершины

мешок вершин - Граф 2

1-й	2-й	3-й	4-й
-	1-2-4	1-2-3	1-2-6
-	1-3-5	1-2-5	2-5-6
-	1-4-5	1-3-4	4-5-6
-	2-3-4	1-3-6	-
-	2-3-5	1-4-6	-
-	2-3-6	1-5-6	-
-	3-5-6	2-4-5	-
-	-	2-4-6	-
-	-	3-4-5	-
-	-	3-4-6	-
0	7	10	3

Таблица 2: Мешок вершин - Граф 2

$$f_2 = \langle 0, 7, 10, 3 \rangle$$

мешок вершин - Граф 3

1-й	2-й	3-й	4-й
2-3-4	1-2-4	1-2-3	1-2-6
3-4-6	1-3-4	1-2-5	2-5-6
-	1-4-5	1-3-5	-
-	1-4-6	1-3-6	-
-	2-3-6	1-5-6	-
-	2-4-6	2-3-5	-
-	-	2-4-5	-
-	-	3-4-5	-
-	-	3-5-6	-
-	-	4-5-6	-
2	6	10	2

Таблица 3: Мешок вершин - Граф 3

$$f_3 = \langle 2, 6, 10, 2 \rangle$$

Графлетовое ядро

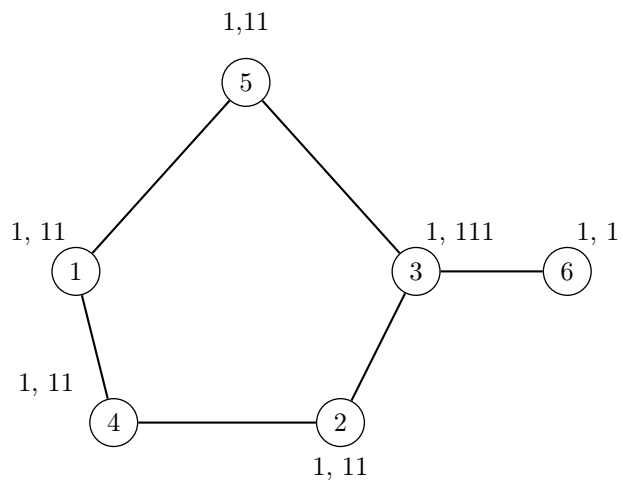
$$k(2, 3) = f_2^T f_3 = 2 * 0 + 6 * 7 + 10 * 10 + 3 * 2 = 42 + 100 + 6 = 148$$

Графы одинакового размера, так что нормализация не требуется

2.2 WL_3 ядро

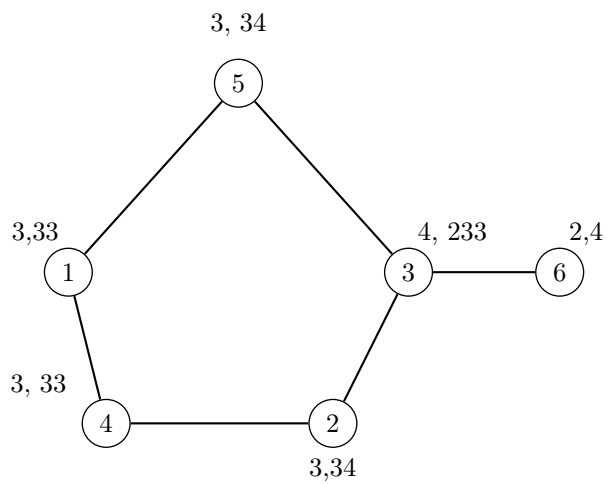
Найду ядро для каждого размера, ограничусь 3 итерациями вручную. Для каждой итерации буду рисовать граф с агрегированными цветами и хеш-таблицу.

Граф 2



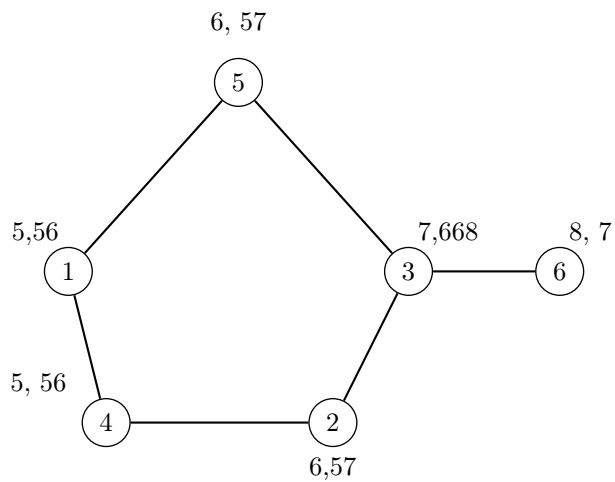
агрегация	хешы	количество
1, 1	2	1
1, 11	3	4
1, 111	4	1

Таблица 4: хеш таблица - граф 2 - итерация 1



агрегация	хешы	количество
2,4	8	1
3,33	5	2
3,34	6	2
4,233	7	1

Таблица 5: хеш таблица - граф 2 - итерация 2

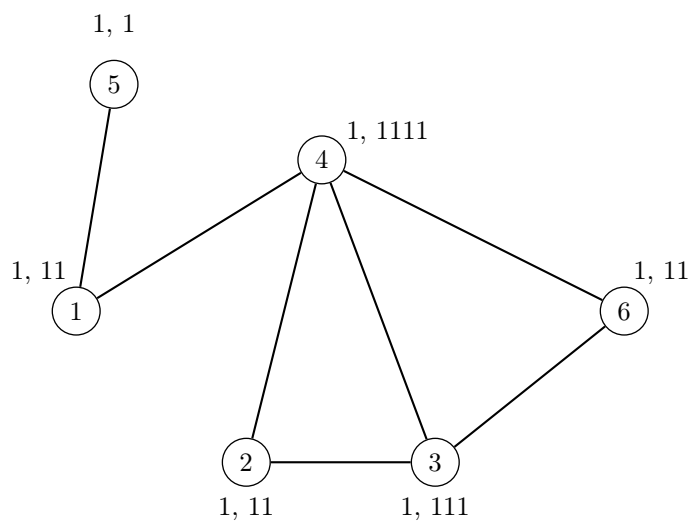


агрегация	хешы	количество
5,56	9	2
6,57	10	2
7,668	11	1
8, 7	12	1

Таблица 6: хеш таблица - граф 2 - итерация 3

Итоговый вектор второго графа

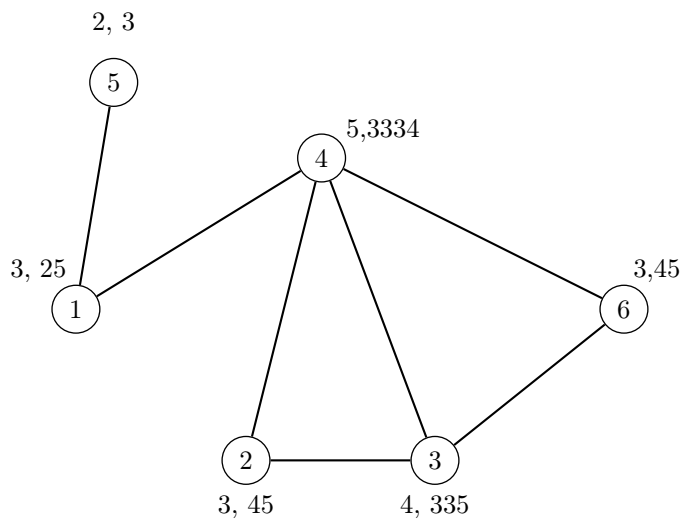
$$f_2 = \langle 6, 1, 4, 1, 1, 2, 2, 1, 2, 2, 1, 1 \rangle$$



Граф 3

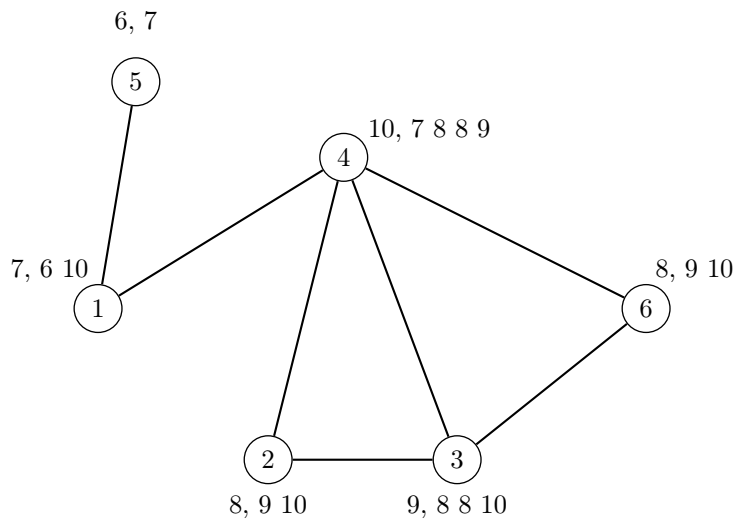
агрегация	хешы	количество
1, 1	2	1
1, 11	3	3
1, 111	4	1
1, 1111	5	1

Таблица 7: хеш таблица - граф 3 - итерация 1



агрегация	хеши	количество
2,3	6	1
3,25	7	1
3,45	8	2
4,335	9	1
5,3334	10	1

Таблица 8: хеш таблица - граф 3 - итерация 2



агрегация	хеши	количество
6,7	11	1
7,6 10	12	1
8,9 10	13	2
9, 8 8 10	14	1
10,7 8 8 9	15	1

Таблица 9: хеш таблица - граф 3 - итерация 3

Итоговый вектор третьего графа

$$f_3 = \langle 6, 1, 3, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1 \rangle$$

3 Задание 3. PageRank. Другая запись

$$r = Gr$$

$$G = \beta M + (1 - \beta) \left[\frac{1}{N} \right]_{N \times N}$$

Подставлю G из второй формулы в первую и распишу r как вектор-столбец, вынесу 1/N из матри

$$\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} = \left(\beta M + \frac{(1 - \beta)}{N} [1]_{N \times N} \right) \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

Раскрою скобки и умножу матрицу на вектор-столбец

$$\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} = \beta M r + \frac{(1 - \beta)}{N} \begin{bmatrix} r_1 + \dots + r_n \\ \vdots \\ r_1 + \dots + r_n \end{bmatrix}$$

Сумма $r_1 \dots r_n$ равна 1, поскольку r - вектор вероятностей, сумма которых равна единице

$$\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} = \beta M r + \frac{1 - \beta}{N} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$r = \beta M r + \frac{1 - \beta}{N} 1_N$$

Что и требовалось доказать

4 Задание 4. PageRank. Алгоритм

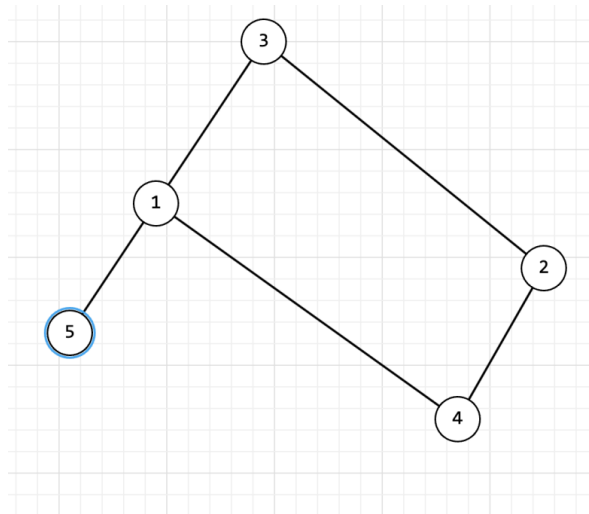


Рис. 3: Граф 4

Буду считать все рёбра графа двунаправленными

4.1 Стохастическая матрица смежности

Матрица смежности

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Стохастическая матрица смежности (Нормирую по строкам)

$$\begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \end{bmatrix}$$

Транспонирую матрицу

$$\begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix}$$

4.2 Матрица перехода

$$r^{(0)} = \langle 1, 0, 0, 0, 0 \rangle$$

$$M = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix}$$

Матрица перехода к следующему состоянию

$$r^{(1)} = Mr^{(0)} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

Матрица перехода на два шага

$$r^{(2)} = Mr^{(1)} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

4.3 PageRank score

Для подсчёта начну с равномерного распределения, а не с одной единицей в начале. Не буду использовать модификацию с телепортами для выхода из циклов, поскольку в данном случае граф двунаправленный и сильно связный.

$$r^{(0)} = \langle 0.2, 0.2, 0.2, 0.2, 0.2 \rangle$$

$$r^{(1)} = Mr^{(0)} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.2 \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{15} \end{bmatrix}$$

Повторяю пока $|r^{(k)} - r^{(k-1)}| < 0.01$ Для этого напишу небольшую программу на python и numpy (Программу прикреплю вместе с решением) В результате получился вектор
 $r^{(51)} = \langle 0.024572, 0.02006296, 0.01487832, 0.01487832, 0.00668765 \rangle$

4.4 альтернативный PageRank

Для решения задачи нужно по-другому составить матрицу смежности. Для этого сначала в матрице смежности я просчитаю все возможные переходы, включая добавленный двойной шаг и отсутствии движений. Важно учесть, что при двойном шаге можно остаться в той же точке.

Матрица смежности

$$\begin{bmatrix} 4 & 2 & 1 & 1 & 1 \\ 2 & 3 & 1 & 1 & 0 \\ 1 & 1 & 3 & 2 & 1 \\ 1 & 1 & 2 & 3 & 1 \\ 1 & 0 & 1 & 1 & 2 \end{bmatrix}$$

Стохастическая матрица смежности (Нормирую по строкам)

$$\begin{bmatrix} \frac{4}{9} & \frac{2}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{2}{5} & \frac{3}{5} & \frac{1}{5} & \frac{1}{5} & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{8} & \frac{2}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{2}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix}$$

Транспонирую матрицу

$$\begin{bmatrix} \frac{4}{9} & \frac{2}{5} & \frac{1}{8} & \frac{1}{8} & \frac{1}{5} \\ \frac{2}{9} & \frac{3}{5} & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{9} & \frac{1}{5} & \frac{3}{8} & \frac{2}{8} & \frac{1}{5} \\ \frac{1}{9} & \frac{1}{5} & \frac{2}{8} & \frac{3}{8} & \frac{1}{5} \\ \frac{1}{9} & 0 & \frac{1}{8} & \frac{1}{8} & \frac{2}{5} \end{bmatrix}$$

Аналогично предыдущей задаче, начну с равномерного распределения, а не с одной единицы в начале. Не буду использовать модификацию с телепортами для выхода из циклов, поскольку в данном случае граф двунаправленный и сильно связный.

$$r^{(0)} = \langle 0.2, 0.2, 0.2, 0.2, 0.2 \rangle$$

$$r^{(1)} = Mr^{(0)} = \begin{bmatrix} \frac{4}{9} & \frac{2}{5} & \frac{1}{8} & \frac{1}{8} & \frac{1}{5} \\ \frac{2}{9} & \frac{3}{5} & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{9} & \frac{1}{5} & \frac{3}{8} & \frac{2}{8} & \frac{1}{5} \\ \frac{1}{9} & \frac{1}{5} & \frac{2}{8} & \frac{3}{8} & \frac{1}{5} \\ \frac{1}{9} & 0 & \frac{1}{8} & \frac{1}{8} & \frac{2}{5} \end{bmatrix} * \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.236 \\ 0.18 \\ 0.216 \\ 0.216 \\ 0.152 \end{bmatrix}$$

Повторяю пока $|r^{(k)} - r^{(k-1)}| < 0.01$ Для этого использую тот же самый алгоритм с изменённым (Программу прикреплю вместе с решением)

В результате получился вектор, который сошёлся уже на 3-м шаге

$$r^{(3)} = \langle 0.24200117, 0.18651428, 0.21699379, 0.21699379, 0.13749696 \rangle$$

5 Задание 5. Дополнительное задание. Индекс Каца

$$S_{Katz}[u, v] = \sum_{i=1}^{\infty} \beta^i A^i[u, v] = \sum_{i=1}^{\infty} (\beta A[u, v])^i$$

Используя формулу $(I - X)^{-1} = \sum_{i=1}^{\infty} X^i$

$$S_{Katz}[u, v] = (I - \beta A[u, v])^{-1}$$

Тогда и только тогда, если $\lambda_1 \leq 1$ и $(I - X)$ - невырожденная