

ОПРЕДЕЛЕНИЕ ЦЕНТРАЛЬНОСТИ ГРАФА АЛГОРИТМОМ PAGERANK С УЧЕТОМ ВЕСОВ СВЯЗЕЙ

Егоркин А. А.¹

(ФГБОУ ВО Российский государственный социальный
университет, Москва)

Работа посвящена нахождению центральности узлов взвешенных графов с учетом веса связей. Актуальность этой задачи обусловлена тем, что игнорирование весов дуг графа при нахождении центральности его узлов недопустимо для ряда прикладных задач, в первую очередь относящихся к задачам из финансовой сферы. В классической постановке алгоритма PageRank происходит потеря части информации о весах связей при формировании матрицы переходных вероятностей из матрицы смежности. Данный эффект был продемонстрирован в настоящей статье. Предложен метод определения центральности узлов сети, базирующийся на алгоритме PageRank, который позволяет учесть веса всех связей. В качестве примера рассматривался граф финансовых транзакций. Узлами графа являются клиенты коммерческого банка в том числе сам банк, а дугами – денежные переводы между узлами. Качество ранжирования определялось путем сравнения различных мер центральности с внешним параметром, который характеризует важность узла и не связан с сетевыми характеристиками графа. По результатам исследования было показано, что предлагаемая мера центральности лучшим образом ранжирует наиболее важные узлы графа по сравнению с иными мерами центральности. Также была продемонстрирована сходимость предлагаемого алгоритма.

Ключевые слова: центральность PageRank, цепи Маркова, взвешенный направленный граф.

1. Введение

На данный момент разработано большое количество мер центральности, позволяющих оценить значимость или важность узлов в графе на основании различных критериев [5]. К классическим мерам центральности традиционно относят центральность по близости, центральность по посредничеству,

¹ Антон Александрович Егоркин, аспирант (2-5@bk.ru).

центральность по степени, центральность по Кацу [13] и PageRank [15].

При этом перечисленные метрики центральности, как правило, используют одну из характеристик узла графа, например, количество соседей, кратчайших путей, проходящих через узел или расстояние до других узлов графа.

Так, в алгоритме PageRank, базирующемся на учете степени влияния соседей, не учитывается вес связи либо учитывается локально, так как при построении матрицы переходных вероятностей указанная информация в почти не используется. В настоящей статье рассматриваются прикладные задачи, для которых необходимо учитывать не только влияние соседей, но и веса связей, а также подходы к решению подобных задач.

1.1. БАЗОВЫЕ ПРИНЦИПЫ ОПРЕДЕЛЕНИЯ ЦЕНТРАЛЬНОСТИ НА ОСНОВЕ ЦЕПЕЙ МАРКОВА

Андрей Андреевич Марков был основоположником исследования случайных процессов, где состояние в будущем зависит только от текущего состояния и названных впоследствии его именем (марковский процесс или марковская цепь). Впервые данный процесс был сформулирован Марковым А.А. в 1907 г.

Данный процесс может быть сформулирован следующим образом. Рассмотрим сеть или граф (размера n на n), где матрицей $\mathbf{P} = (p_{ij})$ описываются вероятности перехода из одного узла-состояния в другой узел ($\sum_{i=1}^n p_{ij} = 1$ и $p_{ij} \geq 0$). Тогда для марковского процесса будет характерно:

- (1) $\mathbf{x}(k+1) = \mathbf{x}(k) \times \mathbf{P}$, $k \rightarrow \infty$,
- (2) $\mathbf{x} = \mathbf{x} \times \mathbf{P}$,

где $\mathbf{x}$ – вектор вероятностей распределения элементов по системе.

Существует предел с положительными элементами:

- (3) $\lim_{k \rightarrow \infty} \mathbf{P}^k$.

При достаточно большом k распределение элементов в сети не будет зависеть от начального вектора $\mathbf{x}$, а будет описываться лишь сходимостью произведений матрицы вероятностей переходов $\mathbf{P}$.

Из уравнения (2) следует, что задача сводится к нахождению левого собственного вектора при собственном значении, равном 1, которое, исходя из теоремы Перрона – Фробениуса, будет положительным.

Для того чтобы выполнялись условия формул (2)–(3), необходимо, чтобы матрица **P** была стохастической, неприводимой и апериодической, т.е. регулярной [1]. Данные требования являются следствием теоремы Перрона – Фробениуса. В литературе подобные матрицы иногда называются SIA (Stochastic Indecomposable Aperiodic) [6, 11].

Однако для реальных сетей условия неразложимости и апериодичности не всегда выполняются, особенно для интернет-сетей с «висячими узлами» – страницами без исходящих ссылок. С указанных страниц пользователь не может перейти на другие страницы в рамках заданной сети ($p_{ij} = 0 \forall j$). В таких случаях предполагается, что пользователь может случайным образом перейти на любую другую страницу, т.е. строка матрицы переходов заполняется значениями $1/n$ [14].

1.2. АЛГОРИТМ PAGERANK

Суть алгоритма заключается в том, что блуждание осуществляется не только по ссылкам самой сети, пользователь случайным образом может перемещаться в пространстве сети. Параметр $\alpha \in [0 \dots 1]$ – фактор телепортации – отвечает за случайное перемещение.

$$(4) \quad \mathbf{x}(t+1) = \mathbf{x}(t)[\alpha \cdot \mathbf{P} + (1 - \alpha) \cdot \mathbf{J}] .$$

Чем больше компонента $(1 - \alpha) \cdot \mathbf{J}$, отвечающая за случайное блуждание по графу (где **J** – квадратная матрица размера $n \times n$, состоящая из единиц), тем быстрее сходится процесс (4), но за счет потери информации о структуре исходного графа. И наоборот, чем меньше параметр α , тем медленнее сходится алгоритм (4), поскольку блуждание осуществляется по исходному графу. Оптимальный баланс обеспечивается при $\alpha = 0,85$ [10].

Существуют модификации алгоритма PageRank, в том числе персонализированный PageRank или PageRank с перезапуском [7, 17], которые предлагают различные способы «обработки» висячих узлов.

2. Потеря информации при формировании стохастической матрицы

Рассмотрим три матрицы смежности:

$$(5) \quad A_1 = \begin{bmatrix} 0 & 100 & 100 & 200 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix},$$

$$(6) \quad A_2 = \begin{bmatrix} 0 & 10 & 10 & 20 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix},$$

$$(7) \quad A_3 = \begin{bmatrix} 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Для преобразования матрицы смежности A в матрицу переходных вероятностей P можно использовать следующие подходы:

При наличии связи между узлами заполнять их значением $1/\sum_{i=1}^n \delta_{ij}$, где δ_{ij} – символ Кронекера:

$$(8) \quad P_1 = \begin{bmatrix} 0 & 0,33 & 0,33 & 0,33 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0,5 & 0,5 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0,33 & 0,33 & 0,33 & 0 \end{bmatrix}.$$

При наличии связи между узлами заполнять их значением $1/\sum_{i=1}^n a_{ij}$:

$$(9) \quad P_2 = \begin{bmatrix} 0 & 0,25 & 0,25 & 0,5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0,5 & 0,5 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0,33 & 0,33 & 0,33 & 0 \end{bmatrix}.$$

Матрица переходных вероятностей P_1 может быть получена из любой матрицы смежности A_1, A_2, A_3 при заполнении элементов матрицы P_1 значениями $1/\sum_{i=1}^n \delta_{ij}$.

Аналогично и для матрицы переходных вероятностей P_2 , которая может быть получена из любой матрицы смежности A_1 , A_2 , A_3 при заполнении элементов матрицы P_2 значениями $1/\sum_{i=1}^n a_{ij}$.

Таким образом, происходит потеря важной информации об исходной матрице смежности. Для задач анализа интернет-сетей важен в первую очередь сам факт наличия ссылки и ее направление, но для многих других задач также важен и вес связи или дуги в терминах теории графов.

Существуют модификации PageRank для матриц вероятностей типа P_2 , где взвешенность учитывается на уровне отдельных узлов. Однако из-за стохастического характера этих матриц «глобальное» распределение весов не может быть учтено в рамках всего графа.

В каких задачах важно учитывать вес связей между узлами? Например, в задачах оценки финансовых рисков [9, 18] в которых вес дуги может представлять собой сумму транзакций в денежном выражении между клиентами, являющимися узлами графа [4]. В данном случае это является критически важной информацией, поскольку платеж на один рубль или один миллиард рублей – это существенно разные денежные потоки. В описанных выше подходах данная информация будет потеряна при создании стохастической матрицы вероятностей перехода.

3. Предложение по учету весов связей в алгоритме PageRank

Для решения описанной выше проблемы предлагается сначала выполнить кластеризацию дуг графа на основе их весов, далее рассчитать показатели центральности для сформированных кластеров, а затем агрегировать полученные вектора центральностей.

Как показано в [3], веса дуг графа финансовых транзакций подчиняются степенному закону распределения [19, 21]. В работе [2] нами было продемонстрировано, что логарифмирование исходных данных, подчиняющихся степенному закону распределения [8, 16], способствует более равномерной кластеризации данных. При этом результат не зависит от выбора основания

логарифма, и при использовании, например, алгоритма кластеризации *k-means* [12] кластеризация в одномерном случае осуществляется вокруг среднегеометрических значений соответствующего кластера.

Графическая интерпретация агрегирования векторов центральности представлена на рис. 1.

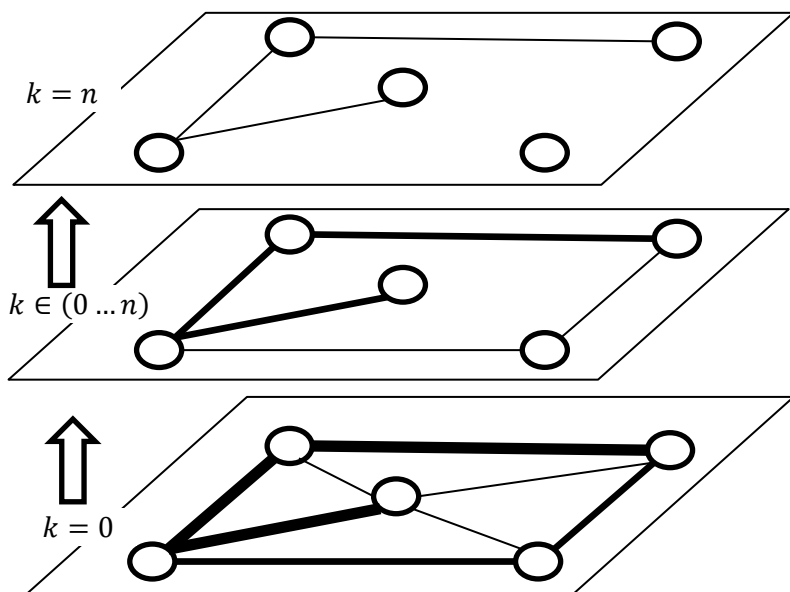


Рис. 1. Графическая интерпретация алгоритма

На шаге $k = 0$ определяется вектор центральности исходного графа s_0 . На следующих шагах $k \in [1, \dots, n - 1]$ определяется вектор центральности s_k графа, для которого выполняется условие, что веса дуг больше определенного порога $a_{ij} > b_k$. С каждым последующим шагом порог b_k увеличивается, а количество дуг уменьшается. Появляются висячие узлы, центральность которых минимальна и определяется исключительно случайным блужданием. Таким образом, для шага $k \in [1, \dots, n - 1]$ интерес представляет не абсолютная величина центральности, а ее отклонение от минимального значения.

На каждом из шагов центральность определяется для взвешенного графа. Принимается, что веса всех дуг в одном кластере одинаковы и равны среднегеометрическому значению $\overline{a_k}$.

Исходя из описанных выше предпосылок, предлагаемый алгоритм можно описать следующим образом:

$$(10) \mathbf{c} = \mathbf{c}_0 + \sum_{k=1}^{n-1} [\mathbf{c}_k - \min(\mathbf{c}_k)] \frac{\overline{a_k} - \overline{a_{k-1}}}{\overline{a_0}}$$

где $\mathbf{A} = \|a_{ij}\|$ – матрица смежности; $\mathbf{c}_k$ – вектор центральности PageRank матрицы смежности, в которой веса дуг больше порога b_k : $a_{ij} > b_k$; $\mathbf{c}_0$ – вектор центральности PageRank исходной матрицы смежности; $\overline{a_k}$ – среднегеометрическое значение весов дуг матрицы смежности с весами более порога b_k ; $\overline{a_{k-1}}$ – среднегеометрическое значение весов дуг матрицы смежности с весами более порога b_{k-1} , при этом $b_k > b_{k-1}$; $\overline{a_0}$ – среднегеометрическое значение весов дуг исходной матрицы смежности; n – количество шагов алгоритма.

Порядок определения порогов b_k в простейшем случае осуществляется следующим образом:

$$(11) b_k = 10^{d_k},$$

$$\text{где } d_k = \lg \min(a_{ij} > 0) + k \cdot \frac{\lg \max(a_{ij}) - \lg \min(a_{ij} > 0)}{n}$$

Итоговое значение вектора $\mathbf{c}$ нормируется на сумму значений вектора $\mathbf{c}$.

4. Пример использования алгоритма

Рассмотрим матрицу смежности

$$\mathbf{A} = \begin{bmatrix} 0 & 10 & 20 & 200 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 20 & 1 \\ 0 & 0 & 80 & 0 & 1 \\ 0 & 10 & 1 & 1 & 0 \end{bmatrix}.$$

Используя в алгоритме $n = 3$, имеем следующие промежуточные вычисления (таблица 1).

Таблица 1. Промежуточные расчеты

k	Матрица смежности	Вид графа	b_k	$\bar{a}_k$	c_k
0	$\begin{bmatrix} 0 & 10 & 20 & 200 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 20 & 1 \\ 0 & 0 & 80 & 0 & 1 \\ 0 & 10 & 1 & 1 & 0 \end{bmatrix}$		-	6,3	$\begin{bmatrix} 0,03 \\ 0,14 \\ 0,24 \\ 0,24 \\ 0,35 \end{bmatrix}$
1	$\begin{bmatrix} 0 & 10 & 20 & 200 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 20 & 0 \\ 0 & 0 & 80 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \end{bmatrix}$		5,8	29,4	$\begin{bmatrix} 0,05 \\ 0,08 \\ 0,41 \\ 0,42 \\ 0,04 \end{bmatrix}$
2	$\begin{bmatrix} 0 & 0 & 0 & 200 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 80 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$		34,2	126,5	$\begin{bmatrix} 0,13 \\ 0,13 \\ 0,35 \\ 0,25 \\ 0,13 \end{bmatrix}$

Итоговое значение центральности получается путем агрегирования векторов c_k в соответствии с формулой (10):

$$c = \begin{bmatrix} 0,03 \\ 0,14 \\ 0,24 \\ 0,24 \\ 0,35 \end{bmatrix} + \begin{bmatrix} 0,04 - 0,04 \\ 0,08 - 0,04 \\ 0,41 - 0,04 \\ 0,42 - 0,04 \\ 0,04 - 0,04 \end{bmatrix} \frac{29,4 - 6,3}{6,3} + \begin{bmatrix} 0,13 - 0,13 \\ 0,13 - 0,13 \\ 0,35 - 0,13 \\ 0,25 - 0,13 \\ 0,13 - 0,13 \end{bmatrix} \frac{126,5 - 29,4}{6,3}.$$

После нормирования итоговый вектор центральности имеет вид

$$c = \begin{bmatrix} 0,003 \\ 0,03 \\ 0,55 \\ 0,39 \\ 0,02 \end{bmatrix}.$$

В стандартном алгоритме PageRank (вектор c_k на шаге 0) наиболее важным узлом является узел №5. В предложенной мере центральности наиболее важными узлами являются узлы №3 и №4, так как в них заходят дуги с наибольшим весом.

5. Результаты применения алгоритма на реальных данных

В качестве примера применения алгоритма рассмотрим граф финансовых транзакций коммерческого банка. Узлами графа являются клиенты банка, дугами – денежные переводы между узлами за период с 2022 по 2023 год. Вес дуги – сумма транзакций в денежных единицах [21].

Рассмотрим следующие меры центральности для направленного графа [20]:

- центральность по близости (Closeness);
- центральность по посредничеству (Betweenness);
- центральность по степени (Degree);
- PageRank;
- предлагаемая мера (Financial PageRank).

Для того чтобы оценить качество ранжирования узлов рассматриваемыми мерами, сравним полученные значения центральностей с внешними данными, которые характеризуют степень важности узла. Узлами в графе являются экономические агенты, целесообразно использовать в качестве внешних данных экономические показатели, например такие, как размер собственных средств, валюта баланса, объем выручки и т.п. Так как узлы в графе в большей своей части являются пользователями услуг банка, то в качестве экзогенных данных будем использовать объем бизнеса, который клиент имеет с банком.

Объем бизнеса (*Volume*) определяется как средние дневные остатки по следующим банковским продуктам: гарантии, выданные кредиты, привлеченные депозиты и средства на текущих

счетах. Поскольку не все узлы графа имеют перечисленные выше продукты, то в анализе используем только тех клиентов у кого объем $Volume > 0$.

Рассмотрим граф, описанный выше. Граф имеет следующие характеристики:

- количество узлов – 4,5 тыс. шт.;
- количество дуг – 6,5 тыс. шт.

Определим для узлов данного графа все меры центральности, перечисленные выше. Выберем из графа узлы, которые имеют показатель $Volume > 0$, и отсортируем их по убыванию. Количество узлов, у которых $Volume > 0$, равно 2,1 тыс. шт.

Ценностью меры центральности является способность ранжирования узлов. Для оценки эффективности ранжирования, разобьем узлы графа на корзины по показателю $Volume$ нарастающим итогом. Далее оценим, какое количество верно отранжированных узлов попадет в соответствующую корзину для каждой меры центральности.

Показатель $Volume$ имеет неравномерное распределение. На первые 2,5% узлов, упорядоченных по данному показателю, приходится 63% всего суммарного показателя $Volume$, на первые 10% узлов – 82%. Исходя из этого, основное требование к мере центральности в данном примере – это корректное ранжирование наиболее важных узлов – узлов с высокими показателями $Volume$ (таблица 2).

Таблица 2. Сравнение мер центральности

№	Доля узлов, упорядоченных по показателю Volume	Доля Volume	Доля правильно определенных узлов				
			Betweenness	Closeness	Degree	Financial Pagerank	Pagerank
1	2.5%	63%	30%	30%	43%	58%	34%
2	5.0%	73%	39%	32%	50%	57%	40%
3	10%	82%	50%	40%	54%	57%	43%
4	15%	88%	55%	44%	63%	57%	49%
5	25%	94%	61%	51%	68%	64%	55%
6	50%	99%	68%	61%	72%	68%	65%
7	100%	100%	100%	100%	100%	100%	100%

В приведенных результатах видно, мера центральности *Financial Pagerank* лучше ранжирует узлы с высокими показателями *Volume*. Среди узлов с самыми крупными значениями объема бизнеса (строка №1 в таблице 2) предложенная мера корректно определяет 58% из первых 2,5% узлов графа, что является лучшим результатом среди рассматриваемых мер центральности.

Далее по мере увеличения количества узлов, упорядоченных по *Volume*, увеличиваются и доли корректно предсказываемых узлов по всем мерам центральности. Начиная с первых 15% узлов графа лучшими ранжирующими способностями обладает мера центральности по степени (*Degree*).

За счет того, что анализируемая мера учитывает весь спектр весов дуг графа, распределение центральностей по сравнению со стандартным алгоритмом *PageRank* становится более полярным, т.е. наиболее важные узлы имеют большее абсолютное значение центральности, а менее важные – меньшее. Данный эффект демонстрирует следующий график, показывающий сравнение значений центральностей (рис. 2).

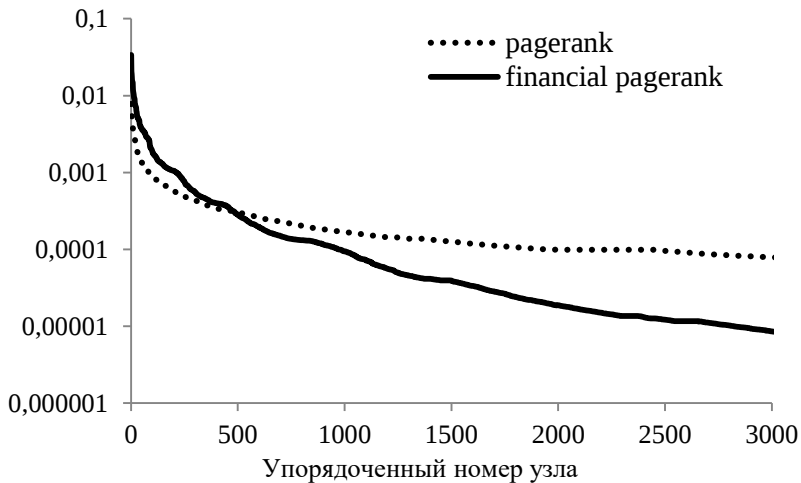


Рис. 2. Сравнение значений центральности
(в логарифмических координатах)

5.1 СХОДИМОСТЬ АЛГОРИТМА

Ключевой внешней переменной алгоритма является количество кластеров n , на которое целесообразно разбивать дуги графа. Нахождение оптимального значения n является отдельной задачей, например, в статье [2] был описан подход к определению оптимального количества кластеров, исходя из поиска локального максимума коэффициента силуэта.

При этом ожидаемо, что при увеличении показателя n алгоритм должен показывать значения, близкие к $n - 1$. Ниже приведен график корреляции Спирмена между центральностями Financial PageRank, рассчитанными для количества кластеров n и $n - 1$.

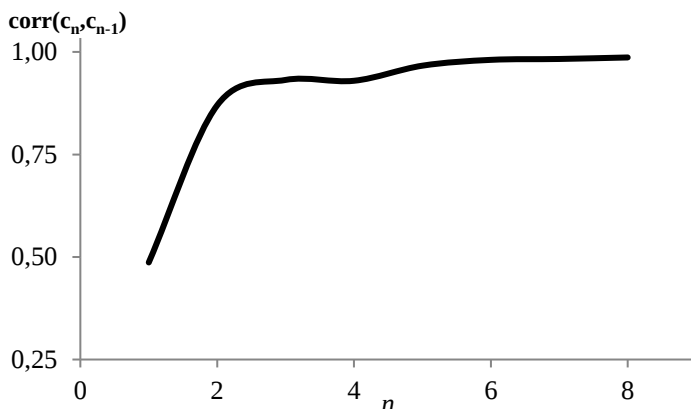


Рис. 3. Корреляция Спирмена для количества кластеров n и $n - 1$

На графике мы видим, что начиная с количества кластеров $n \geq 5$ дальнейшее увеличение количества кластеров не ведет к изменению итоговой оценки.

6. Заключение

Учет весов связей при определении центральности узлов в ряде прикладных задач может иметь критически важный характер. Как нами было показано, это особенно важно в области финансовых вычислений. В представленной работе предложен

подход к реализации данной задачи, показавший возможность применения на реальных данных.

За счет учета весов дуг, предложенная мера центральности имеет большую полярность в абсолютных значениях центральности по сравнению с мерой *PageRank*. Это помогает предложенной мере центральности наилучшим образом осуществлять ранжирование наиболее важных узлов.

Отдельно требуется рассмотреть задачу сходимости предложенного алгоритма при увеличении количества кластеров. Из приведенных расчетов видно, что итоговая центральность сильно отличается от центральности на первом шаге, которая представляет собой центральность *PageRank*. Но при достаточном количестве кластеров (пять и более) дальнейшее увеличение количества кластеров не приводит к существенному изменению предлагаемой меры центральности.

Определение оптимального количества кластеров является отдельной задачей. В [2] были предложен подход к поиску наилучшего количества кластеров, но он не учитывал всех аспектов исследуемой меры центральности.

Также к недостаткам предложенной меры центральности следует отнести сложность ее расчета, особенно при большом количестве кластеров.

Литература

1. ГАНТМАХЕР Ф.Р. *Теория матриц*. – М.: Физматлит. 2004. – 559 с.
2. ЕГОРКИН А.А. *Особенности использования алгоритма классификации k -means для данных, подчиненных степенному закону распределения* // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2023. – №9. – С. 65–69.
3. ЕГОРКИН А.А., АГАЕВ Р.П. *Использование степенного закона в финансовых задачах* // Наука. Производство. Образование: Сборник научных трудов Всероссийской научно-технической конференции, Российский государственный социальный университет, 2023. – С. 58–64

4. ПЫРКИНА О.Е., ЗАДАДАДЕВ С.А. *Разработка критериев безопасности сложных систем электронного денежного оборота с помощью графовой аналитики* // Управление развитием крупномасштабных систем (MLSD'2021): Труды Четырнадцатой международной конференции: Институт проблем управления им. В.А. Трапезникова РАН, 2021. – С. 765–771.
5. ЧЕБОТАРЕВ П.Ю., ГУБАНОВ Д.А. *How to choose the most appropriate centrality measure?* // ArXiv. – 2020. – P. 1–25.
6. ALI M., ISLAM M., SALMAN A. *Average Convergence for Directed & Undirected Graphs in Distributed Systems* // Computer Systems Science and Engineering, 2021. – P. 399–413.
7. AVRACHENKOV K., HOFSTAD R., SOKOL M. *Personalized PageRank with Node-dependent Restart* // Research report, 2014. – 12 p.
8. BARABASI A.L., BONABEAU E. *Scale Free Networks* // Scientific American. – 2003. – P. 60–69.
9. BARDOSCIA M., BARUCCA P., BATTISTON S. et al. *The physics of financial networks* // Nat Rev Phys. – 2021. – No. 3. – P. 490–507.
10. BRIN S., PAGE L. *The anatomy of a large-scale hypertextual web search engine* // Proc. of the 7th Int. Conf. on World Wide Web. ACM, 1998. – P. 107–117.
11. CHEVALIER P-Y, GUSEV V.V., HENDRICKX J.M. et al. *Sets of Stochastic Matrices with Converging Products: Bounds and Complexity.* // ArXiv. – 2017. – 19 p.
12. GHADIRI M., SAMADI S., VEMPALA S. *Socially fair k-means clustering.* // ACM Conf. on Fairness, Accountability, and Transparency. – 2021. – P. 438–448.
13. KATZ L. *A new status index derived from sociometric analysis* // Psychometrika. – 1953. – P. 39–43.
14. LANGVILLE N., MEYER C.D. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* – Princeton University Press, 2006. – 224 p.
15. NEWMAN M.E.J. *Networks: An Introduction* // Oxford University Press, 2012. – 241 p.

16. NEWMAN M.E.J., AARON C., COSMA R.S. *Power-law distributions in empirical data* // SIAM Review. – 2009. – No. 4 – P. 661–703.
17. PANCHENDRARAJAN R., SAXENA A. *Topic-based influential user detection: a survey* // Applied Intelligence. – 2023. – 53 p.
18. PEI Y., FANG L., IPENBURG W. et al. *Subgraph anomaly detection in financial transaction networks*. // ACM Int. Conf. on AI in Finance, 2020. – P. 1–8.
19. SAXENA A. *Evolving Models for Dynamic Weighted Complex Networks*. In *Principles of Social Networking*. – Springer, 2023. – 246 p.
20. SAXENA A., IYENGAR S. *Centrality Measures in Complex Networks: A Survey* // ArXiv. – 2020. – 75 p.
21. SAXENA A., PEI Y., VELDSINK J. et al. *The Banking Transactions Dataset and its Comparative Analysis with Scale-free Networks* // ACM Int. Conf. on Advances in Social Networks Analysis and Mining, 2021. – P. 283–296.

PAGERANK CENTRALITY, ACCOUNTING FOR THE WEIGHTS OF LINKS

Anton Egorkin, Russian State Social University, Moscow, Graduate Student (2-5@bk.ru).

Abstract: The work is devoted to finding the centrality of nodes of weighted graphs. The relevance of this task is due to the fact that ignoring the weights of the arcs of the graph when finding the centrality of its nodes is unacceptable for a number of applied tasks, primarily related to tasks from the financial sphere. In the classical formulation of the PageRank algorithm, part of the information about the weights of connections is lost when forming a matrix of transient probabilities from the adjacency matrix. This effect has been demonstrated in this article. A method for determining the centrality of network nodes is proposed, based on the PageRank algorithm, which allows taking into account all the weights of the links. The graph of financial transactions was considered as an example. The nodes of the graph are the clients of a commercial bank, including the bank itself, and the arcs are money transfers between nodes. The ranking quality was determined by comparing various centrality measures with an external node parameter unrelated to the network characteristics of the transaction graph. According to the results of the study, it was shown that the proposed centrality measure ranks the most important nodes of the

graph in the best way compared to other centrality measures. The convergence of the proposed algorithm was also demonstrated.

Keywords: PageRank, centrality, Markov chains, weighted directed graph.

УДК 519.8

ББК 22.18

DOI: 10.25728/ubs.2024.111.3

*Статья представлена к публикации
членом редакционной коллегии Г.А. Угольниковым.*

Поступила в редакцию 20.04.2024.

Опубликована 30.09.2024.