DOI: 10.17323/2587-814X.2024.1.52.64

Непараметрическая процедура сравнения эффективности работы подразделений сетевой организации

П.А. Колданов [a]
E-mail: pkoldanov@hse.ru

В.А. Колданов [b]
E-mail: vlad.kold@gmail.com

[a] Лаборатория алгоритмов и технологий анализа сетевых структур, Национальный исследовательский университет «Высшая школа экономики»
Адрес: Россия, 603093, г. Н. Новгород, ул. Родионова, 136

[b] Нижегородский государственный инженерно-экономический университет
Адрес: Россия, 606340, Нижегородская область, г. Княгинино, ул. Октябрьская, д. 22А, корпус 2

Аннотация

Для решения задачи сравнительного анализа эффективности работы филиалов по небольшому объему наблюдений случайной природы актуальным является непараметрический подход, не требующей вероятностной модели наблюдений. Актуальной также является задача сравнения результатов непараметрического подхода с результатами, полученными в рамках традиционно применяемой Гауссовской модели. Кроме того, актуальной является проблема получения непротиворечивого сравнения группы (не менее трех) филиалов. В настоящее время непараметрический подход и соответствующее сравнение с известными результатами решения рассматриваемой в настоящей работе задачи, полученного в рамках нормальной модели, отсутствуют. Кроме того, поиску методов получения непротиворечивых решений уделяется недостаточно внимания. Настоящая работа в определенном степени закрывает эти пробелы. Для решения этих задач в настоящей работе используются методы непараметрической статистики и теории одновременной проверки многих гипотез. В работе предлагается процедура сравнительного анализа эффективности работы нескольких подразделений сетевой организации по небольшому объему наблюдений, основанная на тестах Манна–Уитни. Проводится сравнение результатов предлагаемой непараметрической процедуры с результатами, основанными на обобщениях тестов Стьюдента. Предлагается способ уменьшения числа возникающих проблем несовместимости, основанный на поиске подходящего уровня значимости. Приводится пример полностью непротиворечивого сравнения эффективности работы филиалов.

Ключевые слова: сетевая организация, эффективность деятельности филиалов, тесты Манна–Уитни, проблема несовместимости

Цитирование: Колданов П.А., Колданов В.А. Непараметрическая процедура сравнения эффективности работы подразделений сетевой организации // Бизнес-информатика. 2024. Т. 18. № 1. С. 52–64.
DOI: 10.17323/2587-814X.2024.1.52.64

Введение

Различные аспекты проблемы сравнения эффективности деятельности организаций обсуждаются во многих работах, например [1–3]. При этом, как правило, сравнения проводятся по многим показателям. Очевидно, успех таких сравнений зависит от того, насколько адекватно и качественно удается провести сравнения по отдельным показателям, особенно если они имеют случайную природу. В этом последнем случае общепринятым является применение методов математической статистики. Такие методы делятся на:

- параметрические, которые основаны на конкретной вероятностной модели анализируемых показателей. При этом в качестве вероятностной модели чаще всего используется нормальное распределение [4, 5],
- непараметрические, свободные от детальной вероятностной модели, а иногда и от предположения о случайном характере анализируемых данных [3, 6–9].

Многие задачи, в том числе и задача настоящей работы, могут быть рассмотрены в рамках обоих подходов. В этом случае возникает необходимость сравнения выводов процедур, опирающихся на параметрическую модель, с выводами непараметрических процедур. Такое сравнение является одной из целей настоящей работы.

Известно много результатов сравнения параметрических и непараметрических тестов проверки гипотезы против альтернативы. При конечном объеме наблюдений такое сравнение проводится на основе анализа функции мощности тестов, определяемой вероятностью отвержения гипотезы. При неограниченном объеме наблюдений сравнение тестов проверки гипотезы против альтернативы основано на вычислении показателей асимптотической эффективности [10]. Специфика рассматриваемой в настоящей работе задачи заключается в необходимости выбора одного из многих решений по небольшому объему наблюдений.

В настоящей работе предлагается процедура сравнительного анализа эффективности деятельности подразделений сетевой организации. Результаты применения такой процедуры могут быть использованы для принятия обоснованных управленческих решений руководством сетевой организации. При этом под эффективностью деятельности подразделения понимается отношение числа продаж определенного товара (например, числа автомобилей) к числу потенциальных покупателей. Под сетевой организацией понимается совокупность подразделений, работающих по общей схеме. Примерами таких организаций являются сеть филиалов крупной автомобильной компании или сеть магазинов «Пятерочка» и т.д.

В настоящей работе в качестве иллюстративного примера рассматривается задача поддержки управленческих решений руководством сети филиалов крупного университета. На подготовительные курсы таких филиалов принимаются все желающие. Естественной характеристикой эффективности работы персонала таких филиалов является отношение числа слушателей подготовительных курсов к числу потенциальных абитуриентов. Информация о сравнительной эффективности является основой для принятия стратегических решений по развитию сети филиалов.

Сетевую организацию удобно представить в виде графа. Вершины этого графа соответствуют подразделениям. Специфика рассматриваемого нами графа заключается в том, что в нем могут быть как направленные, так и ненаправленные ребра. Ненаправленное ребро между вершинами i, j добавляется в граф тогда и только тогда, когда принимается решение, что i-ое и j-ое подразделения работают одинаково эффективно. Направленное ребро от вершины i к вершине j добавляется в граф тогда и только тогда, когда принимается решение, что i-ое подразделение работает более эффективно, чем j-ое. Заметим, что обычно в графах используются либо только направленные, либо только ненаправленные ребра [11–15]. Мы будем пользоваться и теми, и другими просто потому, что они лучше позволяют подчеркнуть некоторые структуры анализируемого графа, которые характеризуют специфику анализируемой сетевой организации. К числу таких структур мы будем относить клики [16] (совокупность вершин, любые две из которых соединены ненаправленным ребром), которые характеризуют совокупность подразделений, относительно которых принимается решение о том, что они работают одинаково эффективно. В дальнейшем такие клики мы будем называть классами эквивалентности. Другим примером структуры является полный подграф только с направленными ребрами. Такую структуру мы будем называть структурой упорядочения или доминирования.

Во многих задачах, в частности, в анализируемом нами примере, число продаж естественно рассматривать как случайную величину. При этом анализ реальных данных, особенно когда их мало, может приводить к появлению противоречивых выводов. В этом случае соответствующий граф содержит некоторые логически противоречивые структуры, например подграфы из трех вершин, два ребра между которыми ненаправленные, а одно – направленное. Такого типа проблема возникала в [17] при обсуждении задачи проверки гипотез однородности не менее трех совокупностей и называлась проблемой несовместимости. Прикладные задачи, в которых возникает проблема несовместимости (проблема непротиворечивого объединения результатов сравнения эффективности пар подразделений), рассматривались в работах [4, 5] в рамках нормальной модели. Решение проблемы несовместимости было основано на введении дополнительного параметра Δ и переходе к задачам сравнения эффективности двух подразделений с точностью Δ. При этом если эффективности подразделений i, j отличались меньше, чем на Δ, то принималось решение, что их эффективность одинакова (с точностью Δ). Такой прием позволяет решить проблему несовместимости, но приводит к дополнительной проблеме выбора Δ. Другая цель настоящей работы заключается в поиске способов решения проблемы несовместимости без введения вспомогательного параметра Δ.

В настоящей работе, в отличии от [4, 5], предположение о нормальном распределении числа продаж не используется. Попарное сравнение эффективности двух подразделений основано на применении тестов Манна–Уитни. Процедура сравнительного анализа подразделений по эффективности основана на комбинировании непараметрических тестов попарного сравнения двух подразделений. При этом используется графическое представление, которое оказывается удобным для визуализации возникающих проблем несовместимости. Предлагаемая непараметрическая процедура применяется к анализу данных, приведенных в [4], и приводится пример, в котором проблему несовместимости удается решить за счет анализа p-значений и подходящего выбора уровней значимости тестов попарных сравнений. Проводится сравнение с результатами, полученными в рамках нормальной модели.

Статья организована следующим образом: в разделе 1 приведены основные обозначения и постановка задачи; в разделе 2 описана непараметрическая процедура сравнительного анализа подразделений по эффективности и ее графическое представление; в разделе 3 приводятся иллюстративный пример, пример решения проблемы несовместимости и проводится сравнение с результатами, полученными в [4].

1. Постановка задачи

Данные о числе продаж удобно представить в виде матрицы $\|x_{ji}\|$, где $\|x_{ji}\|$ – отношение числа продаж к числу потенциальных покупателей в подразделении j в i-ый временной период, $j = 1, ..., N$, где N – количество подразделений сетевой организации, $i = 1, ..., m_j$, где m_j – количество анализируемых временных периодов работы j-ого подразделения. Будем считать, что наблюдения x_{ji} представляют собой значения случайных величин X_{ji}, которые описывают отношение численности продаж к числу потенциальных покупателей в подразделении j во временной период i. Предположим, что все временные периоды одинаковы, а случайные величины X_{ji} независимы при всех $j = 1, \ldots, N$; $i = 1, \ldots, m_j$ и при фиксированном j одинаково распределены, как X_j. Пусть $F_j(x)$ – функция распределения случайной величины X_j.

Задача, рассматриваемая в настоящей работе, заключается в построении и применении для анализа конкретных данных статистической процедуры различения гипотез вида:

$$\begin{aligned} &H_1 : F_1(x) = F_2(x) = \cdots = F_N(x), \quad \forall x \\ &H_2 : F_1(x) < F_2(x) = \cdots = F_N(x), \quad \forall x \\ &\vdots \\ &H_L : F_1(x) < F_2(x) < \cdots < F_N(x), \quad \forall x \end{aligned} \tag{1}$$

При этом гипотеза H_1 означает, что эффективность всех подразделений одинакова, гипотеза H_2 означает, что подразделение 1 работает более эффективно, чем остальные подразделения, эффективность которых одинакова, и т.д. Заметим, что соотношения (1) не описывают все возможные соотношения между функциями распределения $F_j(x)$, $j = 1, \ldots, N$. Мы ограничиваемся рассмотрением только этих гипотез, так как нас интересует только наличие систематического сдвига, к которому может приводить разная эффективность работы различных подразделений сетевой организации.

Как и в [4] будем использовать метод построения процедур со многими решениями, предложенный в [17]. Этот метод основан на сведении многоальтернативной задачи к совокупности соответствующим образом подобранных двухальтернативных порождающих задач. В нашем случае для различения (1) естественно рассматривать двухальтернативные задачи проверки гипотез $h_{ij} : F_{ij}(x) \geq F(x), \forall x, \forall i, j = 1, \ldots, N$.

При фиксированных i, j комбинация тестов φ_{ij}, φ_{ji} одновременной проверки гипотез h_{ij} и h_{ji} с ненулевой вероятностью может приводить к логически несостоятельному (при данном x) решению об отвержении обеих гипотез, т.е. к проблеме несовместимости. Как показано в [17], для исключения такого противоречия достаточно потребовать, чтобы уровни значимости α_{ij}, α_{ji} тестов φ_{ij}, φ_{ji} удовлетворяли условию $\alpha_{ij} + \alpha_{ji} < 1$. В этом случае комбинация тестов φ_{ij}, φ_{ji} приводит к совместной процедуре различения трех гипотез

$$\begin{aligned} &H_{ij}^1 : F_i(x) < F_j(x); \\ &H_{ij}^2 : F_i(x) = F_j(x); \\ &H_{ij}^3 : F_i(x) > F_j(x). \end{aligned} \tag{2}$$

Однако объединение таких процедур с тремя решениями при различных i, j может привести к противоречию, а именно: с ненулевой вероятностью может быть принято решение о том, что (например):

$$F_1(x) = F_2(x) \text{ и } F_2(x) = F_3(x), \text{ но } F_1(x) \neq F_3(x).$$

Для исключения указанного противоречия в [4], следуя предложению [17], рассматривалась несколько измененная система порождающих гипотез. При этом исследования были ограничены случаем, когда $F_i(x)$ представляет собой нормальное распределение. В обозначениях настоящей работы измененная система порождающих гипотез имеет вид:

$$h'_{ij} : F_{ij}(x) + \Delta \geq F(x), \forall x, \forall i, j = 1, \ldots, N.$$

При комбинировании тестов φ'_{ij}, φ'_{ji} одновременной проверки гипотез h'_{ij}, h'_{ji} получаем процедуру различения трех гипотез:

$$\begin{aligned} &H_{ij}'^1 : F_i(x) + \Delta < F_j(x); \\ &H_{ij}'^2 : |F_i(x) - F_j(x)| < \Delta; \\ &H_{ij}'^3 : F_i(x) > F_j(x) + \Delta. \end{aligned} \tag{3}$$

При этом проблема получения противоречивых выводов не возникает. Вместе с тем введение параметра Δ формально изменяет исходную задачу.

В настоящей работе параметр Δ не вводится и предположение о нормальном распределении не делается. При этом один из интересных вопросов заключается в поиске вариантов непротиворечивого объединения статистических правил с тремя решениями без введения параметра Δ.

Заметим, что в настоящее время в интенсивно развивающейся теории одновременной проверки многих гипотез не делается акцент на необходимости решения проблемы несовместимости [18–21]. При этом начиная с работы [22] проблема несовместимости рассматривается как слишком сильное требование, накладываемое на процедуру одновременной проверки многих гипотез. В рамках теории одновременной проверки многих гипотез в основном изучаются подходы к построению процедур, контролирующих вероятность хотя бы одной ошибки первого рода, доли ошибок первого рода и некоторые другие. В настоящей работе напротив, мы делаем акцент на решении проблемы несовместимости, что позволяет провести адекватное сравнение с результатами, полученными в [4].

2. Непараметрическая процедура сравнительного анализа и ее визуализация

2.1. Процедура с тремя решениями

Одна из наиболее эффективных непараметрических процедур различения гипотез (2) основана на статистике Манна–Уитни [6, 9]. Статистика Манна–Уитни имеет вид:

$$W_{ij}(x_i, x_j) = \sum_{r=1}^{m_j} \sum_{s=1}^{m_i} I(x_{is} < x_{jr}), \qquad (4)$$

где $I(A)$ – индикатор события A.

При фиксированных i, j процедура с тремя решениями различения гипотез (2) в терминах p-значений может быть записана в виде

$$\varphi(i,j) = \begin{cases} d_{ij}^1 & p_{ij}^1 \le \alpha_1 \\ d_{ij}^3 & p_{ij}^3 \le \alpha_3 \\ d_{ij}^2 & p_{ij}^1 > \alpha_1,\ p_{ij}^3 > \alpha_3, \end{cases} \qquad (5)$$

где d_{ij}^k — решение о принятии гипотезы H_{ij}^k $(k = 1, 2, 3)$; p_{ij}^1, p_{ij}^3 — соответствующие p-значения, а именно:

$$\begin{aligned} p_{ij}^1 &= P_{F_i = F_j}\Big(W_{ij}(X_i, X_j) \le W_{ij}(x_i, x_j)\Big) \\ p_{ij}^3 &= P_{F_i = F_j}\Big(W_{ij}(X_i, X_j) > W_{ij}(x_i, x_j)\Big), \end{aligned} \qquad (6)$$

где α_1, α_3 – уровни значимости тестов проверки гипотез H_{ij}^1 и H_{ij}^3 соответственно.

При этом предполагается, что среди наблюдаемых значений $x_{j1}, \ldots, x_{j_{m_j}}$ и $x_{i1}, \ldots, x_{i_{m_i}}$ нет равных. Необходимые корректировки в случае равных наблюдений можно внести, опираясь на методику, изложенную в [9].

Таблицы распределения статистики (4) при малых m_i, m_j приведены в [9]. При больших m_i, m_j можно использовать нормальное распределение

$$N\left(\frac{m_i \cdot m_j}{2}, \frac{m_i \cdot m_j \cdot (m_i + m_j + 1)}{12}\right),$$

которое рекомендуется использовать при

$$\min(m_i, m_j) > 50\ [9].$$

Из (6) очевидно, что при фиксированных i, j справедливо $p_{ij}^1 + p_{ij}^3 = 1$. Поэтому для применения процедуры с тремя решениями (5) достаточно информации о минимальном p-значении:

$$p_{ij} = \min(p_{ij}^1, p_{ij}^3). \qquad (7)$$

2.2. Процедура со многими решениями и ее графическое представление

Процедуру различения гипотез (1) будем получать комбинированием процедур (5).

Такая процедура может быть записана в виде:

$$\delta = \Big(\varphi(1,2), \varphi(2,3), \ldots, \varphi(N-1, N)\Big). \qquad (8)$$

Через f_i, $i = 1, \ldots, N$ обозначим эффективность работы i-го подразделения. В рассматриваемой задаче между эффективностями работы подразделений сетевой организации $(f_1, \ldots, f_N)$ возможны два типа отношений (доминирование или эквивалентность). Запись $f_i > f_j$ означает, что i-ое подразделение работает более эффективно (доминирует), чем j-ое подразделение. Запись $f_i = f_j$ означает, что i-ое и j-ое подразделения работают одинаково эффективно (эквивалентность). Для наглядного анализа результатов применения процедуры (8) воспользуемся методикой, предложенной в [23–25].

При данном векторе $(f_1, \ldots, f_N)$ введем матрицу D с элементами

$$d_{ij} = \begin{cases} 1, & f_i > f_j \\ 0, & f_i = f_j \\ -1, & f_i < f_j \end{cases}$$

и матрицу B с элементами

$$b_{ij} = \begin{cases} 1, & f_i > f_j \\ 0, & \text{иначе}. \end{cases}$$

Легко показать [24], что матрица B связана с матрицей D соотношением

$$D = B - B^T,$$

где B^T – транспонированная матрица B.

Соотношения, описываемые матрицей B, удобнее интерпретировать, если строки и столбцы матрицы B переставлены таким образом, чтобы получить верхнюю треугольную форму (т.е. собрать по возможности все единицы над главной диагональю матрицы B). Для получения верхней треугольной формы матрицы B можно упорядочить по убыванию строки (и столбцы) матрицы D по суммам элементов строк. Заменив элементы –1 полученной матрицы на 0, получим матрицу B, наиболее согласованную с верхней треугольной формой. Матрица B, наиболее согласованная с верхней треугольной формой, позволяет выделить так называемые «классы эквивалентности» («undifferentiated classes» [25]). Термин «класс эквивалентности» будет использоваться для обозначения наибольшего множества подразделений, эффективности работы которых значимо неотличимы друг от друга. Термин «наибольший» означает, что для любого подразделения i, которое не принадлежит данному классу эквивалентности, существует по крайней мере одно подразделение j из этого класса, такое что эффективности работы подразделений i и j значимо различимы. Матрица B, наиболее согласованная с верхней треугольной формой, показывает классы эквивалентности, как квадратные подматрицы, симметричные относительно главной диагонали, все элементы которых равны 0.

Очевидно, что пересекающиеся классы эквивалентности, получающиеся при применении процедуры δ, означают существование проблемы несовместимости. Матрицу D удобно визуализировать в виде графа $G = (V, E)$, где $V = \{1, 2, ..., N\}$ – множество вершин графа, $E = \{e_{ij}\}$ – множество ребер графа. Если элемент d_{ij} матрицы D равен 1, то в граф G добавляется направленное ребро от вершины i к вершине j. При этом вершина i доминирует вершину j. В [11] вершина i называется родителем вершины j, а вершина j – ребенком вершины i. Если вершина i соединена направленным путем длины больше 1 с некоторой вершиной k, то вершина i называется предком вершины k, а вершина k – потомком вершины i. Если элемент d_{ij} матрицы D равен 0, то в граф G добавляется ненаправленное ребро между вершинами i и j. Если элемент $d_{ij} = -1$, то, так как $d_{ij} = -d_{ji}$, граф G уже содержит направленное ребро от вершины j к вершине i. При этом очевидно, что все вершины этого графа, соответствующие подразделениям из некоторого класса эквивалентности, соединены между собой ненаправленными ребрами и поэтому представляют собой клики графа G. Ниже мы будем отдельно изображать подграфы только с направленными ребрами и только с ненаправленными ребрами.

Заметим, что предлагаемое представление будет явно отображать существование проблем несовместимости полученных выводов, в случае их наличия. Очевидно, что если в представлении различные классы эквивалентности содержат одинаковые вершины, то проблема несовместимости имеет место.

3. Иллюстративный пример

Рассмотрим задачу сравнительного анализа эффективностей работы филиалов университета, которая была кратко описана во введении. Обозначим 1f – *первый* филиал университета, 2f – *второй* филиал университета и т.д. Данные для анализа заимствованы из [4] и приведены в *таблице 1.*

Минимальные p-значения (7) тестов (5) приведены в *таблице 2.*

Таблица 1.

Данные о числе слушателей подготовительных курсов в различных филиалах

1f	2f	3f	4f	5f	6f	7f	8f
103	131	187	154				
92	212	262	92	151	99	235	
122	197	376	129	164	268	338	77
48	143	283	146	141	217	239	63
86	95	231	125	140	231	187	59
89	70	203	127	173	175	123	78
147	92	276	183	141	137	139	82
134	95	258	213	187	242	185	28
Количество потенциальных абитуриентов в i-ом филиале							
6390	7090	28900	6320	6320	11130	4660	2530

Таблица 2.

Минимальные *p*-значения

	$1f$	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$
$2f$	p^1_{12} 0,4392						
$3f$	p^3_{13} 0,0023	p^3_{23} 0,0103					
$4f$	p^1_{14} 0,0364	p^1_{24} 0,0652	p^1_{34} 0,0001				
$5f$	p^1_{15} 0,0006	p^1_{25} 0,0469	p^1_{35} 0,0002	p^1_{45} 0,1678			
$6f$	p^1_{16} 0,3063	p^1_{26} 0,4775	p^1_{36} 0,0070	p^3_{46} 0,0760	p^3_{56} 0,0020		
$7f$	p^1_{17} 0,0002	p^1_{27} 0,0011	p^1_{37} 0,0002	p^1_{47} 0,0200	p^1_{57} 0,0012	p^1_{67} 0,0003	
$8f$	p^3_{18} 0,0147	p^3_{28} 0,0539	p^1_{38} 0,0007	p^1_{48} 0,1725	p^1_{58} 0,1830	p^1_{68} 0,0256	p^3_{78} 0,0175

3.1. Построение классов эквивалентности

Рассмотрим прежде всего традиционный уровень значимости $\alpha_{ij} = 0{,}05,\ \forall\, i, j = 1, \ldots, 8$. Матрица $D_{0,05}$ приведена в *таблице 3*.

Таблица 3.

Матрица $D_{0,05}$

	$1f$	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$	$8f$	Σ
$1f$	–	0	1	–1	–1	0	–1	–1	–3
$2f$	0	–	1	0	–1	0	–1	0	–1
$3f$	–1	–1	–	–1	–1	–1	–1	–1	–7
$4f$	1	0	1	–	0	0	–1	0	1
$5f$	1	1	1	0	–	1	–1	0	3
$6f$	0	0	1	0	–1	–	–1	–1	–2
$7f$	1	1	1	1	1	1	–	1	7
$8f$	1	0	1	0	0	1	–1	–	2

Графическое представление матрицы $D_{0,05}$ показано на *рисунке 1*.

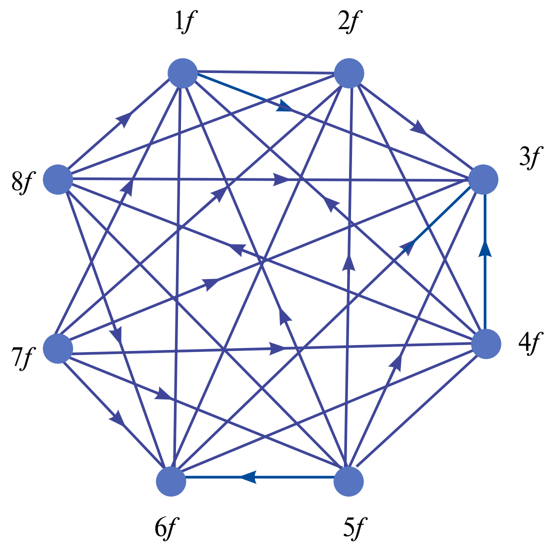

Рис. 1. Графическое представление матрицы $D_{0,05}$.

Матрица $B_{0,05}$, полученная из матрицы $D_{0,05}$, приведенная к верхней треугольной форме, показана в *таблице 4*.

Таблица 4.

Матрица $B_{0,05}$,
приведенная к верхней треугольной форме

	$7f$	$5f$	$8f$	$4f$	$2f$	$6f$	$1f$	$3f$
$7f$	–	1	1	1	1	1	1	1
$5f$		–	0	0	1	1	1	1
$8f$			–	0	0	1	1	1
$4f$				–	0	0	1	1
$2f$					–	0	0	1
$6f$						–	0	1
$1f$							–	1
$3f$								–

Графическое представление матрицы $B_{0,05}$, приведенной к верхней треугольной форме, показано на *рисунке 2*.

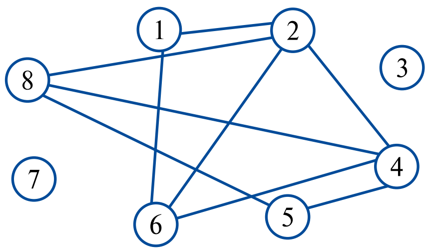

Рис. 2. Графическое представление матрицы $B_{0,05}$, приведенной к верхней треугольной форме. Номер филиала обозначен цифрой.

На *рис. 2* легко видеть, что в данном графе шесть клик: {7}, {4, 5, 8}, {2, 4, 8}, {2, 4, 6}, {2, 6, 1}, {3}. Отметим, что эти клики имеют общие вершины, например, вершина 4 принадлежит трем кликам. Это свидетельствует о существовании проблемы несовместимости. Таким образом, при $\alpha_{ij} = 0{,}05$, $\forall\, i, j = 1, \ldots, 8$ непересекающиеся классы эквивалентности не выделяются.

Хорошо известно, что строгих правил выбора уровня значимости α не существует. Возможный способ уменьшения числа возникающих проблем несовместимости может быть основан на изменении уровня значимости α. Анализ p-значений, приведенных в *таблице 2*, подсказывает целесообразность построения и изучения графических моделей при выборе α_{ij}, $\forall\, i, j = 1, \ldots, 8$ из интервала (p^3_{46} =0,076; p^1_{45} =0,1678). Для определенности выберем $\alpha_{ij} = 0{,}1$, $\forall\, i, j = 1, \ldots, 8$. Матрица $D_{0,1}$ приведена в *таблице 5*.

Таблица 5.

Матрица $D_{0,1}$

	$1f$	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$	$8f$	Σ
$1f$	–	0	1	–1	–1	0	–1	–1	–3
$2f$	0	–	1	–1	–1	0	–1	–1	–3
$3f$	–1	–1	–	–1	–1	–1	–1	–1	–7
$4f$	1	1	1	–	0	1	–1	0	3
$5f$	1	1	1	0	–	1	–1	0	3
$6f$	0	0	1	–1	–1	–	–1	–1	–3
$7f$	1	1	1	1	1	1	–	1	7
$8f$	1	1	1	0	0	1	–1	–	3

Графическое представление матрицы $D_{0,1}$ показано на *рисунке 3*.

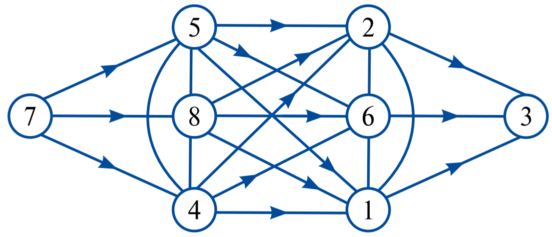

Рис. 3. Графическое представление матрицы $D_{0,1}$. Номер филиала обозначен цифрой.

Матрица $B_{0,1}$, полученная из матрицы $D_{0,1}$, приведенная к верхней треугольной форме, показана в *таблице 6*.

Таблица 6.

Матрица $B_{0,1}$,
приведенная к верхней треугольной форме

	$7f$	$5f$	$8f$	$4f$	$2f$	$6f$	$1f$	$3f$
$7f$	–	1	1	1	1	1	1	1
$5f$		–	0	0	1	1	1	1
$8f$			–	0	1	1	1	1
$4f$				–	1	1	1	1
$2f$					–	0	0	1
$6f$						–	0	1
$1f$							–	1
$3f$								–

Графическое представление матрицы $B_{0,1}$, приведенной к верхней треугольной форме, показано на *рисунке 4*.

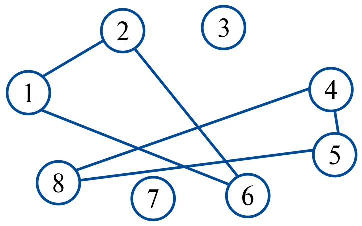

Рис. 4. Графическое представление матрицы $B_{0,1}$, приведенной к верхней треугольной форме. Номер филиала обозначен цифрой.

На *рисунке 4* легко видеть, что в данном графе четыре клики {7}, {4, 5, 8}, {2, 6, 1}, {3} причем эти клики не имеют общих вершин. Это свидетельствует о том, что при $\alpha_{ij} = 0{,}1,\ \forall\, i, j = 1, \ldots, 8$ проблема несовместимости не возникает, т.е. классы эквивалентности не пересекаются.

3.2. Построение структур упорядочения

Графическое представление матрицы $D_{0,05}$ показывает структуру упорядочения филиалов по эффективности их работы, приведенную на *рисунке 5*.

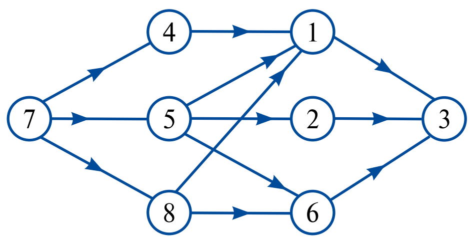

Рис. 5. Непараметрическое упорядочение $\alpha = 0{,}05$. Номер филиала обозначен цифрой.

Показаны только связи доминирования, связи эквивалентности и связи между предками и потомками не показаны. В частности, не проведено направленное ребро между вершинами 7 и 3, т.к. существует направленный путь $7 \to 4 \to 1 \to 3$ от предка 7*f* к потомку 3*f*, означающий строгое упорядочение: 7*f* эффективнее 4*f*, 4*f* эффективнее 1*f*, 1*f* эффективнее 3*f*. Заметим, что в данном графе нет направленных путей $7 \to 4 \to 2 \to 3$; $7 \to 4 \to 6 \to 3$; $7 \to 8 \to 2 \to 3$. Это свидетельствует об отсутствии полного упорядочения между эффективностями работы филиалов при $\alpha = 0{,}05$. Учитывая, что {4*f*, 5*f*, 8*f*} принадлежат одному классу эквивалентности, отсутствие таких путей приводит к логическим противоречиям. Подчеркнем, что при $\alpha = 0{,}1$ полное упорядочение имеет место (см. *рис. 6*) и логические противоречия не возникают.

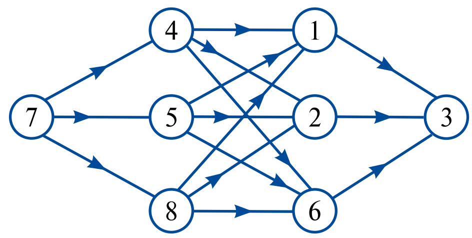

Рис. 6. Непараметрическое упорядочение $\alpha = 0{,}1$. Номер филиала обозначен цифрой.

3.3. Сравнение

На *рисунке 7* показаны графы, построенные по результатам сравнения эффективностей работы филиалов как в предположении нормального распределения изучаемых случайных величин [4], так и в непараметрической постановке.

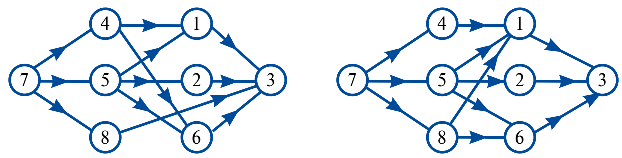

Рис. 7. Параметрическое (слева), непараметрическое (справа) упорядочения при $\alpha = 0{,}05$. Номер филиала обозначен цифрой.

Графы упорядочения, представленные на *рисунке 7*, отличаются тремя ребрами, а именно: ребро (4, 6) присутствует при параметрическом упорядочении, и отсутствует при непараметрическом упорядочении; ребра (8, 1); (8, 6) присутствует при непараметрическом упорядочении, и отсутствует при параметрическом упорядочении.

Линейное упорядочение, построенное по схеме, предложенной в [4], соответствующее непараметрическому упорядочению (*рис. 7*, справа), имеет вид:

$$f_3 < f_1 \le f_6 \le f_2 \le f_4 \le f_8 \le f_5 < f_7 \text{ с точностью } \Delta. \qquad (9)$$

Получение такого линейного упорядочения, формально предложенного в [4] в несколько иной постановке (упорядочение с точностью Δ), основано на анализе числа направленных связей, выходящих из конкретной вершины или входящих в конкретную вершину. А именно, так как вершины 4, 5, 8 (*рис. 7*, справа) не соединяются направленными ребрами, то на первый взгляд кажется, что может быть вынесено решение $f_4 = f_8 = f_5$ с точностью Δ. Однако, так как вершина 4 доминирует только одну вершину 1 (из вершины 4 выходит одно направленное ребро), вершина 8 доминирует две вершины 1 и 6 (из вершины 8 выходит два направленных ребра), и вершина 5 доминирует три вершины 1, 2, 6 (из вершины 5 выходит три направленных ребра), то выносится решение $f_4 \le f_8 \le f_5$ с точностью Δ. Запись $f_4 \le f_8$ с точностью Δ означает, что $f_4 + \Delta < f_8$ или $|f_4 - f_8| < \Delta$. Аналогично, так как вершины 1, 2, 6 (*рис. 7*, справа) не соединяются направленными ребрами, то на первый взгляд кажется, что может быть вынесено решение $f_1 = f_2 = f_6$ с точностью Δ. Однако, так как вершина 1 – потомок всех вершин 4, 5, 8 (в вершину 1 входит три направленных ребра), вершина 6 – потомок двух вершин 5 и 8 (в вершину 6 входит два направленных ребра), вершина 2 – потомок одной вершины 5 (в вершину 2 входит одно направленное ребро), то выносится решение $f_1 \le f_6 \le f_2$ с точностью Δ.

Линейное упорядочение, полученное в [4] имеет вид:

$$f_3 < f_1 = f_6 \le f_2 \le f_8 \le f_4 \le f_5 \le f_7 \text{ с точностью } \Delta. \quad (10)$$

Упорядочения (9) и (10) отличаются очень незначительно. В самом деле, крайние элементы упорядочений совпадают, изменился знак $f_1 \le f_6$ по сравнению со знаком $f_1 = f_6$, и нестрогое упорядочение $f_4 \le f_8$ изменилось на упорядочение $f_8 \le f_4$. Но в обоих случаях значимого различия эффективности работы филиалов f_8 и f_4 не обнаружено. Это косвенно свидетельствует о приемлемости нормальной модели, предложенной в [4].

Классы эквивалентности, представленные на *рисунке 8*, отличаются тремя ребрами, а именно: ребро (4, 6) присутствует при непараметрическом построении классов эквивалентности, и отсутствует при параметрическом построении; ребра (8, 1); (8, 6) присутствует при параметрическом построении классов эквивалентности, и отсутствует при непараметрическом построении. Это вполне согласуется со сравнением структур упорядочения (см. *рис. 7*).

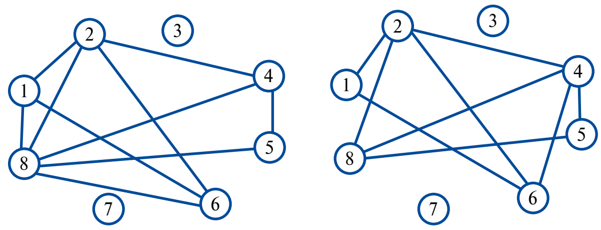

Рис. 8. Параметрические (слева), непараметрические (справа) классы эквивалентности при $\alpha = 0{,}05$. Номер филиала обозначен цифрой.

Заключение

В работе построена непараметрическая процедура сравнительного анализа эффективности работы нескольких подразделений сетевой организации по небольшому объему наблюдений. Приведен пример применения предлагаемого подхода к сравнительному анализу эффективностей работы филиалов ВУЗа. Выполнено сравнение результатов сравнительного анализа, полученных предлагаемой непараметрической процедурой с результатами, полученными в рамках нормальной модели [4]. Показано, что результаты непараметрического упорядочения без введения дополнительного параметра неопределенности Δ и результаты упорядочения, полученные в рамках нормальной модели с введением Δ, достаточно близки. Приведен пример полного непротиворечивого сравнения эффективности работы нескольких подразделений сетевой организации. ■

Благодарности

Результаты разделов 1 и 2 подготовлены в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ). Результаты раздела 3 подготовлены при поддержке гранта Российского научного фонда (проект 22-11-00073).

Литература

1. Алескеров Ф.Т., Белоусова В.Ю. Эффективное развитие филиальной сети коммерческого банка // Управление в кредитной организации. 2007. № 6. С. 23–34.
2. Крюков А.М. Анализ эффективности деятельности филиалов и подразделений – необходимое условие устойчивости бизнеса // Стратегические решения и риск-менеджмент. 2010. Т. 4. С. 84–87. https://doi.org/10.17747/2078-8886-2010-4-84-87

3. Мызникова М.А. Качество стратегического управления в условиях неопределенности: оценка в контексте устойчивого развития // Бизнес-информатика. 2022. Т. 16. № 3. С. 36–52. https://doi.org/10.17323/2587-814X.2022.3.36.52

4. Колданов А.П., Колданов П.А. Статистические процедуры со многими решениями в задаче анализа итогов приема в филиалы ВУЗа // Бизнес-информатика. 2012. Т. 19. № 1. С. 24–31.

5. Koldanov P.A. Efficiency analysis of branch network // Models, Algorithms, and Technologies for Network Analysis (eds. B.I. Goldengorin, V.A. Kalyagin, P.M. Pardalos). Springer Proceedings in Mathematics & Statistics. 2013. Vol. 59. P. 71–83. https://doi.org/10.1007/978-1-4614-8588-9_5

6. Хеттманспергер Т. Статистические выводы, основанные на рангах. М.: Финансы и статистика, 1987.

7. Кендалл М. Ранговые корреляции. М.: Статистика, 1975.

8. Холлендер М., Вулф Д. Непараметрические методы статистики. М.: Финансы и статистика, 1983.

9. Lehmann E.L. Nonparametrics: Statistical methods based on ranks. San Francisco: Hoden-Day, 1975. https://doi.org/10.1002/zamm.19770570922

10. Кендалл М., Стюарт А. Статистические выводы и связи. М.: Наука, 1973.

11. Jordan M.I. Graphical models // Statistical Science. 2004. Vol. 19. No. 1. P. 140–155. https://doi.org/10.1214/088342304000000026

12. Newman M.E.J. Networks. An introduction. New York: Oxford University Press, 2010.

13. Opsahl T., Panzarasa P. Clustering in weighted networks // Social Networks. 2009. Vol. 31. No. 2. P. 155–163.

14. Horvath S. Weighted network analysis. Applications in genomics and systems biology. New York: Springer, 2011. https://doi.org/10.1007/978-1-4419-8819-5

15. Li S., He J., Zhuang Y. A network model of the interbank market // Physica A: Statistical Mechanics and its Applications. 2010. Vol. 389. No. 24. P. 5587–5593. https://doi.org/10.1016/j.physa.2010.08.057

16. Boginski V., Butenko S., Pardalos P.M. Statistical analysis of financial networks // Computational Statistics & Data Analysis. 2005. Vol.48. No. 2. P. 431–443. https://doi.org/10.1016/j.csda.2004.02.004

17. Lehmann E.L. A theory of some multiple decision problems, I // The Annals of Mathematical Statistics. 1957. Vol. 28. No. 1. P. 1–25. https://doi.org/10.1214/aoms/1177707034

18. Bretz F., Hothorn T., Westfall P. Multiple comparisons using R. New York: Chapman and Hall/CRC, 2010. https://doi.org/10.1201/9781420010909

19. Hochberg Y., Tamhane A.C. Multiple comparison procedures. New York: John Wiley & Sons, 1987. https://doi.org/10.1002/9780470316672

20. Handbook of Multiple Comparisons / [Edited by X. Cui, T. Dickhaus, Y. Ding, J.C. Hsu]. New York: Chapman and Hall/CRC, 2021. https://doi.org/10.1201/9780429030888

21. Dickhaus T. Simultaneous statistical inference. With applications in the life sciences. Heidelberg: Springer Berlin, 2014. https://doi.org/10.1007/978-3-642-45182-9

22. Gabriel K.R. Simultaneous test procedures – some theory of multiple comparisons // The Annals of Mathematical Statistics. 1969. Vol. 40. No. 1. P. 224–250. https://doi.org/10.1214/aoms/1177697819

23. Cliff N. Complete orders from incomplete data: Interactive ordering and tailored testing // Psychological Bulletin. 1975. Vol. 82. No. 2. P. 289–302. https://doi.org/10.1037/h0076373

24. Cliff N. Dominance statistics: Ordinal analyses to answer ordinal questions // Psychological Bulletin. 1993. Vol. 114. No. 3. P. 494–509. https://doi.org/10.1037/0033-2909.114.3.494

25. Donoghue J.R. Implementing Shaffer's multiple comparison procedure for a large number of groups // Recent Developments in Multiple Comparison Procedures (Eds. Benjamini, Y., Bretz, F. and Sarkar, S.), Institute of Mathematical Statistics, Beachwood, Ohio, 2004. P. 1–23.

Об авторах

Колданов Петр Александрович

д.ф.-м.н.;

старший научный сотрудник, лаборатория алгоритмов и технологий анализа сетевых структур, Национальный исследовательский университет «Высшая школа экономики», Россия, 603093, г. Н. Новгород, ул. Родионова, 136;

E-mail: pkoldanov@hse.ru

ORCID: 0000-0001-5961-0282

Колданов Владимир Александрович

к.ф.-м.н.;

доцент, кафедра математики и вычислительной техники, Нижегородский государственный инженерно-экономический университет, Россия, 606340, Нижегородская область, г. Княгинино, ул. Октябрьская, д. 22А, корпус 2;

E-mail: vlad.kold@gmail.com

ORCID: 0009-0006-1924-746X

Nonparametric procedure for comparing the performance of divisions of a network organization

Petr A. Koldanov [a]
E-mail: pkoldanov@hse.ru

Vladimir A. Koldanov [b]
E-mail: vlad.kold@gmail.com

[a] Laboratory of Algorithms and Technologies for Network Analysis, HSE University
Address: 136, Rodionova St., Nizhny Novgorod 603093, Russia

[b] Nizhny Novgorod State University of Engineering and Economics
Address: 22A, building 2, Oktyabrskaya St., Nizhny Novgorod Region, Knyaginino 606340, Russia

Abstract

To solve the problem of comparative efficiency analysis of branch operations for a small volume of randomly observed data, a non-parametric approach is relevant, since it does not require a probabilistic model of observations. Comparing the results of the non-parametric approach with the results obtained within the traditionally used Gaussian model is also relevant. Additionally, obtaining a consistent comparison of a group (of no less than three) branches is important. Currently, the non-parametric approach and the corresponding comparison with the known results of solving the problem considered in this work obtained within the framework of the normal model are absent. In addition, insufficient attention is paid to the search for methods of obtaining consistent solutions. This work to some extent fills these gaps. This work uses non-parametric statistical methods and theory of simultaneous hypothesis testing to address these problems. This paper proposes a procedure for comparative analysis of the efficiency of several units within a network organization with a small volume of observations based on the Mann–Whitney tests. We carry out a comparison of the results obtained from the proposed non-parametric procedure with results based on extensions of Student's t-tests. We propose a method for reducing the number of compatibility problems based on the search for an appropriate significance level. We provide an example of a fully consistent comparison of the efficiency of branch operations.

Keywords: network organization, efficiency of branch operations, Mann–Whitney tests, incompatibility problem

Citation: Koldanov P.A., Koldanov V.A. (2024) Nonparametric procedure for comparing the performance of divisions of a network organization. *Business Informatics*, vol. 18, no. 1, pp. 52–64. DOI: 10.17323/2587-814X.2024.1.52.64

References

1. Aleskerov F.T., Belousova V. (2007) Effective development of the branch network of a commercial bank. *Upravleniye v kreditnoy organizatsii (Management in a credit institution)*, no. 6, pp. 23–34 (in Russian).

2. Kryukov A.M. (2010) Efficiency analysis of branch and division activities is a prerequisite for business stability. *Strategic decisions and risk management*, vol. 4, pp. 84–87 (in Russian). https://doi.org/10.17747/2078-8886-2010-4-84-87

3. Myznikova M.A. (2022) Quality of strategic management under ambiguity: Assessment within the framework of sustainable development. *Business Informatics*, vol. 16, no. 3, pp. 36–52. https://doi.org/10.17323/2587-814X.2022.3.36.52

4. Koldanov A.P., Koldanov P.A. (2012) Multiple decision procedures for the analysis of higher school entry selection results. *Business Informatics (Biznes-informatika)*, vol. 19, no. 1, pp. 24–31 (in Russian).

5. Koldanov P.A. (2013) Efficiency analysis of branch network. *Models, Algorithms, and Technologies for Network Analysis* (eds. B.I. Goldengorin, V.A. Kalyagin, P.M. Pardalos). Springer Proceedings in Mathematics & Statistics, vol. 59, pp. 71–83. https://doi.org/10.1007/978-1-4614-8588-9_5

6. Hettmansperger T. (1987) *Statistical Inference Based on Ranks*. Moscow: Finance and Statistics (in Russian).

7. Kendall M. (1975) *Rank correlations*. Moscow: Statistics (in Russian).

8. Hollander M., Wolfe D. (1983) *Nonparametric statistical methods*. Moscow: Finance and Statistics (in Russian).

9. Lehmann E.L. (1975) *Nonparametrics: Statistical methods based on ranks.* San Francisco: Hoden-Day. https://doi.org/10.1002/zamm.19770570922

10. Kendall M., Stuart A. (1973) *Statistical Inferences and relationships*. Moscow: Science (in Russian).

11. Jordan M.I. (2004) Graphical models. *Statistical Science*, vol. 19, no. 1, pp. 140–155. https://doi.org/10.1214/088342304000000026

12. Newman M.E.J. (2010) *Networks. An introduction*. New York: Oxford University Press.

13. Opsahl T., Panzarasa P. (2009) Clustering in weighted networks. *Social Networks*, vol. 31, no. 2, pp. 155–163.

14. Horvath S. (2011) *Weighted network analysis. Applications in genomics and systems biology*. New York: Springer. https://doi.org/10.1007/978-1-4419-8819-5

15. Li S., He J., Zhuang Y. (2010) A network model of the interbank market. *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 24, pp. 5587–5593. https://doi.org/10.1016/j.physa.2010.08.057

16. Boginski V., Butenko S., Pardalos P.M. (2005) Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, vol.48, no. 2, pp. 431–443. https://doi.org/10.1016/j.csda.2004.02.004

17. Lehmann E.L. (1957) A theory of some multiple decision problems, I. *The Annals of Mathematical Statistics*, vol. 28, no. 1, pp. 1–25. https://doi.org/10.1214/aoms/1177707034

18. Bretz F., Hothorn T., Westfall P. (2010) *Multiple comparisons using R.* New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781420010909

19. Hochberg Y., Tamhane A.C. (1987) *Multiple comparison procedures*. New York: John Wiley & Sons. https://doi.org/10.1002/9780470316672

20. *Handbook of Multiple Comparisons* (2021) (eds. X. Cui, T. Dickhaus, Y. Ding, J.C. Hsu). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429030888

21. Dickhaus T. (2014) *Simultaneous statistical inference. With applications in the life sciences.* Heidelberg: Springer Berlin. https://doi.org/10.1007/978-3-642-45182-9

22. Gabriel K.R. (1969) Simultaneous test procedures – some theory of multiple comparisons. *The Annals of Mathematical Statistics*, vol. 40, no. 1, pp. 224–250. https://doi.org/10.1214/aoms/1177697819

23. Cliff N. (1975) Complete orders from incomplete data: Interactive ordering and tailored testing. *Psychological Bulletin*, vol. 82, no. 2, pp. 289–302. https://doi.org/10.1037/h0076373

24. Cliff N. (1993) Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, vol. 114, no. 3, pp. 494–509. https://doi.org/10.1037/0033-2909.114.3.494

25. Donoghue J.R. (2004) Implementing Shaffer's multiple comparison procedure for a large number of groups. *Recent Developments in Multiple Comparison Procedures* (Eds. Benjamini, Y., Bretz, F. and Sarkar, S.), Institute of Mathematical Statistics, Beachwood, Ohio, pp. 1–23.

About the authors

Petr A. Koldanov

Dr. Sci. (Phys.-Math.);

Senior Research Fellow, Laboratory of Algorithms and Technologies for Networks Analysis, HSE University (campus in Nizhny Novgorod), 136, Rodionova St., Nizhny Novgorod 603093, Russia;

E-mail: pkoldanov@hse.ru

ORCID: 0000-0001-5961-0282

Vladimir A. Koldanov

Cand. Sci. (Phys.-Math.);

Associate Professor, Department of Mathematics and Computer Science, Nizhny Novgorod State University of Engineering and Economics, 22A, building 2, Oktyabrskaya St., Nizhny Novgorod Region, Knyaginino 606340, Russia;

E-mail: vlad.kold@gmail.com

ORCID: 0009-0006-1924-746X