

**МЕТОДИЧЕСКИЙ ПОДХОД К ОЦЕНИВАНИЮ СОСТОЯНИЯ  
СИСТЕМЫ УПРАВЛЕНИЯ РАСПРЕДЕЛЕННОГО ОБЪЕКТА  
НА ОСНОВЕ РЕШЕНИЯ ЗАДАЧИ ОПРЕДЕЛЕНИЯ ТОПОЛОГИЧЕСКИХ  
СВОЙСТВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ**

**METHODICAL APPROACH TO ASSESSING THE STATE OF THE CONTROL  
SYSTEM OF A DISTRIBUTED OBJECT BASED ON SOLVING THE PROBLEM OF  
DETERMINING TOPOLOGICAL PROPERTIES INFORMATION  
AND TELECOMMUNICATION NETWORKS**

*Канд. техн. наук Ф.Л. Шуваев, канд. воен. наук Р.В. Васильев*

*Ph.D. F.L. Shuvaev, Ph.D. R.V. Vasil'ev*

*ВКА им. А.Ф. Можайского*

В данной статье приведен методический подход, применяемый при решении задач оценивания топологических свойств сетей с учетом определения важности ее признаков. Данная задача может решаться в интересах распознавания структур информационно-телекоммуникационных сетей с целью оценивания состояния системы управления объекта исследования. Подход основан на использовании методов теории математической статистики, а именно — метода главных компонент. Апробация предложенных решений проведена с использованием моделей случайных графов. Полученные результаты свидетельствуют о необходимости анализа признакового пространства для повышения точности систем распознавания при решении задач оценивания состояния системы управления сложного объекта.

**Ключевые слова:** метод главных компонент, граф, признак, центральность, важность, распознавание, топологические свойства.

This article presents a methodological approach used in solving problems of evaluating the topological properties of networks, taking into account the determination of the importance of its features. This problem can be solved in the interests of recognizing the structures of information and telecommunications networks in order to assess the state of the control system of the object of study. The approach is based on the use of methods of the theory of mathematical statistics, namely, the method of principal components. The proposed solutions were tested using random graph models. The obtained results indicate the need to analyze the feature space to improve the accuracy of recognition systems when solving problems of assessing the state of the control system of a complex object.

**Keywords:** principal component analysis, graph, feature, centrality, importance, recognition, topological properties.

### **Введение**

Определение режимов и способов функционирования известных объектов (явлений) на-

блюдения зачастую носит трудноформализуемый характер в силу отсутствия достаточного количества данных об указанном объекте или неоднозначности их интерпретаций. Одним из

возможных вариантов решения данной задачи может служить оценивание состояния системы управления объекта исследования и соотнесение полученных результатов с вероятными режимами его функционирования. В этом случае, основным объектом наблюдения и анализа является информационно-телекоммуникационная сеть, которая лежит в основе исследуемой системы управления.

В настоящее время активно развивается научное направление анализа сетей, базирующееся на положениях теории графов. Это обусловлено развитием вычислительных мощностей ЭВМ, появлением языков высокого уровня и специального программного обеспечения, позволяющего эффективно исследовать сети. При этом под сетями подразумеваются сети различной природы: указанные информационно-коммуникационные сети (ИКС), сети взаимодействия белков в живых организмах, сети цитирования, социальные сети и др. Для исследования сетей к настоящему времени разработан достаточно богатый инструментарий, базирующийся на графовых моделях [1–3]. С помощью таких моделей становится возможным тестировать гипотезы о сетевых структурах, процессах их формирования и динамике эволюции, или иными словами — оценивать топологические свойства сети.

Топологические свойства применяются в разработке систем распознавания исследуемых сетей.

При распознавании одной из основных задач является отбор признаков, заключающийся в определении их важности, выявлении коллинеарных, что, в свою очередь, помогает лучшим образом понять структуру данных. Таким образом, актуальность работы обусловлена необходимостью разработки такого методического подхода, который позволит обосновать состав признаков для решения задачи определения топологических свойств информационно-телекоммуникационных сетей с целью оценивания состояния системы управления исследуемого объекта.

### Постановка задачи

Система управления описывается в виде графа ее системы связи, который представляет собой совокупность вершин, интерпретируемых как множество элементов системы управ-

ления, и ребер, показывающих связи между ними. В общем виде, при выборе в качестве объекта исследования какой-либо сети, в состав характеризующих ее признаков включают следующие характеристики, описанные в теории графов [4–7]: среднее расстояние пути (СРс), плотность (ПГ), ассортативность (АС), централизация по степени (СЦ), централизация по близости (СБ), централизация по посредничеству (СП), централизация по собственному вектору (СВЦ), авторитарность Кляйнберга (АЦ), диаметр (Ди), центральность Кляйнберга (КЦ). Данное подмножество можно представить в виде:

$$F = (f_{\text{СРс}}, f_{\text{ПГ}}, f_{\text{АС}}, f_{\text{Ди}}, f_{\text{СЦ}}, f_{\text{СБ}}, f_{\text{СП}}, f_{\text{СВЦ}}, f_{\text{КЦ}}, f_{\text{АЦ}}).$$

Элементы  $F$  формируют образ рассматриваемой модели графа или графа реальной сети в пространстве параметров. В результате расчета  $F$  для  $N$  реализаций графов, получаем матрицу наблюдений  $\mathbf{Y}_{[N,n]} = \|\mathbf{f}_{ij}\|_N^n$ , где  $n$  — номер характеристики графа.

Решение задачи определения состава наиболее важных признаков графа сети может быть построено с использованием метода главных компонент (МГК), который представляет собой совокупность статистических приемов обработки данных, позволяющих сконцентрировать информацию, содержащуюся в исходном массиве данных, за счет перехода к меньшему числу наиболее информативных факторов — главных компонент (ГК) [8].

Исходными данными являются стандартизированные значения  $f_{ij}$  из матрицы наблюдений  $\mathbf{Y}_{[N,n]}$ . Каждую строку матрицы можно представить, как реализацию  $n$ -мерного случайного вектора:

$$\hat{\mathbf{Y}}_{(n)} = \langle \hat{y}_1, \hat{y}_2, \dots, \hat{y}_n \rangle.$$

Свойства данного случайного вектора с достаточной для практики точностью описываются вектором математических ожиданий:

$$\bar{\mathbf{Y}}_{(n)} = \langle \bar{y}_1, \bar{y}_2, \dots, \bar{y}_n \rangle$$

и корреляционной матрицей, содержащей линейные коэффициенты корреляции Пирсона:

$$K_{[n]} = \|K_{ij}\|_n^n, \text{ где } K_{ij} = \frac{\sum (\hat{y}_i - \bar{y}_i)(\hat{y}_j - \bar{y}_j)}{\sqrt{\sum (\hat{y}_i - \bar{y}_i)^2 \sum (\hat{y}_j - \bar{y}_j)^2}}.$$

Метод ГК основывается на предположении, что любой  $j$ -й признак может быть представлен в виде линейной комбинации ГК  $f_i$ :

$$y_j = a_{1,j}f_1 + a_{2,j}f_2 + \dots + a_{n,j}f_n, [j = 1(1)n],$$

где  $f_1, f_2, \dots, f_n$  — главные компоненты;  
 $a_{ij}$  — вес  $i$ -й ГК в  $j$ -й переменной.

Задача анализа главных компонент сводится к тому, чтобы найти такое линейное ортогональное преобразование  $n$  наблюдаемых признаков, которое позволит получить совокупность  $n$  некоррелированных нормированных переменных  $f_i, [i = 1(1)n]$ , дисперсии  $\sigma^2$  которых обладают следующим свойством:

$$\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_n^2.$$

По дисперсии, соответствующей каждой ГК, можно оценить вклад этой компоненты в формирование общей дисперсии совокупности результатов расчета мер центральности и тем самым выделить существенные компоненты. При этом существуют следующие варианты использования результатов вычисления ГК:

- расчет вклада каждого наблюдаемого признака в ГК и их ранжирование по степени вклада;

- выявление мультиколлинеарных признаков, то есть имеющих тесную корреляционную взаимосвязь;

- разделение наблюдаемых признаков по группам (кластеризация).

Для расчета удельного вклада каждой меры центральности используется следующее выражение [8]:

$$Con_i = \frac{f_1(y_i)f_2(y_i)}{\sum_{k=1}^r f_1(y_k)f_2(y_k)}, [i = 1(1)N],$$

где  $f_1(y_i)$  и  $f_2(y_i)$  — значение вклада  $i$ -й меры в ГК1 и ГК2.

Таким образом, значение удельного вклада  $Con$  (от англ. «contribution») является произве-

дением вклада в ГК1 и ГК2  $i$ -й меры центральности, нормированной на сумму вкладов остальных мер в ГК1 и ГК2.

### Основные преобразования, осуществляемые в рамках методического подхода

На рис. 1 представлена структурная схема методического подхода по определению важности признаков различных моделей сетей. Она состоит из 3-х основных этапов, которые в свою очередь делятся на подэтапы. Рассмотрим их.

#### Этап подготовки данных

В настоящее время при решении задач машинного обучения и анализа данных около 30–40 % времени уходит на подготовку данных. Некорректно подготовленные данные могут существенно исказить результаты эксперимента.

1.1 В ходе преобработки данных заполняются пропущенные значения признаков, удаляются петли графа (при наличии), рис. 1.

1.2 Расчет признаков графов для каждого рассматриваемого графа (табл. 1). Полученные значения  $y_{ij}$  помещаются в матрицу наблюдений. При этом каждая строка матрицы соответствует одной исследуемой сети, рис. 1.

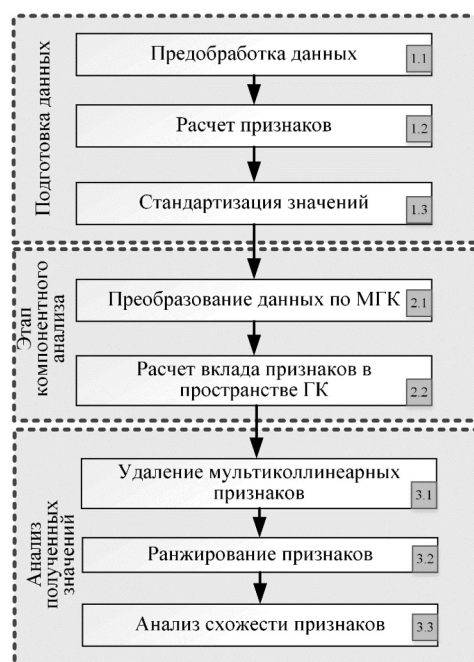


Рис. 1. Структурная схема методического подхода по определению важности признаков

Исходные данные для моделирования

Класс модели/Характеристика	Число вершин	Число ребер	Количество формируемых моделей графов
1 (БА)	100	переменно	50
2 (УС)	100	переменно	50
3 (ЭР)	100	300	50

1.3 Стандартизация значений. Результаты вычислений признаков имеют различную размерность, поэтому выполняется их нормировка по величине среднеквадратического отклонения для каждой меры, рис. 1.

### Преобразование данных МГК

Осуществляются математические преобразования, согласно представленного математического аппарата МГК. Производится расчет удельного вклада  $Con$ .

### Анализ полученных значений

Заключительный этап на котором осуществляется анализ полученных результатов.

3.1 На этом этапе осуществляется идентификация и удаление коллинеарных значений, что позволяет оптимизировать признаковое пространство, рис. 1.

3.2 На этапе ранжирования признаки упорядочиваются по их удельному вкладу  $Con$  в формирование главных компонент, рис.1.

3.3 На данном этапе на основе корреляционных окружностей осуществляется анализ схожести признаков, разделение их на группы, рис. 1.

### Апробация методического подхода

Апробацию будем осуществлять на базе математических моделей случайных графов Барабаши-Альберта (БА), Уоттса-Строгатца (УС) и Эрдеша-Реньи (ЭР), достаточно широко освещенных в литературе [2, 5, 8]. Исходные параметры моделей представлены в табл. 1.

Целью апробации методического подхода является демонстрация работоспособности МГК и влияния отбора признаков на оценивание топологических свойств сетей в задачах распознавания.

На этап 1.1 поступает 150 моделей графов (табл. 1), после расчетных этапов 1.2–1.3 формируется матрица наблюдений  $Y_{150,10}$ .

Далее над этой матрицей осуществляются преобразования 2.1–2.2, в ходе которых получают матрицы удельного вклада  $Con$  в ГК для признаков графов ( $n=10$ ) и для самих наблюдений ( $N=150$ ). Проведем анализ удельного вклада признаков на основе корреляционной окружности, представленной на рис. 2.

Корреляционная окружность — способ визуализации результатов анализа признаков методом главных компонент. Окружность имеет единичный радиус, что соответствует суммарному вкладу первой и второй компонент. Чем ближе к ней значение координат вектора, характеризующего тот или иной признак графа, тем выше его вклад в указанные компоненты. Схожие по степени вклада в ГК1 и ГК2 признаки расположены рядом.

Исходя из корреляционной окружности, представленной на рис. 2, можно сделать следующие выводы:

– признаки центральность и авторитарность Кляйнберга коллинеарны, видно, что они представлены одним вектором. Следовательно, один из них можно удалить, как дублирующий;

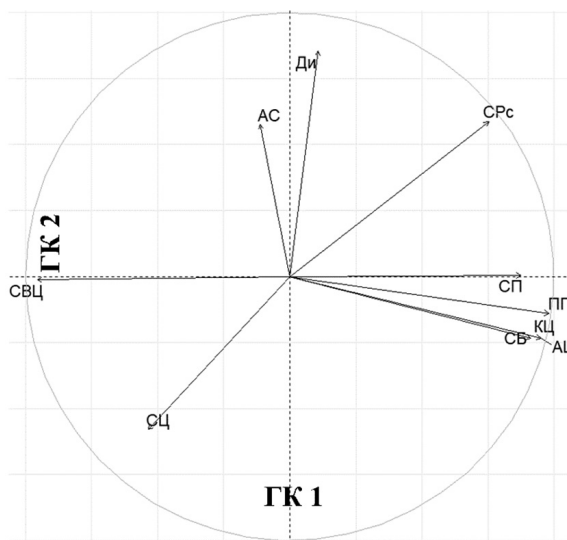


Рис. 2. Корреляционная окружность

– ассортативность и диаметр графа являются схожими характеристиками, о чем говорить соседнее расположение их векторов;

– уникальными, не имеющими соседей признаками, являются центральность по степени, собственному вектору, и среднее расстояние пути графа.

Корреляционная окружность позволяет визуально понять, какие признаки похожи друг на друга, и какие группы они создают. На рис. 3 представлена диаграмма, показывающая ранжированный вклад признаков в ГК1 и ГК2.

На рис. 3 видно, что наибольший вклад вносит плотность графа, следующая по вкладу — авторитарность Кляйнберга. Таким образом, получены ранжированные по важности значения признаков, оценивающих топологические свойства сети.

### Оценивание влияния важности признаков в задачах распознавания на примере решения задачи кластеризации

В разделе апробации методического подхода была сформирована матрица наблюдений, состоящая из 3-х различных по своим топологическим свойствам типам сетей. Эти сети представлены математическими моделями Барабаши-Альберта, Уоттса-Строгатца и Эрдеша-Реньи имеющими разный алгоритм формирования. Таким образом, при решении задачи кластеризации теоретически возможно разделить их на классы. Но неясно — какие же признаки позволят наилучшим образом разделить смесь трех классов моделей. Воспользуемся

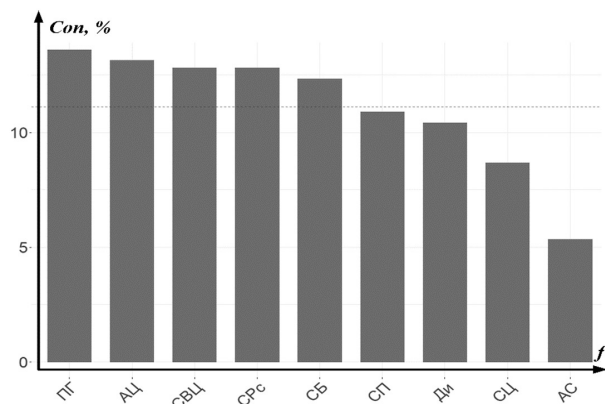


Рис. 3. Диаграмма вклада признаков в главные компоненты

данными анализа на основе МГК, представленными на рис. 2 и 3.

Итак, проанализируем 150 наблюдений трех моделей, представленных после преобразования по МГК в пространстве ГК1 и ГК2. При этом рассмотрим три варианта использования признаков:

– 1 вариант. Наблюдения представлены двумя признаками с наименьшим вкладом, то есть СЦ и АС;

– 2 вариант. Наблюдения представлены девятью признаками (удален признак АЦ, ввиду его коллинеарности КЦ), то есть СРс, ПГ, АС, Ди, СЦ, СБ, СП, СВЦ, КЦ;

– 3 вариант. Наблюдения представлены двумя признаками с наибольшим вкладом, то есть ПГ и АЦ.

Результаты наблюдений представлены в пространстве первой и второй главной компонент по вариантам на рис. 4, а, б, в.

Исходя из визуального анализа рис. 4 можно сделать вывод о том, какое влияние оказывает на разрешающую способность использование различных комбинаций признаков. Наилучшая способность к разделению данных достигается путем использования наиболее важных признаков, худшая — наименее важных.

Помимо визуального анализа влияния выбора признаков осуществим анализ путем измерения статистики Хопкинса ( $H$ ) [9], которая является одним из индикаторов тенденции данных к группированию. Значения статистики  $H \geq 0,5$  говорят о том, что данные распределены случайно и равномерно. Значения статистики  $H \leq 0,25$  на 90 % в доверительном интервале указывает на имеющуюся тенденцию к группированию данных.

Кроме измерения статистики Хопкинса осуществим иерархическую кластеризацию данных для трех классов [9] и измерим качество кластеризации на основе коэффициента Фулкса-Мэллова ( $T$ ) [10]. Коэффициент Фулкса-Мэллова — это мера сходства между двумя результатами кластеризации (в этом исследовании). Между истинными значениями классов, полученных на этапе моделирования (реальных данных) и значениями классов, полученных в результате кластеризации. Результаты измерений представлены в табл. 2.

Представленные в табл. 2 результаты свидетельствуют о влиянии выбора признака на

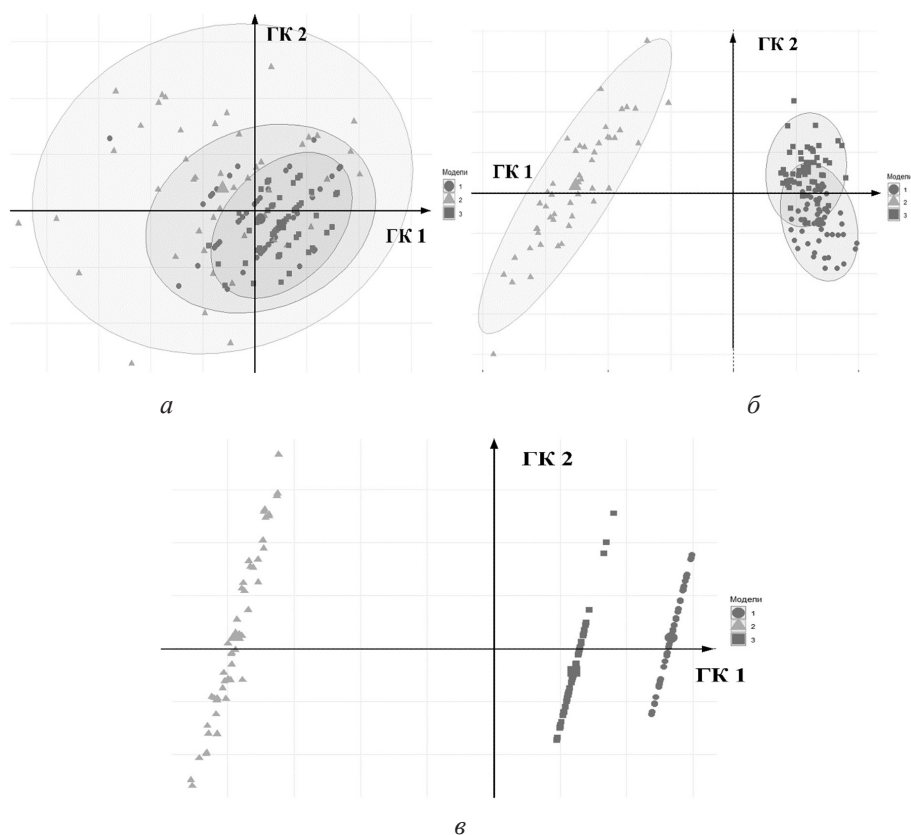


Рис. 4. Представление наблюдений в пространстве GK1 и GK2:  
а — вариант 1; б — вариант 2; в — вариант 3

Таблица 2

**Оценка результатов кластеризации и тенденции данных к группированию**

Варианты использования признаков	$H$	$T$
1 вариант	0,24	0,40
2 вариант	0,11	0,69
3 вариант	0,03	1

результаты кластеризации, а равно и иной задачи распознавания, например, классификации и регрессии. Полученные результаты для варианта 3, когда выбраны два наиболее важных признака, показывают, что данные четко сгруппированы и кластеризация при таком их расположении происходит без ошибок. Использование наименее важных признаков приводит к низкой группировке данных и высоким ошибкам кластеризации.

**Выводы**

Данное исследование показало, что при решении широкого спектра задач оценивания топологических свойств и распознавании структур

сетей, важным аспектом является выбор признакового пространства. Признаковое пространство выбирается на основе анализа важности признаков, в результате которого удаляются коллинеарные, а остальные ранжируются на основе их удельного вклада. Проведенный натурный эксперимент по кластеризации графов подтвердил теоретические результаты анализа признаков, что подтверждает работоспособность методического подхода.

Разработанный методический подход может быть использован в ходе решения задач распознавания структур информационно-телекоммуникационных сетей с целью оценивания состояния системы управления объекта исследования.

Литература

1. Chen P-Y., Choudhury S., Hero A. Multi-centrality graph spectral decompositions and their application to cyber intrusion detection // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016. doi: 10.1109/ICASSP.2016.7472539.
2. Newman M.E.J. Networks an introduction. — N.Y.: Oxford University Press Inc. 2010. 1042 p.
3. Шуваев Ф.Л., Татарка М.В. Анализ динамики мер центральности математических моделей случайных графов // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 2. С. 249–256. doi: 10.17586/2226-1494-2020-20-2-249-256
4. Hartmann A., Mézard M. Distribution of diameters for Erdős-Rényi random graphs // Phys. Rev. 2018. Vol. 97. № 3. doi: 10.1103/PhysRevE.97.032128
5. Gibson H., Vickers P. Using adjacency matrices to lay out larger small-world networks // Applied soft computing. 2016. Vol. 42. P. 80–92.
6. Minoo A., Salehzadeh-Yazdi A., Razaghi-Moghadam Z., Hennig H., Wolkenhauer O. A Systematic Survey of Centrality Measures for Protein-Protein Interaction Networks, Systems Biology. 2018.
7. Шуваев Ф.Л., Татарка М.В. Анализ математических моделей случайных графов, применяемых в имитационном моделировании информационно-коммуникационных сетей // Вестник Санкт-Петербургского университета ГПС МЧС России. 2020. № 2. С. 67–77.
8. Еремеев И.Ю., Татарка М.В., Шуваев Ф.Л., Цыганов А.С. Анализ мер центральности узлов сетей на основе метода главных компонент // Информатика и автоматизация. 2020. № 19 (6). С. 1307–1331. doi: org/10.15622/ia.2020.19.6
9. Kassambara A. Practical guide to cluster analysis in R // STDHA. 2017. 187 p.
10. Depaolini M., Ciucci D., Calegari S., Dominoni M. External Indices for Rough Clustering. Lecture Notes in Computer Science. 2018. Vol. 11103.

References

1. Chen P-Y., Choudhury S., Hero A. Multi-centrality graph spectral decompositions and their application to cyber intrusion detection // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016. doi: 10.1109/ICASSP.2016.7472539
2. Newman M.E.J. Networks an introduction. — N.Y.: Oxford University Press Inc. 2010. 1042 p.
3. Shuvaev F.L., Tatarka M.V. Analysis of the dynamics of measures of centrality of mathematical models of random graphs. Scientific and technical bulletin of information technologies, mechanics and optics. 2020. Vol. 20. № 2. P. 249–256. doi: 10.17586/2226-1494-2020-20-2-249-256.
4. Hartmann A., Mézard M. Distribution of diameters for Erdős-Rényi random graphs // Phys. Rev. 2018. Vol. 97. № 3. doi: 10.1103/PhysRevE.97.032128.
5. Gibson H., Vickers P. Using adjacency matrices to lay out larger small-world networks // Applied soft computing. 2016. Vol. 42. P. 80–92.
6. Minoo A., Salehzadeh-Yazdi A., Razaghi-Moghadam Z., Hennig H., Wolkenhauer O. A Systematic Survey of Centrality Measures for Protein-Protein Interaction Networks, Systems Biology. 2018.
7. Shuvaev F.L., Tatarka M.V. Analysis of mathematical models of random graphs used in the simulation of information and communication networks. Bulletin of St. Petersburg University State Fire Service of the Ministry of Emergencies of Russia. 2020. № 2. P. 67–77.
8. Eremeev I., Tatarka M., Shuvaev F., Cyganov A. Comparative analysis of centrality measures of network nodes based on principal component analysis. Informatics and automation. 2020. Vol. 19 (6). P. 1307–1331. doi: org/10.15622/ia.2020.19.6.
9. Kassambara A. Practical guide to cluster analysis in R // STDHA. 2017. 187 p.
10. Depaolini M., Ciucci D., Calegari S., Dominoni M. External Indices for Rough Clustering. Lecture Notes in Computer Science. 2018. Vol. 11103.