



Определение оптимального набора метрик подобия графов в задачах распознавания топологий техногенных космических объектов

Н. В. Груздев^а, канд. воен. наук, доцент, orcid.org/0009-0000-6125-1362, NikGruzdev@rambler.ru

Ф. Л. Шуваев^а, канд. техн. наук, начальник лаборатории, orcid.org/0000-0001-7502-9455

^аВоенно-космическая академия им. А. Ф. Можайского, ул. Ждановская, 13, Санкт-Петербург, 197198, РФ

Введение: в условиях техногенного засорения околоземного космического пространства актуальной остается задача анализа результатов наблюдений за техногенными космическими объектами в целях выявления однотипных по топологии групп объектов. Для ее решения предлагается использовать инструментарий теории графов и теории распознавания для оценивания топологий техногенных космических объектов различного типа. **Цель:** выполнить сравнительный анализ метрик подобия графов, заключающийся в поиске оптимальной метрики в рамках показателей качества распознавания техногенных космических объектов. **Результаты:** проведенные исследования показали, что существует три основных подхода к расчету метрик подобия графов. Каждый из подходов основан на вычислении вектора параметров, характеризующих граф с различных сторон. Разработаны метрики, позволяющие различать группы техногенных космических объектов, представленных в виде графов. Вычислительный эксперимент показал, что оптимальный по критерию качества кластеризации состав признакового пространства отличается для различных типов графовых моделей. Это свойство играет важную роль при решении задач распознавания топологии техногенных космических объектов на новом качественном уровне. Применяя предложенные методы сравнительного анализа метрик подобия графов, возможно выбрать наилучшую метрику для распознавания топологии техногенных космических объектов. **Практическая значимость:** использование предлагаемых методов анализа метрик подобия графов позволит оценить изменения в группах однотипных техногенных космических объектов. Предложенные методы могут использоваться при решении задач контроля и каталогизации техногенных космических объектов.

Ключевые слова – техногенные космические объекты, граф, метрики подобия графов, топология, мера сходства, кластеризация.

Для цитирования: Груздев Н. В., Шуваев Ф. Л. Определение оптимального набора метрик подобия графов в задачах распознавания топологий техногенных космических объектов. *Информационно-управляющие системы*, 2024, № 1, с. 9–19. doi:10.31799/1684-8853-2024-1-9-19, EDN: OWCDGM

For citation: Gruzdev N. V., Shuvaev F. L. Determination of the optimal set of graph similarity metrics in problems of recognition of topologies of technogenic space objects. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2024, no. 1, pp. 9–19 (In Russian). doi:10.31799/1684-8853-2024-1-9-19, EDN: OWCDGM

Введение

В настоящее время активно увеличивается число областей, где находит практическое применение теория графов для исследования сложных объектов или групп объектов, связанных в определенную структуру.

Теория графов интересна тем, что позволяет наглядно отображать и исследовать различные природные, техногенные и социальные процессы. Она позволяет представить сложные объекты исследования (сети связи, сети электроснабжения, космические объекты, атомы молекул, социальные сети) в виде графа (G) – совокупности вершин (V , т. е. простых объектов) и ребер (E , связей между ними). Типы взаимосвязей, образующих граф, неслучайны. Они отражают лежащие в основе сложных сетевых объектов процессы, которые можно исследовать с помощью различных разделов теории графов. Основные направления анализа сетей включают:

- визуализацию и формализацию сетей с использованием аппарата теории графов;
- оценивание структуры сети в целом, а также оценивание параметров структурных элементов – отдельных подсетей и узлов, связей между ними;
- создание математических и статистических моделей сетевых структур и сетевой динамики.

Техногенные космические объекты (отделяемые части ракет космического назначения, разгонные блоки, космические аппараты) представляются в виде графа. В работе [1] предложено решающее правило для выбора наиболее опасного (с точки зрения вероятного столкновения) техногенного объекта из множества наблюдаемых по результатам измерений их частных признаков. В целях выявления изменений, происходящих в группе наблюдаемых опасных техногенных космических объектов, возникает задача анализа сетевой структуры наблюдаемой группы объ-

ектов. Изменения в сетевой структуре техногенных космических объектов, проявляющиеся либо в результате визуализации, либо в результате структурного анализа, зачастую не видны человеку, но легко выявляются машинным зрением, и интерпретация полученных результатов дает понимание основных закономерностей, имеющих место в наблюдаемой системе, и причин их возникновения [2–4].

Авторами [5] выделяется два типичных варианта трансформации динамических графов [6, 7]. По первому варианту со временем изменяется число вершин, что соответствует изменению структуры исследуемой сети, что характерно для развивающихся, эволюционирующих или деградирующих сетей. По второму варианту число вершин графа неизменно, однако изменяется характер связей между ними, что характерно для устоявшихся систем, для систем с ограниченным числом элементов. Наибольший интерес в прикладном смысле имеет второй вариант, отвечающий таким направлениям, как исследование вариативности социальных связей в ограниченном коллективе, исследование процессов самоорганизации робототехнических систем с ограниченным числом объектов, исследование космических объектов и др. При таком варианте анализируется стационарная система (стационарная в смысле фиксированного числа вершин графа), в которой в равные интервалы времени ребра между вершинами меняются, тем самым изменяя топологию графа.

Исследование таких графовых структур состоит в выявлении условий образования или исчезновения ребер и распознавании тенденций трансформации связей между вершинами. В этом и заключается актуальность задачи распознавания графов по их топологической структуре. Важнейшим этапом задачи распознавания является определение оптимального признакового пространства; применительно к данному исследованию признаковым пространством являются метрики топологической близости графов.

С учетом ограниченности числа вершин задача упорядочивания графов по их топологической структуре является конечномерной. Однако с ростом числа вершин число топологических структур растет катастрофически [8], что требует для их перечисления огромных вычислительных мощностей [9]. В настоящее время известен целый ряд подходов к решению задачи упорядочивания графов по их топологической структуре на основе сокращения размерности параметрического описания графовой структуры с использованием различных метрик [10–25].

Наличие различных методов и метрик для оценивания топологической близости графов порождает задачу их сравнительного анализа

с целью выбрать оптимальное сочетание по критерию «качество кластеризации» в конкретных условиях. Решению этой задачи посвящено данное исследование.

Основные этапы сравнительного анализа

В основе сравнительного анализа метрик топологической близости графов лежат следующие базовые положения:

1) задача кластеризации графов различной природы может быть решена только в условиях различимости графовых структур;

2) для порождения различных по природе возникновения графовых структур возможно использовать модели случайных графов различных классов [11–13, 25–27];

3) в исследовании используются метрики топологической близости, основанные на расчете триадного вектора, интегральных характеристиках графов и топологических индексах. Все метрики широко освещены в научной литературе и опубликованы в работах последнего времени [14–25];

4) для метрик, в состав которых входят неортогональные компоненты, необходимо провести предварительный анализ информативности компонент для сокращения признакового пространства;

5) чтобы оценить результаты анализа информативности компонент, необходимо сравнить их с реальными показателями, а для этого требуется перебрать все возможные комбинации признаков и рассчитать качество кластеризации для них.

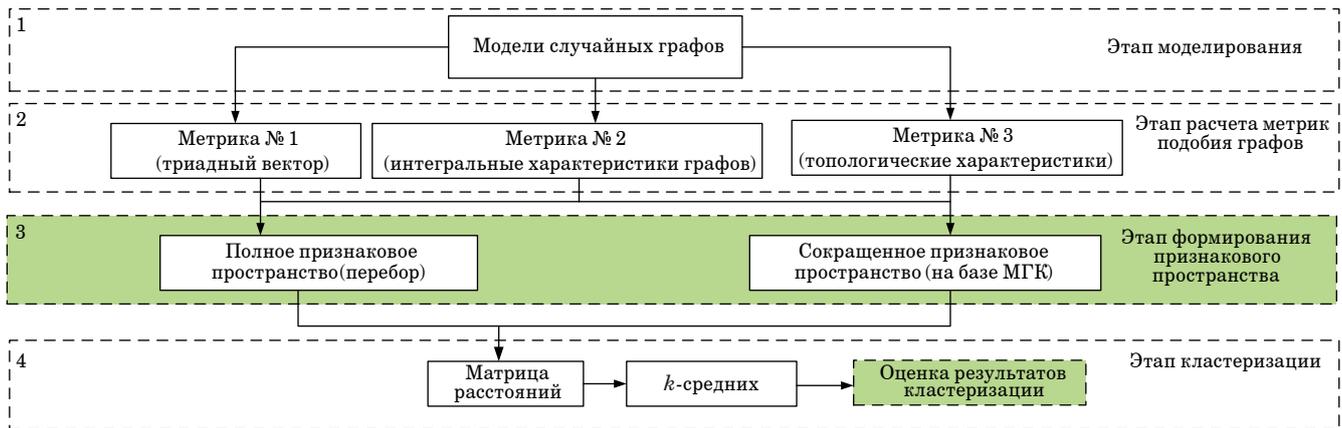
Структурная схема проводимого в рамках исследования сравнительного анализа метрик подобия графов изображена на рис. 1. Каждый из четырех основных этапов подразделяется на вспомогательные подэтапы.

Этап моделирования

Этап моделирования предназначен для формирования матриц смежности случайных графов. В настоящее время в науке наиболее активно применяемыми считаются модели случайного графа Эрдеша – Реньи (ЭР), модель Уоттса – Строгатца (УС) и модель Барабаши – Альберта (БА) [11–13, 25, 26]. Модель УС является модификацией модели ЭР, поэтому в дальнейшем при проведении эксперимента ограничимся моделями БА и ЭР. Алгоритм построения моделей подробно описан в работе [25].

Выбор и обоснование параметров моделей.

При формировании исходных данных для исследования различных метрик схожести графов



■ **Рис. 1.** Структурная схема исследования
 ■ **Fig. 1.** Block diagram of the study

удобно использовать модели случайных графов, порождающие непересекающиеся классы графов для распознавания. Для этого необходимо рассчитать вероятность изоморфизма графов $P_{из}$, порождаемых различными математическими моделями. Логично предположить, что вероятность появления изоморфных друг другу графов $P_{из}$, порождаемых моделями разных классов, уменьшается при увеличении числа вершин.

В рамках этого предположения сформированы по 10 000 реализаций случайных графов для каждой из рассматриваемых моделей (табл. 1) и проведено попарное сравнение графов на изоморфизм внутри каждого класса и между классами.

При этом оценка вероятности появления изоморфных графов рассчитывается как $P_{из} = \frac{n_{из}}{n_{общ}}$, где $n_{из}$ – число пар изоморфных графов; $n_{общ}$ – общее число пар графов. Для 500 графов получим $n_{общ} = 124\ 750$. Полученные результаты свидетельствуют, что для графов

с числом вершин 12 и более величина $P_{из} \approx 0$ при значении доверительной вероятности на уровне 0,95. На этом основании выберем число вершин для формирования случайных графов не менее 12, например 60. Чтобы результаты эксперимента были адекватными, параметр «число ребер» моделей подобран в равные интервалы. Для этого измерены граничные значения числа ребер для моделей БА.

Вычислительный эксперимент был разделен на два этапа: 1-й этап – проведение кластеризации для моделей БА и ЭР; 2-й этап – проведение кластеризации для двух моделей БА. Таким образом, на каждом этапе эксперимента проводилась кластеризация для двух классов. На каждом этапе формировалось по 10 000 графов каждого класса, при этом на этапе 1 параметры модели 1 « p_1, p_2, p_3 » задавались случайно с равномерным распределением в интервале от 0,1 до 0,9 с шагом 0,1, а параметр модели 2 «число ребер» задавался случайно в интервале от 100 до 200.

На этапе 2 параметр модели 3 « $p_2 = 0,1$ » был фиксирован, параметры « p_3 » и « p_4 » задавались случайно с равномерным распределением в интервале от 0,1 до 0,9 с шагом 0,1. Параметр модели 4 « $p_2 = 0,9$ » был фиксирован, « p_3 » и « p_4 » задавались аналогично модели 3.

Модели ЭР подобны социальным структурам с низкой централизацией, отсутствием узлов с высокой центральностью, а модели БА, напротив, формируют графы с низкой централизацией и строгой иерархической структурой, тем самым они подобны социальным структурам с высокой организацией. Изучение возможности распознавания сетей с низкой и высокой централизацией осуществляется на этапе 1, а распознавание сетей с высокой централизацией, но разными формами ее порождения, – на этапе 2.

■ **Таблица 1.** Параметры моделей графов
 ■ **Table 1.** Parameters of graph models

| Этап | Номер модели | Тип модели | Число вершин | Число ребер | Дополнительные параметры |
|------|--------------|------------|--------------|-------------|---|
| 1 | 1 | БА | 60 | 76:117 | $p_{2,3} = (0,1:0,9)$ с шагом 0,1 |
| | 2 | ЭР | 60 | 76:117 | – |
| 2 | 3 | БА | 60 | 76:117 | $p_2 = 0,1$ $p_3 = (0,1:0,9)$ $p_4 = (0,1:0,9)$ |
| | 4 | БА | 60 | 59:102 | $p_2 = 0,1$ $p_3 = (0,1:0,9)$ $p_4 = (0,1:0,9)$ |

Этап расчета метрик подобия графов

На данном этапе рассчитываются метрики подобия, проявляющие различие или близость топологической структуры графов. Метрики разделены на три группы.

Метрика № 1. Триадный вектор.

В основе метрики № 1 лежит утверждение, что минимально различимой структурой графа является диада: две вершины, между которыми либо есть, либо нет ребра изучаемого типа, т. е. наличие или отсутствие связи между вершинами. Анализ публикаций в области теории графов за последние пять лет показал, что наиболее предпочтительной минимальной единицей в социальной группе является триада [14–17]. Триада – это граф, состоящий из трех вершин. Комбинации ребер между вершинами в триаде определяются 16 возможными вариантами построения, количество классов таких комбинаций называется триадным вектором.

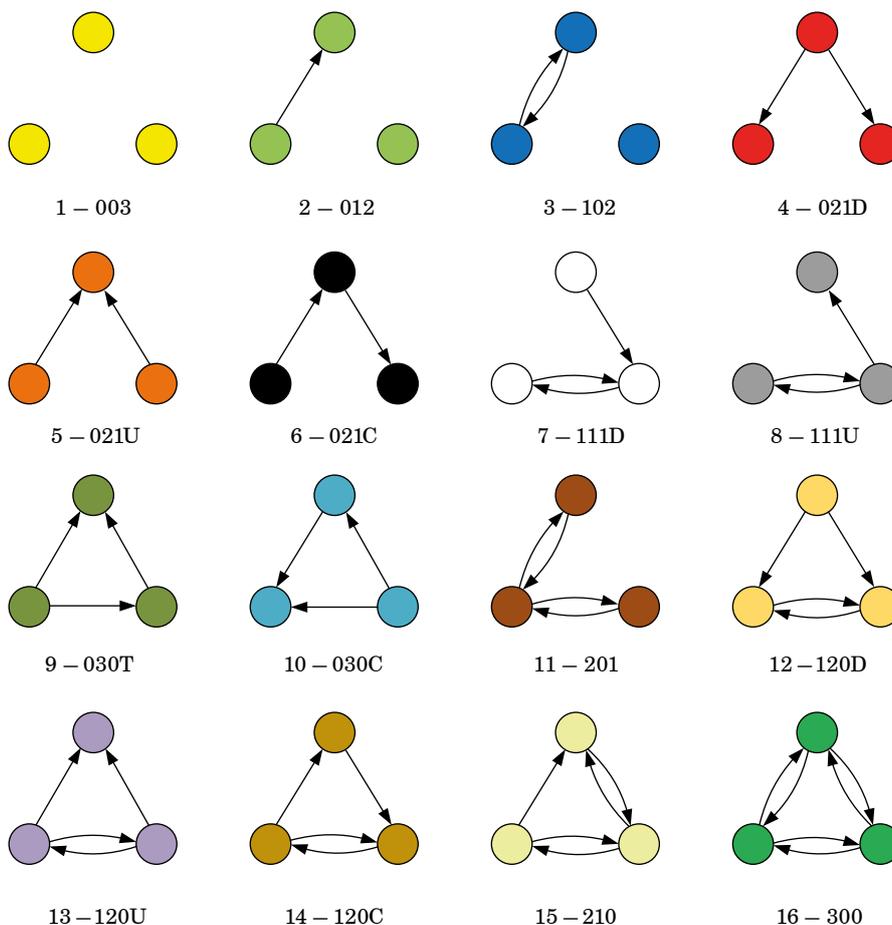
Рассчитав триадные векторы для графов, исследователь получает метрику или пространство признаков, на основе которого рассчитывается

расстояние между графами и определяется степень их подобия. Таким образом, триадный вектор может быть использован как метрика расстояния между графами. На рис. 2 [14] показаны триады и соответствующие обозначения.

Для различения комбинаций употребляются символы латинского алфавита, например: «D» – вниз (асимметричная связь или связи направлены вниз); «U» – вверх (асимметричная связь или связи направлены вверх); «C» – триада циклическая; «T» – триада транзитивная [14–17].

Метрика № 2. Интегральные характеристики графов.

В теории графов имеются различные характеристики, описывающие те или иные свойства графов, среди них основные: среднее расстояние пути (СРс), плотность графа (ПГ), ассортативность (АС), централизация по степени (ЦС), централизация по близости (ЦБ), централизация по посредничеству (ЦП), централизация по собственному вектору (СВЦ), централизация Кляйнберга (КЦ), централизация по авторитетности (АЦ) [10, 22–25]. Рассчитав для каждого



■ **Рис. 2.** Комбинации триад [14]
 ■ **Fig. 2.** Combinations of triads [14]

графа набор значений таких параметров, возможно перейти к матрице расстояний между ними, т. е. представить граф в признаковом пространстве, описанном образами – интегральными характеристиками графов.

Метрика № 3. Топологические индексы.

Топологический индекс – некоторое обычно числовое значение или упорядоченный набор значений, характеризующих структуру графа. Широкое применение топологические индексы получили в хромометрии для исследования химических свойств веществ. Для исследования были выбраны три наиболее известных топологических индекса: Винера (Ви), сумм строк (ССт), эффективной сложности (ЭС) [19–21].

Полученные значения параметров y_{ij} помещаются в матрицу данных, где i – номер метрики; j – номер графа. Таким образом, ячейки матрицы заполняют N метрик для n графов:

$$Y_{[N,n]} = \|y_{ij}\|_N^n.$$

Этап формирования признакового пространства

В нашем исследовании рассматривается два варианта формирования признакового пространства: 1) полное признаковое пространство, при котором перебираются все комбинации признаков и для каждой оценивается качество кластеризации; 2) сокращенное признаковое пространство, причем сокращение пространства осуществляется за счет выявления наиболее информативных признаков на основе компонентного анализа.

Компонентный анализ.

Компонентный анализ в рамках метода главных компонент (МГК; Principal Component Analysis – PCA) представляет собой совокупность статистических приемов обработки данных, которые позволяют сконцентрировать содержащуюся в исходном массиве данных информацию за счет перехода к меньшему числу наиболее информативных факторов – главных компонент (ГК) [23].

Для каждого из исследуемых графов осуществляется расчет вклада различных характеристик графов в пространстве ГК. Для расчета удельного вклада каждой интегральной характеристики используется следующее выражение [23]:

$$Con_j = \frac{f_1(y_j)f_2(y_j)}{\sum_{k=1}^I f_1(y_k)f_2(y_k)}, [j = 1(1)n],$$

где $f_1(y_j)$ и $f_2(y_j)$ – значение вклада j -го признака в ГК1 и ГК2.

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

вклада в ГК1 и ГК2 j -го признака графа, нормированного на сумму вкладов остальных мер в ГК1 и ГК2. На основе удельного вклада возможно выбирать наиболее информативные признаки, позволяющие сокращать признаковое пространство и проводить кластеризацию с наивысшим качеством.

Этап кластеризации

Процедура кластеризации в исследовании используется для оценивания разделяющей способности сравниваемых метрик подобия графов. Учитывая, что цель исследования состоит в сравнении различных метрик подобия графов, для решения задачи кластеризации графов, порождаемых различными исследуемыми моделями, можно использовать алгоритм кластеризации с известным числом классов по методу k -средних [28].

Метод k -средних, один из самых распространенных методов кластеризации, был предложен в 1950-х годах математиком Гуго Штейнгаузом. Суть метода заключается в минимизации S – суммарного квадратичного отклонения координат элементов кластеров от центров этих кластеров – на основе итеративной процедуры пересчета координат центров кластеров и состава кластеров [28].

Этап оценки результатов кластеризации

Для оценки результатов кластеризации был выбран коэффициент Фоулкса – Мэллова (КФМ) [25, 29] – мера сходства между двумя результатами кластеризации (в настоящем исследовании между истинными значениями классов $A1$ и значениями классов $A2$, полученными в результате кластеризации). Коэффициент изменяется в пределах $[0; 1]$. Чем его значение выше, тем выше качество кластеризации [23].

Результаты экспериментальных исследований

Анализ методом ГК метрик подобия графов показал, что в метрике № 2 коллинеарными являются характеристики АЦ и КЦ, поэтому исходное количество признаков сокращено за счет удаления КЦ. При вычислении метрики № 1 было установлено, что триады с шестой по 16 в рассматриваемых моделях графов отсутствуют, поэтому в метрике № 1 количество признаков сокращено с 16 до пяти. Таким образом, матрица наблюдений приобретает вид $Y_{[16,20000]}$. Итоговое число признаков: метрика № 1 – 5, метрика № 2 – 8, метрика № 3 – 3, итого 16.

Качество кластеризации в ходе эксперимента рассчитывалось для комбинаций из 1:16 признаков с целью выявить комбинацию, позволяющую разделить смеси моделей наилучшим образом. При этом наилучшей считается при-

знак или комбинация признаков, обладающая самым высоким значением КФМ. Вычисления осуществлялись на сервере HPProliantDL380, на базе двух процессоров INTEL Xeon silver 4214. Моделирование осуществлялось средствами языка R, распараллеливание работы процессоров обеспечивалось библиотекой doParallel [30].

1-й этап (рис. 3, а, б). Корреляционная окружность представляет собой вариант визуализации результатов компонентного анализа. Радиус корреляционной окружности равен единице и интерпретирует суммарный вклад первой и второй компонент. Ближайшие к радиусу векторы признаков соответствуют наибольшему вкладу в первую и вторую компоненты. Чем ближе расположены друг к другу признаки (обозначены квадратом красной пунктирной линией на корреляционной окружности), тем больше у них совместный вклад в ГК1 и ГК2.

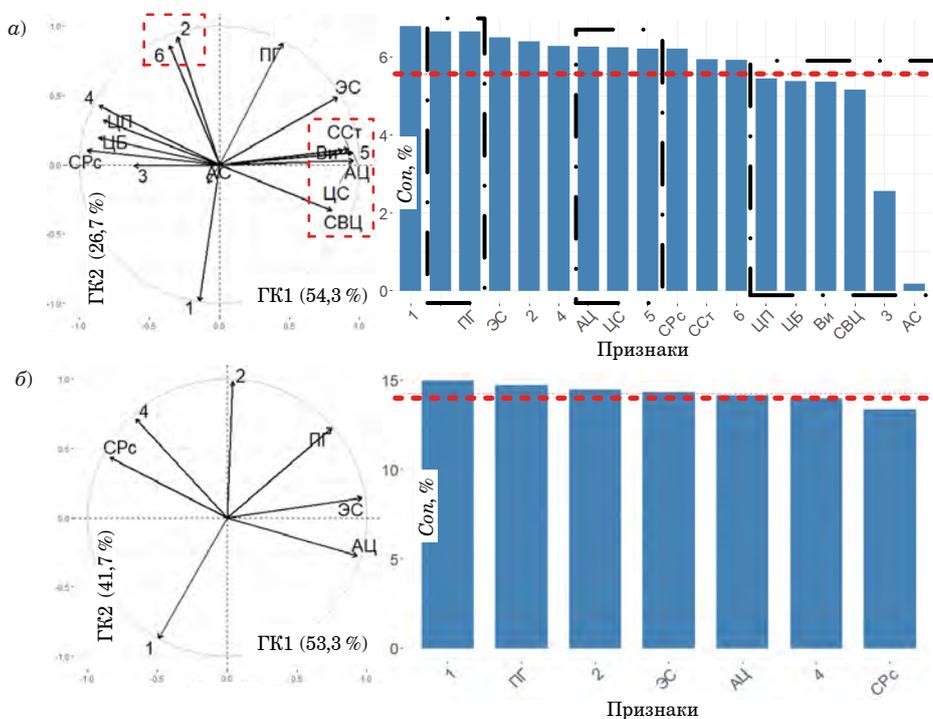
Красная пунктирная, проведенная через признаки, — это средний удельный вклад в ГК1 и ГК2. Рассчитывается как деление всего вклада (100 %) на количество признаков. В соответствии с методикой применения МГК признаковое пространство сокращается за счет удаления коллинеарных признаков, удаления признаков с *Con* ниже среднего значения (выделены на признаках рис. 3, а черной штрихпунктирной линией) и перколяции (т. е. удаления из группы одно-

типных признаков тех, у которых ниже значение *Con*), сходных по степени вклада признаков.

На рис. 3, б представлено сокращенное в соответствии с этой методикой до семи признаков пространство. Корреляционная окружность рис. 3, б показывает, что суммарный вклад в ГК1 и ГК2 вырос до 95 % (41,7 + 53,3) по сравнению с 81 % (54,3 + 26,7) на корреляционной окружности рис. 3, а. Это позитивное наблюдение подчеркивает правильность выбранных шагов по сокращению признаков.

Результаты сравнения качества кластеризации с использованием признакового пространства, выбранного на основе компонентного анализа и метода перебора, представлены в табл. 2. Обозначение столбцов таблицы: 1 — количество признаков; 2 — тип признакового пространства; 3 — количество комбинаций признаков; 4 — количество комбинаций с наилучшим КФМ; 5 — наилучший КФМ; 6 — наихудший КФМ; 7 — средний КФМ; 8 — лучшая по МГК комбинация; 9 — КФМ лучшей по МГК комбинации.

Анализ результатов, представленных в табл. 2, сделал возможным выявить характерную особенность, заключающуюся в том, что на этапе 1 при числе признаков $N = \{2:7\}$ при полном переборе КФМ = 1. Для $N = 2$ это комбинации ЦС-ЭС и ЦП-ЭС, при $N = 7$ это комбинация ПГ-АС-ЦС-СВЦ-ЦБ-ЦП-ЭС. При $N = 3, 4, 5, 6$ признаки не выходят



■ **Рис. 3.** Результаты применения МГК на этапе 1 до (а) и после (б) сокращения признакового пространства: слева — корреляционная окружность; справа — ранжированные по значению удельного вклада признаки
 ■ **Fig. 3.** The results of the use of PCA at stage 1 before (a) and after (б) the reduction of the feature space: left — correlation circle; on right — features ranked by the value of the specific contribution

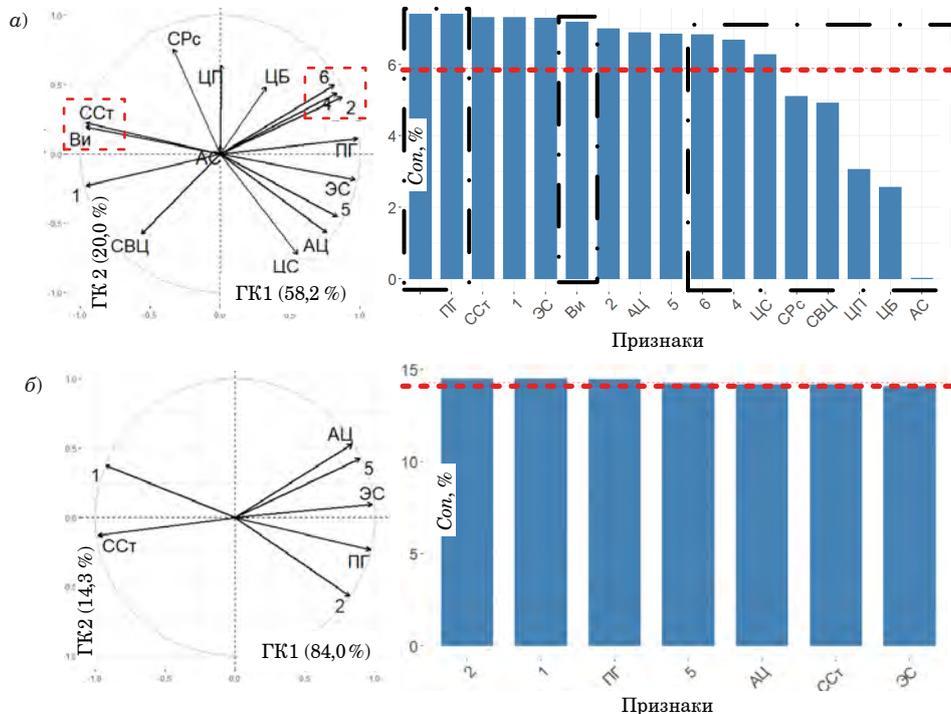
■ **Таблица 2.** Результаты вычислительного эксперимента на этапе 1
 ■ **Table 2.** Results of the computational experiment at stage 1

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-------------|--------|----|------|------|------|-------------------|------|
| 1 | Сокращенное | 7 | 1 | 0,95 | 0,50 | 0,73 | 1 | 0,56 |
| | Полное | 16 | 1 | 0,95 | 0,50 | 0,77 | | |
| 2 | Сокращенное | 21 | 2 | 0,95 | 0,49 | 0,70 | 1-ПГ | 0,58 |
| | Полное | 120 | 2 | 1 | 0,49 | 0,78 | | |
| 3 | Сокращенное | 35 | 1 | 0,95 | 0,49 | 0,63 | 1-ПГ-2 | 0,56 |
| | Полное | 560 | 9 | 1 | 0,49 | 0,77 | | |
| 4 | Сокращенное | 35 | 4 | 0,93 | 0,49 | 0,58 | 1-ПГ-2-5 | 0,56 |
| | Полное | 1820 | 15 | 1 | 0,49 | 0,76 | | |
| 5 | Сокращенное | 21 | 1 | 0,93 | 0,49 | 0,54 | 1-ПГ-2-5-ЭС | 0,61 |
| | Полное | 4368 | 14 | 1 | 0,49 | 0,75 | | |
| 6 | Сокращенное | 7 | 1 | 0,55 | 0,49 | 0,51 | 1-ПГ-2-5-ЭС-4 | 0,55 |
| | Полное | 8008 | 6 | 1 | 0,49 | 0,75 | | |
| 7 | Сокращенное | 1 | 1 | 0,49 | 0,49 | 0,49 | 1-ПГ-2-5-ЭС-4-СРс | 0,49 |
| | Полное | 11 440 | 1 | 1 | 0,49 | 0,75 | | |

из границ множества ПГ-АС-ЦС-СВЦ-ЦБ-ЦП-ЭС. Видны граничные пределы по распознаванию графов, выраженные в КФМ: нижний – 0,49; верхний – 1. Из столбцов 8 и 9 видно, что комбинации, определенные по МГК, не являются оптимальны-

ми по критерию КФМ. Особенностью является также и то, что увеличение размерности признакового пространства не приводит к повышению КФМ.

2-й этап (рис. 4, а, б). На корреляционной окружности рис. 4, а видно, что тенденции по



■ **Рис. 4.** Результаты применения МГК на этапе 2 до (а) и после (б) сокращения признакового пространства: слева – корреляционная окружность; справа – ранжированные по значению удельного вклада признаки
 ■ **Fig. 4.** The results of the use of PCA at stage 2 before (a) and after (b) the reduction of the feature space: left – correlation circle; on right – features ranked by the value of the specific contribution

■ **Таблица 3.** Результаты вычислительного эксперимента на этапе 2

■ **Table 3.** Results of the computational experiment at stage 2

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-------------|--------|-----|------|------|------|--------------------|------|
| 1 | Сокращенное | 7 | 1 | 0,90 | 0,57 | 0,77 | 1 | 0,88 |
| | Полное | 16 | 1 | 0,90 | 0,50 | 0,68 | | |
| 2 | Сокращенное | 21 | 1 | 0,89 | 0,57 | 0,79 | 1-ПГ | 0,88 |
| | Полное | 120 | 1 | 0,90 | 0,49 | 0,72 | | |
| 3 | Сокращенное | 35 | 4 | 0,89 | 0,70 | 0,82 | 1-ПГ-2 | 0,88 |
| | Полное | 560 | 12 | 0,89 | 0,49 | 0,75 | | |
| 4 | Сокращенное | 35 | 5 | 0,89 | 0,70 | 0,84 | 1-ПГ-2-ССТ | 0,88 |
| | Полное | 1820 | 66 | 0,89 | 0,49 | 0,77 | | |
| 5 | Сокращенное | 21 | 4 | 0,89 | 0,71 | 0,86 | 1-ПГ-2-ССТ-ЭС | 0,88 |
| | Полное | 4368 | 221 | 0,89 | 0,49 | 0,79 | | |
| 6 | Сокращенное | 7 | 1 | 0,89 | 0,81 | 0,87 | 1-ПГ-2-ССТ-ЭС-5 | 0,88 |
| | Полное | 8008 | 494 | 0,89 | 0,49 | 0,80 | | |
| 7 | Сокращенное | 1 | 1 | 0,88 | 0,88 | 0,88 | 1-ПГ-2-ССТ-ЭС-5-ЦС | 0,88 |
| | Полное | 11 440 | 789 | 0,89 | 0,51 | 0,80 | | |

группированию признаков схожи с рис. 3, а лишь отчасти. Так, схожи по-прежнему Ви и ССт, триады 6 и 2. Удельный вклад ниже среднего также имеют признаки ЦБ, СВЦ, ЦП и АС.

Признаковое пространство на этапе 2 сокращено аналогично этапу 1 до семи признаков (см. рис. 4, б). На корреляционной окружности рис. 4, б видно, что суммарный вклад в ГК1 и ГК2 вырос до 98,3 % (84,0 + 14,3), по сравнению с 78 % (20,0 + 58,2) на корреляционной окружности рис. 4, а.

Результаты вычислительного эксперимента для этапа 2 приведены в табл. 3.

Минимальный КФМ на этапе 2 составляет 0,49, как и на этапе 1, максимальный – 0,90. Комбинации признаков, отобранные на основе МГК, имеют стабильное значение КФМ = 0,88. В результате перебора установлено, что на этапе 2 наилучшим признаком является признак ПГ и комбинация ПГ-ЦБ с КФМ = 0,90.

Заключение

В ходе проведенного сравнительного анализа получены убедительные результаты, позволяющие выбирать наилучшую метрику для распознавания топологии техногенных космических объектов. Определены требования к числу вершин случайных графов, формируемых по моделям ЭР и БА, обеспечивающие близкую к нулевой вероятность изоморфизма порождаемых графов, что необходимо для обеспечения условия непересечения классов графов для распознавания. Методом ГК выявлена коллинеарность таких интегральных характери-

стик графов, как КЦ и АЦ, применительно к исследуемым моделям случайных графов.

Вычислительный эксперимент показал, что оптимальный по критерию качества кластеризации состав признакового пространства разный для различных типов графовых моделей. На этапе 1 исследовались модели различной природы – с низкой и высокой централизацией, но с равным количеством ребер. Проведенный эксперимент показал, что признак «плотность» графа, являющийся интегральным от количества ребер, в таких условиях обладает низкими разделяющими свойствами. А показатели централизации «центральность по степени и по близости», напротив, – высокими. При этом наилучшими признаками являются АС-ЦС-СВЦ-ЦБ-ЦП-ЭС-ПГ, причем оптимальными по размеру и качеству кластеризации являются комбинации ЦС-ЭС и ЦП-ЭС.

При проведении эксперимента на этапе 2 сравнение признаков, полученных от графовых моделей со схожей природой централизации, выявило наилучшие признаки ПГ и ЦБ, причем использование одного признака ПГ дает такое же качество кластеризации по КФМ, как и совместно с ЦБ. При распознавании моделей БА признаковое пространство может быть сокращено до одного признака.

Таким образом, разработанные методы анализа метрик подобия графов позволят на новом качественном уровне решать задачи выявления изменений, происходящих в группе наблюдаемых опасных техногенных космических объектов, представляющих опасность для функционирующих орбитальных средств.

Литература

1. Арсеньев В. Н., Ядренкин А. А., Слатова Е. В. Принятие решения о наиболее опасном техногенном объекте космическим аппаратом мониторинга околоземного космического пространства. *Информационно-управляющие системы*, 2023, № 4, с. 26–34. doi:10.31799/1684-8853-2023-4-26-34, EDN: ВНХТФ
2. Шевченко Д. Н., Литвин А. Ю., Федянин М. А. Имитационное моделирование графа состояний в задачах анализа надежности технических систем. *ПФМТ*, 2018, № 3(36), с. 101–104.
3. Chen P.-Y., Choudhury S., Hero A. Multi-centrality graph spectral decompositions and their application to cyber intrusion detection. *IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, 20–25 March. IEEE, 2016. doi:10.1109/ICASSP.2016.7472539
4. Newman M. E. J. *Networks an Introduction*. N.Y., Oxford University Press Inc., 2010. 1042 p.
5. Trivedi R., Farajtabar M., Biswal P., Zha H. DyREP: Learning representations over dynamic graphs. *Proc. of the 7th Intern. Conf. on Learning Representations (ICLR 2019)*, 2019, May 6–9, pp. 1–25.
6. Pareja A., Domeniconi G., Chen J., Ma T., Suzumura T., Kanezashi H., Kaler T., Schardl T., Leiserson Ch. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. *The Thirty-Fourth AAAI Conf. on Artificial Intelligence (AAAI-20)*, 2020, pp. 5363–5370. doi:doi.org/10.1609/aaai.v34i04.5984
7. Khan A., Bhowmick S., Bonchi F. Summarizing static and dynamic big graphs. *Proc. of the VLDB Endowment*, 2017, vol. 10, iss. 12, pp. 1981–1984. doi:doi.org/10.14778/3137765.3137825
8. Cai J.-Y., Furer M., Immerman N. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 1992, vol. 12, pp. 389–410. doi:doi.org/10.1007/BF01305232
9. Neuen D., Schweitzer P. Benchmark graphs for practical graph isomorphism. *25th Annual European Symp. on Algorithms (ESA 2017). Leibniz Intern. Proc. in Informatics (LIPIcs)*, 2017, vol. 87, pp. 60:1–60:14. doi:10.4230/LIPIcs.ESA.2017.60
10. 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
11. Hartmann A., Mézard M. Distribution of diameters for Erdős — Rényi random graphs. *Phys. Rev.*, 2018, vol. 97, iss. 3. doi:10.1103 / PhysRevE.97.032128
12. Gibson H., Vickers P. Using adjacency matrices to lay out larger small-world networks. *Applied Soft Computing*, 2016, vol. 42, pp. 80–92.
13. Barabasi A. *Network Science*. Glasgow, Cambridge University Press, 2016. 453 p.
14. Zvereva O. M. Triad census usage for communication network analysis. *The 5th Intern. Conf. on Analysis of Images, Social Networks, and Texts*. Yekaterinburg, 2016, pp. 378–389. http://ceur-ws.org/Vol-1710/paper38.pdf (дата обращения: 15.08.2023).
15. Brunson J. C. Triadic analysis of affiliation networks. *Network Science*, 2015, no. 3 (4), pp. 480–508.
16. Татарка М. В., Шуваев Ф. Л. Метод распознавания топологий информационно-телекоммуникационных сетей на основе расчета триадного вектора. *Труды Военно-космической академии им. А. Ф. Можайского*, 2021, № 676, с. 73–84.
17. Faust K. A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks*, 2010, no. 32, pp. 221–233.
18. Faust K. Comparing social networks: size, density, and local structure. *Metodološki zvezki*, 2006, vol. 2, pp. 185–216.
19. Hayat S. Computing distance-based topological descriptors of complex chemical networks: New theoretical techniques. *Chemical Physics Letters*, 2017, vol. 688, pp. 51–58.
20. Mowshowitz A., Dehmer M. A symmetry index for graphs. *Symmetry: Culture and Science*, 2010, vol. 21, pp. 321–327.
21. Mueller L., Schutte M., Kugler K., Dehmer M. QuACN: an R package for analyzing complex biological networks quantitatively. *Bioinformatics*, 2011, vol. 27, pp. 140–141. doi:10.1093/bioinformatics/btq606
22. Bonchi F., De Francisci G., Riondato M. Centrality measures on big graphs: Exact, approximated, and distributed algorithms. *Proc. of the 25th Intern. Conf. “Companion on World Wide Web”*, 2016, pp. 1017–1020. doi:10.1145/2872518.2891063
23. Brandes U., Borgatti S., Freeman L. Maintaining the duality of closeness and betweenness centrality. *Social Networks*, 2016, vol. 44, pp. 153–159.
24. Mino A., Salehzadeh-Yazdi A., Razaghi-Moghadam Z., Hennig H., Wolkenhauer O. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Systems Biology*, 2018, vol. 12, no. 1, pp. 80:1–80:17.
25. 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, т. 19, № 6, с. 1307–1331. doi:10.15622/ia.2020.19.6.7
26. 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.

27. Lê S., Josse J., Husson F. FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 2008, vol. 25, iss. 1, pp. 1–18.

28. Bai L., Cheng X., Liang J., Shen H., Guo Y. Fast density clustering strategies based on the k-means algorithm. *Pattern Recognition*, 2017, vol. 71, pp. 375–386.

29. Depaolini M., Ciucci D., Calegari S., Dominoni M. External indices for rough clustering. *Lecture Notes in Computer Science*, 2018, vol. 11103.

30. Joy A., Korobova J. Bootstrapping clustered data in R using Imeresampler. arXiv: 2106.06568, 2021. doi:10.48550/arXiv.2106.06568

UDC 004.93

doi:10.31799/1684-8853-2024-1-9-19

EDN: OWCDGM

Determination of the optimal set of graph similarity metrics in problems of recognition of topologies of technogenic space objects

N. V. Gruzdev^a, PhD, Military, Associate Professor, orcid.org/0009-0000-6125-1362, NikGruzdev@rambler.ru

F. L. Shuvaev^a, PhD, Tech., Acting Head of Labs, orcid.org/0000-0001-7502-9455

^aA. F. Mozhaiskii Military Space Academy, 13, Zhdanovskaia Emb., 197198, Saint-Petersburg, Russian Federation

Introduction: The task of analyzing the results of observations of man-made space objects in order to identify groups of objects of the same topological type remains vital in the conditions of technogenic pollution of interplanetary environment. To solve this problem, it is proposed to use the tools of graph theory and recognition theory to evaluate the topologies of man-made space objects of various types. **Purpose:** To perform a comparative analysis of graph similarity metrics, which consists in finding the optimal metric within the framework of quality indicators for recognizing man-made space objects. **Results:** The conducted research has shown that there are three main approaches to the calculation of graph similarity metrics. Each of the approaches has its base on calculating the vector of parameters characterizing the graph from different sides. Developed metrics make it possible to distinguish between groups of man-made space objects represented in the form of graphs. A computational experiment has shown that the optimal composition of the feature space according to the clustering quality criterion is different for different types of graph models. This property plays an important role in solving problems of recognizing the topology of man-made space objects at a new qualitative level. Using the proposed methods of comparative analysis of graph similarity metrics, it is possible to choose the best metric for recognizing the topology of man-made space objects. **Practical relevance:** The use of the proposed methods for analyzing graph similarity metrics will allow us to assess changes in groups of similar man-made space objects. The proposed methods can be used in solving problems of control and cataloging of man-made space objects.

Keywords – technogenic space objects, graph, graph similarity metrics, topology, similarity measure, clustering.

For citation: Gruzdev N. V., Shuvaev F. L. Determination of the optimal set of graph similarity metrics in problems of recognition of topologies of technogenic space objects. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2024, no. 1, pp. 9–19 (In Russian). doi:10.31799/1684-8853-2024-1-9-19, EDN: OWCDGM

References

1. Arseniev V. N., Yadrenkin A. A., Slatova E. V. Making a decision on the most dangerous man-made object by a spacecraft for monitoring near-Earth space. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 4, pp. 26–34 (In Russian). doi:10.31799/1684-8853-2023-4-26-34, EDN: BHIXTF
2. Shevchenko D. N., Litvin A. Yu., Fedianin M. A. Reliability analysis of technical systems by simulation modeling of graph of states. *Problems of Physics, Mathematics and Technics*, 2018, no. 3(36), pp. 101–104 (In Russian).
3. Chen P.-Y., Choudhury S., Hero A. Multi-centrality graph spectral decompositions and their application to cyber intrusion detection. *IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016. IEEE, 2016. doi:10.1109/ICASSP.2016.7472539
4. Newman M. E. J. *Networks an Introduction*. N.Y., Oxford University Press Inc., 2010. 1042 p.
5. Trivedi R., Farajtabar M., Biswal P., Zha H. DyREP: Learning representations over dynamic graphs. *Proc. of the 7th Intern. Conf. on Learning Representations (ICLR 2019)*, 2019, pp. 1–25.
6. Pareja A., Domeniconi G., Chen J., Ma T., Suzumura T., Kanezashi H., Kaler T., Schardl T., Leiserson Ch. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. *The Thirty-Fourth AAAI Conf. on Artificial Intelligence (AAAI-20)*, 2020, pp. 5363–5370. doi:doi.org/10.1609/aaai.v34i04.5984
7. Khan A., Bhowmick S., Bonchi F. Summarizing static and dynamic big graphs. *Proc. of the VLDB Endowment*, 2017, vol. 10, iss. 12, pp. 1981–1984. doi:doi.org/10.14778/3137765.3137825
8. Cai J.-Y., Furer M., Immerman N. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 1992, vol. 12, pp. 389–410. doi:doi.org/10.1007/BF01305232
9. Neuen D., Schweitzer P. Benchmark graphs for practical graph isomorphism. *25th Annual European Symp. on Algorithms (ESA 2017). Leibniz Intern. Proc. in Informatics (LIPIcs)*, 2017, vol. 87, pp. 60:1–60:14. doi:10.4230/LIPIcs.ESA.2017.60
10. Shuvaev F. L., Tatarka M. V. Dynamics of centrality measures of random graph mathematical models. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 2, pp. 249–256 (In Russian). doi:10.17586/2226-1494-2020-20-2-249-256
11. Hartmann A., Mézard M. Distribution of diameters for Erdős — Rényi random graphs. *Phys. Rev.*, 2018, vol. 97, iss. 3. doi:10.1103/PhysRevE.97.032128
12. Gibson H., Vickers P. Using adjacency matrices to lay out larger small-world networks. *Applied Soft Computing*, 2016, vol. 42, pp. 80–92.
13. Barabasi A. *Network Science*. Glasgow, Cambridge University Press, 2016. 453 p.
14. Zvereva O. M. Triad census usage for communication network analysis. *The 5th Intern. Conf. on Analysis of Images, Social Networks, and Texts. Yekaterinburg*, 2016, pp. 378–389. Available at: <http://ceur-ws.org/Vol-1710/paper38.pdf> (accessed 15 August 2023).
15. Brunson J. C. Triadic analysis of affiliation networks. *Network Science*, 2015, no. 3 (4), pp. 480–508.
16. Tatarka M. V., Shuvaev F. L. Method of recognition of topologies of information-telecommunication networks on the

- basis of calculation of triad census. *Proc. of the Mozhaisky Military Space Academy*, 2021, vol. 676, pp. 73–84 (In Russian).
17. Faust K. A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks*, 2010, no. 32, pp. 221–233.
 18. Faust K. Comparing social networks: size, density, and local structure. *Metodološki zvezki*, 2006, vol. 2, pp. 185–216.
 19. Hayat S. Computing distance-based topological descriptors of complex chemical networks: New theoretical techniques. *Chemical Physics Letters*, 2017, vol. 688, pp. 51–58.
 20. Mowshowitz A., Dehmer M. A symmetry index for graphs. *Symmetry: Culture and Science*, 2010, vol. 21, pp. 321–327.
 21. Mueller L., Schutte M., Kugler K., Dehmer M. QuACN: an R package for analyzing complex biological networks quantitatively. *Bioinformatics*, 2011, vol. 27, pp. 140–141. doi:10.1093/bioinformatics/btq606
 22. Bonchi F., De Francisci G., Riondato M. Centrality measures on big graphs: Exact, approximated, and distributed algorithms. *Proc. of the 25th Intern. Conf. "Companion on World Wide Web"*, 2016, pp. 1017–1020. doi:10.1145/2872518.2891063
 23. Brandes U., Borgatti S., Freeman L. Maintaining the duality of closeness and betweenness centrality. *Social Networks*, 2016, vol. 44, pp. 153–159.
 24. Mino A., Salehzadeh-Yazdi A., Razaghi-Moghadam Z., Hennig H., Wolkenhauer O. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Systems Biology*, 2018, vol. 12, no. 1, pp. 80:1–80:17.
 25. Ereemeev 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, no. 6, pp. 1307–1331 (In Russian). doi:10.15622/ia.2020.19.6.7
 26. 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 of the State Fire Service of the Ministry of Emergencies of Russia*, 2020, no. 2, pp. 67–77 (In Russian).
 27. Lê S., Josse J., Husson F. FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 2008, vol. 25, iss. 1, pp. 1–18.
 28. Bai L., Cheng X., Liang J., Shen H., Guo Y. Fast density clustering strategies based on the k-means algorithm. *Pattern Recognition*, 2017, vol. 71, pp. 375–386.
 29. Depaolini M., Ciucci D., Calegari S., Dominoni M. External indices for rough clustering. *Lecture Notes in Computer Science*, 2018, vol. 11103.
 30. Joy A., Korobova J. Bootstrapping clustered data in R using lmerSampler. arXiv: 2106.06568, 2021. doi:10.48550/arXiv.2106.06568

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Текст рукописи должен быть оригинальным, а цитирование и самоцитирование корректно оформлено.

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, ORCID и электронный адрес одного из авторов. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы. Предоставляйте подрисночные подписи и названия таблиц на русском и английском языках.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени.

Простые **формулы** набирайте в Word, сложные с помощью редактора MathType или Equation. Для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта в MathType никогда не пользуйтесь вкладкой Other, Smaller, Larger, используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; пробелы в формуле ставьте только после запятой при перечислении с помощью Ctrl+Shift+Space (пробел); не отделяйте пробелами знаки: + = ×, а также пространство внутри скобок; для выделения греческих символов в MathType полужирным начертанием используйте Style → Other → bold.

Для набора формул в Word никогда не используйте вкладки: «Уравнение», «Конструктор», «Формула» (на верхней панели: «Вставка» — «Уравнение»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Подробнее см. <http://i-us.ru/index.php/ius/author-guide>

Иллюстрации:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (*.vsd, *.vsdx); Adobe Illustrator (*.ai); Coreldraw (*.cdr, версия не выше 15); Excel (*.xls); Word (*.docx); AutoCad, Matlab (экспорт в PDF, EPS, SVG, WMF, EMF); Компас (экспорт в PDF); веб-портал DRAW.IO (экспорт в PDF); Inkscape (экспорт в PDF);

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисночных подписей и названий таблиц на русском и английском языках обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png, *.jpg с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение;

— экспертное заключение.

Список литературы

составляется по порядку ссылок в тексте и оформляется следующим образом:

- для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц, doi;
- для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц, doi;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (<http://i-us.ru/paperrules/>): Литература и References.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Руководство для авторов» — <http://i-us.ru/index.php/ius/author-guide>.

Контакты

Куда: 190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ

Кому: Редакция журнала «Информационно-управляющие системы»

Тел.: (812) 494-70-02

Эл. почта: ius.spb@gmail.com

Сайт: www.i-us.ru