

# Дендрограммы в региональном социально-экономическом анализе: интерпретация и верификация

В.И. Блануца<sup>1</sup>

Институт географии им. В.Б. Сочавы СО РАН

<sup>1</sup> ORCID: 0000-0003-3958-216X, [blanutsa@list.ru](mailto:blanutsa@list.ru)

## Аннотация

В последнее десятилетие вырос интерес к региональному социально-экономическому анализу. Для визуализации его результатов используются различные инструменты. Одним из них является дендрограмма. Однако в мировой науке отсутствует обобщение опыта применения дендрограмм для отображения последовательности объединения регионов в социально-экономические кластеры. Поэтому целью нашего исследования стало обобщение такого опыта с акцентом на интерпретацию и верификацию дендрограмм. На основе восьми библиографических баз данных с помощью специального алгоритма семантического поиска выявлено более восьмидесяти журнальных статей, опубликованных во всем мире в последние два десятилетия. В этих статьях размещены сто тридцать дендрограмм. Их анализ показал, что основным назначением древовидных диаграмм является фиксация последовательности объединения регионов в кластеры и обоснование количества кластеров. Предложено два новых вида интерпретации дендрограмм – выделение ядер в кластерах и определение уровня социально-экономической сплоченности кластеров. Для проверки обоснованности определения количества кластеров предложен авторский алгоритм выявления оптимального варианта кластеризации, опирающийся на представление о сложности дерева в теории графов. Перечислены десять основных проблем визуализации результатов регионального социально-экономического анализа с помощью дендрограмм.

**Ключевые слова:** региональная экономика, региональный анализ, иерархический кластерный анализ, экономическая конвергенция регионов, социально-экономическое районирование, дендрограмма, интерпретация, верификация.

## 1. Введение

Региональный социально-экономический анализ нацелен на изучение функционирования (взаимодействия) единиц административно-территориального деления государства (части государства; иногда административных единиц нескольких государств) для управления социально-экономическим развитием регионов и обоснования региональной экономической политики. По набору оцениваемых параметров и используемых методов социально-экономический анализ на региональном уровне несколько отличается от такого же анализа на локальном (в пределах населенного пункта), межнациональном (несколько государств) и глобальном (все государства) уровнях [1–5]. Если оперировать количеством статей, ежегодно публикуемых в научных журналах во всем мире, то в XXI в. наблюдается рост интереса к проблематике регионального социально-экономического анализа (рис. 1). Результаты такого анализа визуализируются в основном с помощью картосхемы (упрощенная географическая карта, на которой отображены только границы регионов). Гораздо реже для визуализации выявленных межрегиональных социально-экономических различий используются интерактивные онлайн-инструменты картографирования [6],

серии картограмм [7], картограммы в сочетании с самоорганизующимися картами Кохонена [8], картодиаграммы и карты-анаморфозы [9]. Дополнительно можно отметить такой редко встречающийся способ визуализации, как кластерграмма [10]. При отображении результатов социально-экономической кластеризации регионов также используются картосхемы, но в иерархическом кластерном анализе (в отличие от неиерархического) картосхема дополняется или заменяется дендрограммой (древовидной диаграммой), на которой фиксируется последовательность объединения регионов в группы (кластеры). Дендрограмма применяется не только для визуализации результатов кластерного анализа, но еще экономической конвергенции регионов [11, 12] и социально-экономического районирования [13, 14].

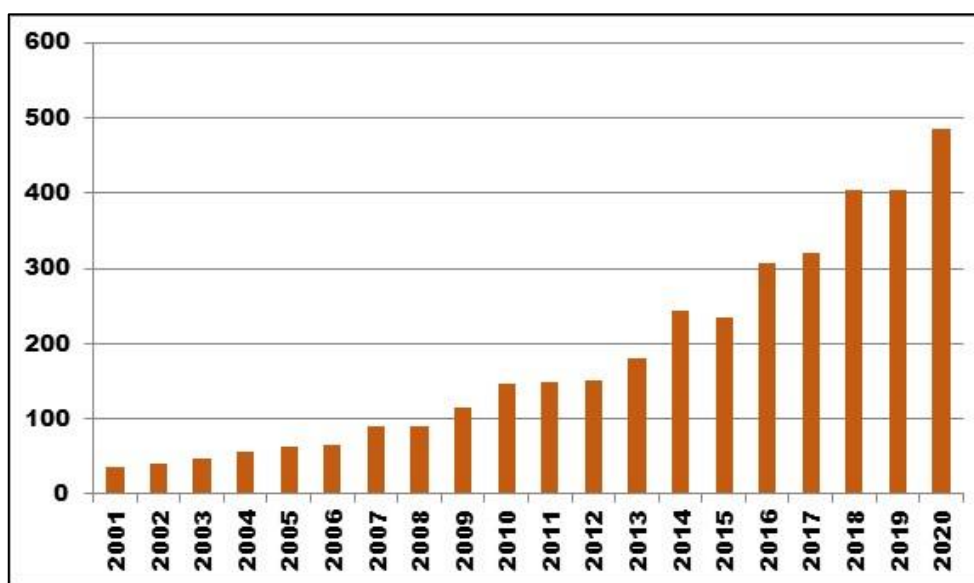


Рис. 1. Изменение ежегодного количества статей по региональному социально-экономическому анализу, опубликованных в научных журналах всего мира в 2001–2020 гг. (по данным Scopus на 15 апреля 2021 г.)

В некоторых научных направлениях уже проведено обобщение опыта применения дендрограмм для визуализации результатов исследования (например, в биоинформатике [15], селекции растений [16] и при изучении функционального разнообразия экосистем [17]). Однако в региональной экономике подобные обобщения не делались. Поэтому целью нашего исследования стало обобщение мирового опыта применения дендрограмм для визуализации результатов регионального социально-экономического анализа. Для достижения цели потребовалось решить следующие задачи: идентифицировать мировой массив (корпус) публикаций по рассматриваемой проблематике; извлечь из него множество дендрограмм; определить общие характеристики выявленных древовидных диаграмм; уяснить существующие и наметить перспективные способы интерпретации дендрограмм; диагностировать применяемые и предложить новые алгоритмы верификации дендрограмм. При реализации поставленных задач использовались следующие ограничения: анализировалась кластеризация регионов (существует еще кластеризация признаков); рассматривалась дендрограмма без ее зависимости от матрицы сходства или различия регионов в признаковом пространстве (в большинстве публикаций такие матрицы отсутствуют); учитывалась последовательность объединения регионов в группы вне зависимости от выбранной меры сходства и метода кластеризации (не во всех публикациях они приводятся); обсуждалась агломеративная схема получения кластеров (объединение отдельных регионов в группы), так как дивизионная схема (деление всех регионов на группы) не получила распространение в региональной экономике; изучалась последовательность объединения регионов в пространстве

экономических или социально-экономических признаков (без кластеризации регионов по исключительно социальным признакам, что выходит за рамки региональной экономики). Вне нашего исследования оказалось извлечение из дендрограмм специфических представлений для машинного обучения [18], так как это перспективное направление относится к проблематике использования алгоритмов искусственного интеллекта в региональной экономике [19] и требует отдельного рассмотрения.

## **2. Материалы и методы**

При решении первой задачи рассматривались только журнальные статьи, так как по ним можно получить все тексты с иллюстрациями, а по другим видам научных публикаций – монографиям, сборникам статей и материалам конференций – не все доступно. Поэтому наши выводы относятся только к массиву статей, опубликованных в научных журналах во всем мире. В качестве хронологического ограничения выбраны последние двадцать лет (2001–2020 гг.). До этого статьи по рассматриваемой проблематике почти не публиковались. Для выявления статей с дендрограммами, относящимися к региональному социально-экономическому анализу, использовались одна отечественная и семь международных библиографических баз данных (Научная электронная библиотека eLIBRARY.RU, издательства Springer, Wiley, Elsevier и SAGE, базы данных Web of Science, Scopus и IDEAS). В этих базах заложен поиск необходимых публикаций по ключевым словам. Однако такой поиск имеет много недостатков. Например, по ключевому слову «regional socio-economic dendrogram» в базе данных Scopus удалось обнаружить только одну статью, посвященную устойчивому развитию сельских территорий.

Поэтому для обнаружения искомых статей использовалась «Самоорганизующаяся система поиска публикаций по заданной теме в библиографической базе данных» [19], представляющая собой алгоритм машинного обучения с постоянным расширением семантического поля. Этот алгоритм применялся к каждой из восьми баз в итерационном режиме: семантическое поле, выявленное в одной базе, применялось и расширялось в следующей базе, после чего происходил возврат к предыдущей базе для дополнительного поиска статей по расширенному полю. Так происходило до тех пор, пока не стабилизировался размер семантического поля. Ограничением алгоритма является отбор публикаций только на кириллице и латинице. Поэтому научные статьи с использованием другого алфавита (к примеру, китайского или арабского) остались вне нашего анализа. Другим ограничением стало использование только восьми баз данных, которые охватывают большинство, но не все статьи в мире.

Извлечение из выявленных статей дендрограмм (вторая задача) не представляет какой-либо сложности, кроме качества иллюстраций, которое не во всех журналах одинаково высокое. При решении оставшихся трех задач использовались дендрограммы, тексты статей с их описанием, а также авторские наработки по обобщению, интерпретации и верификации схем социально-экономического районирования [14].

## **3. Результаты и обсуждение**

Применение алгоритма семантического поиска [19] позволило обнаружить в восьми библиографических базах данных 81 журнальную статью по региональному социально-экономическому анализу, в которых использовались дендрограммы объединения территориальных единиц в группы (кластеры, клубы конвергенции, районы). Эти статьи распределились между 73 журналами. Больше всего статей (три) опубликовано в журнале «European Journal of Operational Research». При фиксации ежегодного количества статей наблюдалось последовательное (с колебаниями) увеличение публикационной активности к концу рассматриваемого периода (рис. 2), в результате

чего на последние четыре года пришлось примерно половина всех статей (42 из 81). При этом удельный вес публикаций в экономических журналах оказался менее 30%. Большинство статей содержало один (60 публикаций) или два (16) рисунка, а на каждом рисунке размещалась одна (98 случаев), две (3), три (4), четыре (2) или шесть (1) древовидных диаграмм. В итоге было идентифицировано 130 дендрограмм, ставших источником последующих обобщений.

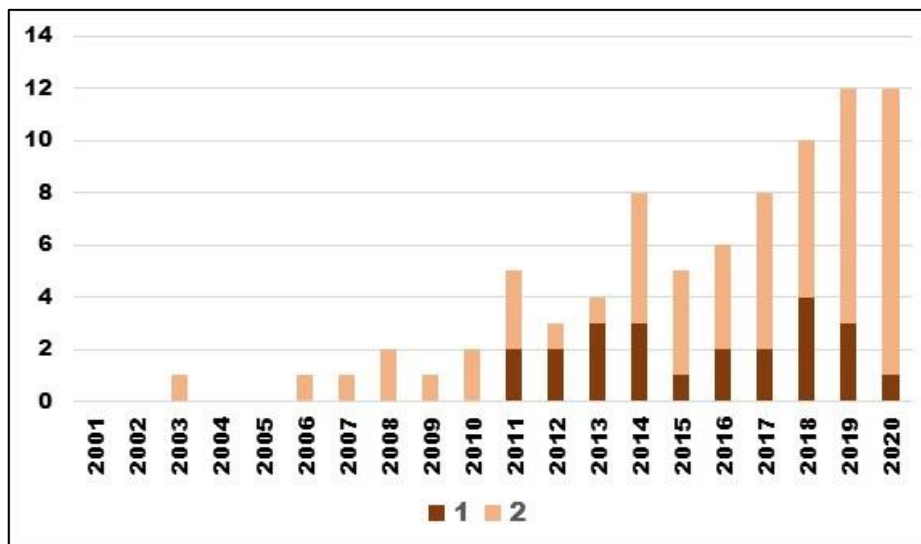


Рис. 2. Изменение ежегодного количества статей с визуализацией результатов регионального социально-экономического анализа с помощью дендрограмм, опубликованных в экономических (1) и других (2) научных журналах во всем мире в 2001–2020 гг. (составлено автором)

Главное назначение дендрограммы как визуализатора результатов регионального социально-экономического анализа заключалось в отображении последовательности объединения регионов в группы по заданным признакам (рис. 3) и подтверждении обоснованности выбора количества групп. В подавляющем большинстве проанализированных статей использовались обе функции, но в 5 статьях древовидная диаграмма применялась только для демонстрации группировки регионов. Территориальные единицы сравнивались между собой по одному (чаще всего оценивался валовой региональный продукт на душу населения) или несколькими признаками. Во втором случае осуществлялся расчет величины сходства или различия между каждой парой регионов (предпочтение отдавалось евклидову расстоянию). На основе этих расстояний проводилось объединение регионов в группы с помощью различных методов. К сожалению, по 72 дендрограммам метод либо не указан, либо приводилась ссылка на стандартные программные продукты, в которых представлено несколько методов. Среди упомянутых алгоритмов доминировал (36 дендрограмм) метод Уорда [20].

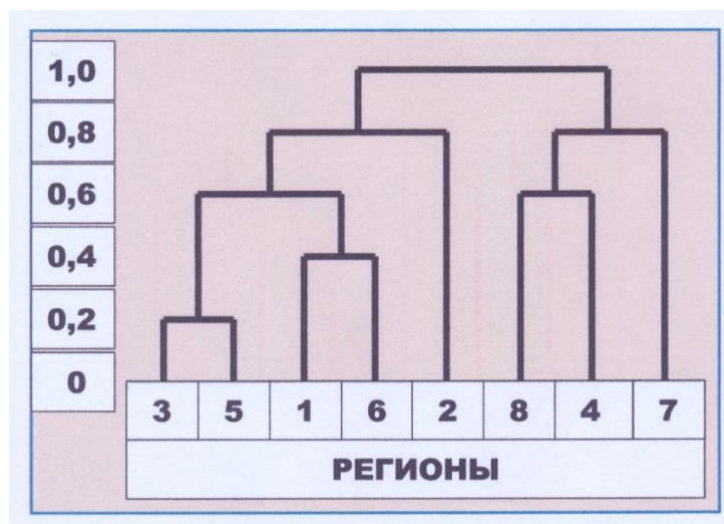


Рис. 3. Условный пример последовательности объединения регионов (1–8) в группы в зависимости от межрегионального расстояния в признаковом пространстве (0–1,0)

Основное визуальное различие между дендрограммами проявлялось в выборе направления сходимости древовидной структуры. Имели место пять альтернативных направлений: вверх (см. рис. 3), вниз, вправо, влево или внутрь. Последнее направление связано с круговой (радиальной) дендрограммой, в которой сходимость направлена к центру круга (внутрь). В анализируемом массиве предпочтение отдавалось графическому представлению группировки регионов по двум альтернативным направлениям (рис. 4).

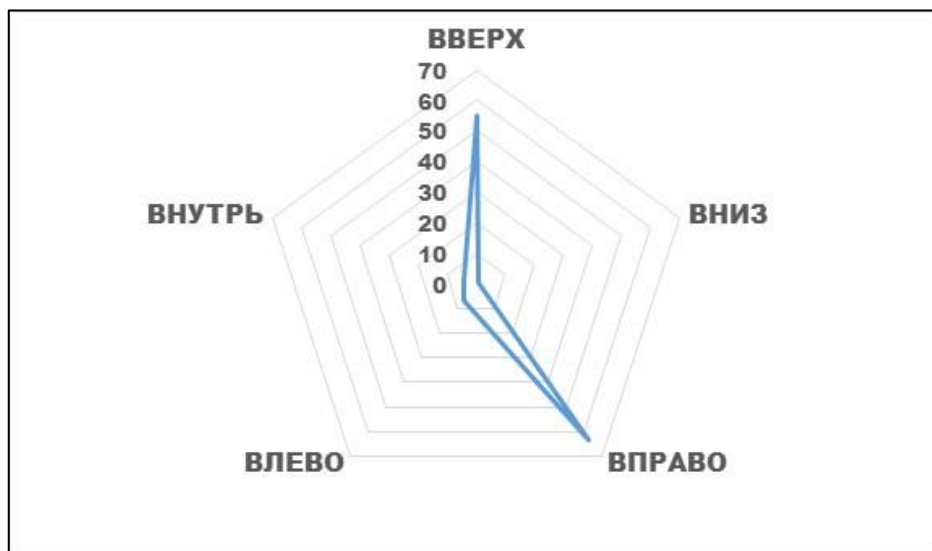


Рис. 4. Количество дендрограмм из журнальных статей по региональному социально-экономическому анализу (2001–2020 гг.), на которых последовательность объединения регионов в группы графически представлено по разным направлениям

Последней общей характеристикой дендрограмм (третья задача) является количество идентифицированных кластеров. В региональном социально-экономическом анализе вне зависимости от исходного количества территориальных единиц чаще других выделялось четыре кластера (рис. 5).

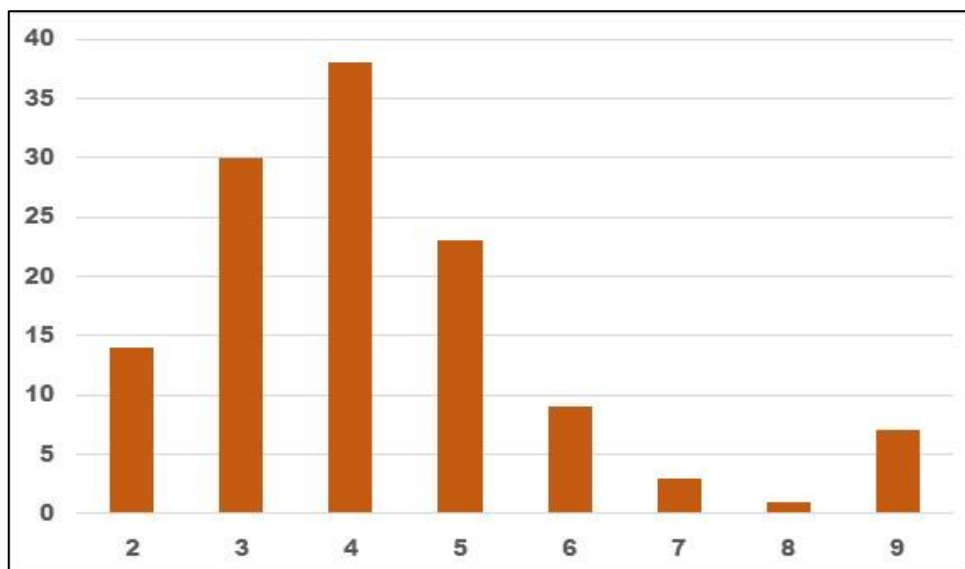


Рис. 5. Количество дендрограмм с разным числом кластеров (от 2 до 9), размещенных в журнальных статьях по региональному социально-экономическому анализу (2001–2020 гг.)

В результате решения четвертой задачи исследования установлено, что при региональном социально-экономическом анализе интерпретация дендрограмм сводилась только к расшифровке двух ее основных функций – характеристике последовательности объединения регионов в группы и подтверждению обоснованности выбора определенного количества групп (кластеров). При этом характеристика объединения регионов заключалась в повторе того, что уже показано на дендрограмме – как один регион на каком-то шаге группировки соединялся с другим регионом, а затем к ним присоединялся третий регион и так далее. Подобное описание не является интерпретацией, поскольку не несет новую информацию и не трактует результаты с позиции другой концепции (подхода). В такой ситуации можно предложить два новых вида интерпретации дендрограмм: выделение ядер в кластерах и определение уровня социально-экономической сплоченности кластеров.

Опираясь на три правила выделения ядер в типах территориальных информационно-коммуникационных сетей [21], можно определить ядра по дендрограмме социально-экономического анализа, понимая под ядром группу наиболее похожих регионов, к которой в ходе объединения присоединяются остальные регионы рассматриваемого кластера. При этом наименьшее сходство или максимальное расстояние достигается на последнем шаге группировки, когда все регионы объединяются в один кластер. Наибольшее сходство между регионами равно единице или нулевому расстоянию. Однако при региональном измерении социально-экономических процессов редко встречаются полностью похожие территориальные единицы [14]. Поэтому устанавливается некоторая величина допустимого сходства, не превышая которую регионы считаются наиболее похожими. Обоснование этой величины – тема отдельного исследования. Если учитывать опыт Европейского союза по экономической конвергенции регионов [22], то такой величиной является допустимое отклонение на 25% от желательной социально-экономической ситуации. Применительно к нашей проблематике это выражается в 75% сходства или 25% расстояния. По условному примеру видно (см. рис. 3), что регионы 3 и 5 объединяются в группу при расстоянии между ними в 0,20, а остальные регионы группируются при расстояниях более 0,25. Допустим, в условном примере выделяются два кластера: регионы 1, 2, 3, 5, 6 и регионы 4, 7, 8. Тогда первый кластер будет одноядерным, а второй – безъядерным. Возможен также многоядерный кластер, когда имеется две и более группы, образованные при сходстве не менее 75% (расстоянии не более 25%).



Идентификация безъядерных, одно- и многоядерных кластеров придает новый смысл интерпретации последовательности объединения регионов в группы. Например, при кластеризации 13 регионов (периферий) Греции по 11 социально-экономическим показателям с помощью метода средней связи (average linkage) в 1995, 2000 и 2007 гг. [23] было выявлено по 5 кластеров (рис. 6). При максимальном расстоянии в 25 единиц ядра формируются до расстояния 6,25 (25% от максимума). В таком случае можно получить следующую интерпретацию: один двухъядерный и четыре безъядерных кластера в 1995 г.; три одноядерных и два безъядерных кластера в 2000 г.; один двухъядерный, один одноядерный и три безъядерных кластера в 2007 г. При ужесточении требований к минимально допустимому расстоянию до 10% (2,5 для данного случая) получается другой расклад: один одноядерный и четыре безъядерных кластера в 1995 г.; один двухъядерный, один одноядерный и три безъядерных кластера в 2000 г.; один двухъядерный и четыре безъядерных кластера в 2007 г. Выбор 25%, 10% или другого максимально допустимого расстояния зависит от специфики социально-экономической ситуации в Греции и требует соответствующего обоснования. Поэтому расстояния в 6,25 и 2,5 единицы приводятся только в качестве иллюстрации возможности идентификации ядер. Подтверждением невозможности применения единого допустимого расстояния (до 25% от максимума) на все дендрограммы служат примеры выделения кластеров при меньших расстояниях в социально-экономическом районировании Сьерра-Невады (Испания) [24] и предпринимательской кластеризации округов Румынии [25]. Поэтому в каждом конкретном случае необходимо не только обосновывать выбор меры сходства регионов и метода кластерного анализа, но еще и допустимого сходства (различия).

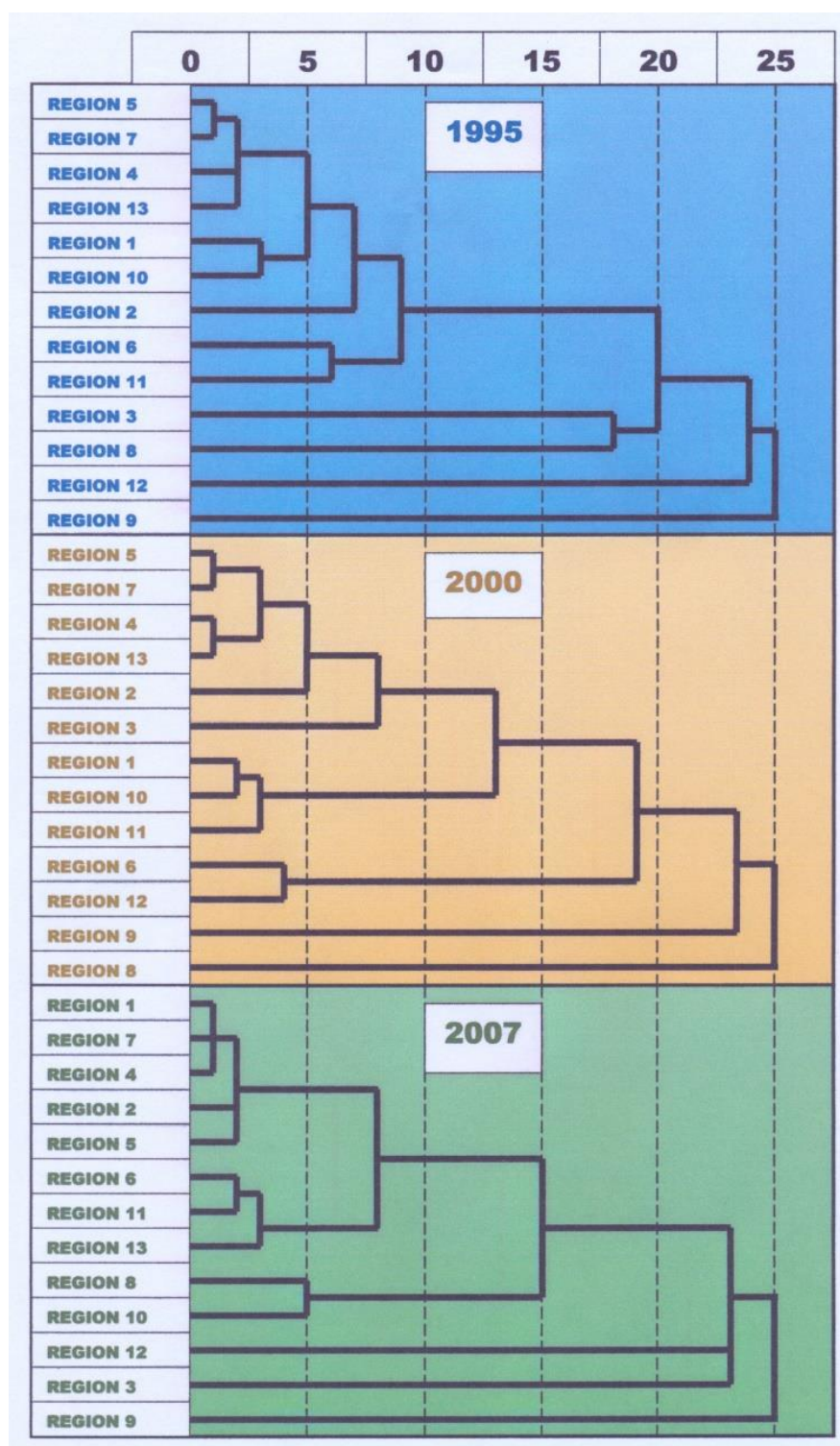


Рис. 6. Социально-экономическая кластеризация регионов Греции при последовательном увеличении межрегионального расстояния в признаковом пространстве

от 0 до 25 единиц (1995, 2000 и 2007 гг.) [23]

Регионы: 1 – Восточная Македония и Фракия, 2 – Центральная Македония, 3 – Западная Македония, 4 – Эпир, 5 – Фессалия, 6 – Ионические острова, 7 – Западная Греция, 8 – Центральная Греция, 9 – Аттика, 10 – Пелопоннес, 11 – Северные Эгейские острова, 12 – Южные Эгейские острова, 13 – Крит.

Определение уровня сплоченности кластеров по социально-экономическим признакам также может опираться на максимально допустимое расстояние. Здесь интерпретация связана с уяснением того, насколько разнородны (однородны)



полученные кластеры. Об этом можно судить по величине расстояния, на котором проходит линия «разрезания» дендрограммы на кластеры. Если в нашем условном примере (см. рис. 3) выделяются два кластера, то разрезание дендрограммы происходит на расстоянии 0,8. Данная величина может оцениваться с учетом максимально допустимого расстояния – при 25% весь интервал (0–100%) делится на четыре квартиля. Условный пример попадает в последний (четвертый) квартиль. Для сравнения, 5 кластеров греческих регионов (см. рис. 6) характеризуются вторым квартилем, что указывает на их более высокую сплоченность по отношению к условному примеру. Можно не оперировать квартилями или другими полосами значений, а сравнивать различные дендрограммы по линии разрезания. Только для обеспечения сравнения следует использовать относительные величины сходства или различия. Сопоставление уровня сплоченности регионов, полученного в ходе проведенного анализа, с аналогичным уровнем в контрольном примере открывает новые возможности интерпретации.

Последний и редко встречающийся вид интерпретации последовательности объединения территорий возникает в случае появления «выброса» – отдельного региона, который настолько не похож на остальные регионы, что его не считают кластером. Поэтому такой регион исключают из кластеризации как «информационный шум». Однако с гносеологической точки зрения анализ целостных территориальных образований не должен содержать исключения [14]. Существование очень специфических регионов как кластеров позволяет интерпретировать результат группировки как неуравновешенную систему, а игнорирование этих регионов приводит к появлению «белых пятен» на исследуемой территории, что искажает реальную социально-экономическую ситуацию. Именно по данному вопросу проходит разграничение между отечественным и зарубежным региональным анализом. В некоторых российских работах предполагаемые выбросы удаляются априори (до кластеризации), а в иностранных исследованиях – апостериори (после кластеризации). Примером может служить исключение из Центрального федерального округа города Москва [26, 27] или Москвы с Московской областью [28], а из Северо-Кавказского округа – Чеченской Республики [29]. За рубежом не стали считать отдельным кластером, к примеру, область Валле-д'Аоста при экономической конвергенции регионов Италии [30] и провинцию Цинхай в американском исследовании режимов социального обеспечения в Китае [31].

Вторая группа видов интерпретации дендрограмм связана с определением оптимального количества кластеров. Это может быть сделано на основе матрицы сходства (различия) регионов по заданным признакам (показателям, параметрам), что позволяет рассчитывать индексы эффективности кластеризации (например, индекс Калинского-Харабаса [32, 33]) и с этих позиций оценивать дендрограммы, путем сравнения с географической картой (предпочтение отдается варианту, в котором регионы внутри каждого кластера являются соседями [34]) или в результате визуального анализа дендрограммы. В мировом массиве статей по региональному социально-экономическому анализу отмеченные матрицы приводятся редко, а требование географической компактности кластеров относится только к районированию. Поэтому ограничимся визуальным анализом.

Процедура визуального анализа довольно субъективна и может проявляться в выборе количества кластеров, например, по «нашему пониманию особенностей индийских штатов» [35] или если получают группы регионов, «поддающиеся экономической интерпретации» [36]. Однако в большинстве случаев разрезание дендрограммы происходило по ее наиболее длинным отрезкам без объединения регионов в группы. В случае с группировкой греческих периферий (см. рис. 6) разрезание проведено примерно на удалении в 10 единиц от исходного состояния. Для 1995 г. отрезок между образованием первого кластера (регионы 1, 2, 4, 5, 6, 7, 10, 11 и 13) и объединением регионов 3 и 8 действительно является наиболее длинным, а для 2000

и 2007 гг. максимальное расстояние без объединений приходится на следующий отрезок, на котором наблюдаются четыре кластера вместо декларируемых пяти групп. Если обратиться к условному примеру (см. рис. 3), то там все отрезки являются одинаковыми (по 0,20 единицы расстояния). Приведенные примеры указывают на то, что выбор на дендрограмме наиболее длинного отрезка не всегда проводится правильно или не позволяет выделить максимальный отрезок. Поэтому необходим эвристический алгоритм, формализующий интуитивные представления о визуализации оптимального количества кластеров.

Еще один вид интерпретации числа кластеров связан с незавершенностью исследования. В большинстве публикаций определение оптимального количества групп регионов завершало исследование (далее следовала только характеристика выделенных кластеров). Однако в нескольких случаях иерархический кластерный анализ использовался для обоснования числа групп в последующем неиерархическом объединении регионов в кластеры с помощью метода  $k$ -средних [37–39]. В таких случаях интерпретировалось не распределение регионов по кластерам, а только число кластеров в иерархической группировке, которое иногда трактовалось как избыточное для неиерархического объединения [37].

Почти во всех выявленных статьях количество и состав кластеров определялись путем прямолинейного разрезания дендрограммы. Однако в одном случае наблюдалось криволинейное разрезание, приводящее к неправильной интерпретации числа кластеров. При изучении цикличности региональных цен на жилье в США была установлена (по дендрограмме) возможность существования 4 кластеров. Однако затем авторы решили разделить один большой кластер на два подкластера на том основании, что это «позволяет выделить интересные детали, касающиеся географического распределения циклов цен на жилье» [40]. В итоге дендрограмма разрезалась по кривой линии, а подкластеры рассматривались наравне с оставшимися тремя кластерами. Непонятно, почему сразу не был сделан выбор 5 кластеров (по дендрограмме это можно сделать), а потребовалось выйти на 4-кластерное решение и затем только в рамках первого кластера вернуться к варианту группировки территорий, соответствующему 5-кластерному решению.

Дендрограммы могут использоваться не только для интерпретации последовательности объединения регионов в кластеры и выбора оптимального количества кластеров, но и для проверки (верификации) результатов регионального социально-экономического анализа. Поскольку одним из основных результатов такого анализа является объединение территорий в группы (кластеры, клубы конвергенции, районы), то в первую очередь необходимо проверить обоснованность объединения. При этом будем оперировать только дендрограммами как визуализатором объединения регионов (как отмечалось выше, воспользоваться матрицами сходства или различия территорий не представляется возможным, так как в большинстве статей они не приводятся).

Один из наиболее распространенных способов верификации результатов группировки регионов – определение оптимального числа кластеров с помощью альтернативного алгоритма. Этой точки зрения придерживаются многие исследователи в области регионального анализа [32, 34, 36, 39, 41–46]. Однако применяемые ими алгоритмы нельзя назвать «альтернативными». К примеру, предлагалось разделить все признаки (переменные) на две группы и с помощью метода Уорда провести сначала кластеризацию греческих регионов по первой группе признаков, а затем сравнить полученные результаты с кластеризацией этих же регионов по обоим группам признаков [41]; посмотреть различия между группировками округов Хорватии по абсолютным и относительным показателям [42]; сопоставить группы территорий по региональному индексу развития информационно-коммуникационных технологий, полученные путем применения метода Уорда и метода полной связи [39]; сравнить группировки итальянских регионов по тенденциям

обеспечения занятости инвалидов, полученные тремя методами – одиночной, полной и средней связи [32], или объединения городов Бангладеш в результате использования четырех методов, отличающихся только критерием присоединения региона к группе [43]; визуально сопоставить географические карты с объединением французских муниципалитетов по социально-экономическому расстоянию между ними и их же объединением с учетом географического расстояния между соседями [34]. В приведенных примерах алгоритмы были однотипными (одна и та же исходная матрица или ее разновидность), что позволяло оценить только некоторые различия. Поэтому для проверки результатов регионального анализа целесообразно использовать метод, никак не связанный с процедурами этого анализа. Таким «независимым» алгоритмом, не применявшимся в выявленном массиве статей, может служить эвристический способ определения наиболее сложного яруса леса в социально-экономическом районировании [14].

Общую структуру альтернативного алгоритма можно представить в виде следующей последовательности действий: проверяемая дендрограмма преобразуется в группировочное дерево; рассчитываются абсолютные и относительные значения сложности каждого яруса леса (шага группировки регионов); выбирается наиболее сложный ярус (вариант кластеризации) и сравнивается с проверяемым кластерным решением. Для визуализации выбора оптимального числа кластеров исходная дендрограмма трансформируется в одну из своих разновидностей – группировочное дерево, в котором вершины соединены ребрами без образования петель (циклов, замкнутых контуров). В нашем условном примере (см. рис. 3) можно выделить 5 группировочных шагов (один шаг равен 0,20 единицы расстояния), приводящих, соответственно, к пяти вариантам объединения 8 регионов в 7, 6, 4, 2 и 1 кластер (рис. 7). Последний вариант охватывает все регионы и в этом смысле не является кластеризацией. Остаются четыре варианта, среди которых необходимо выбрать оптимальное решение.

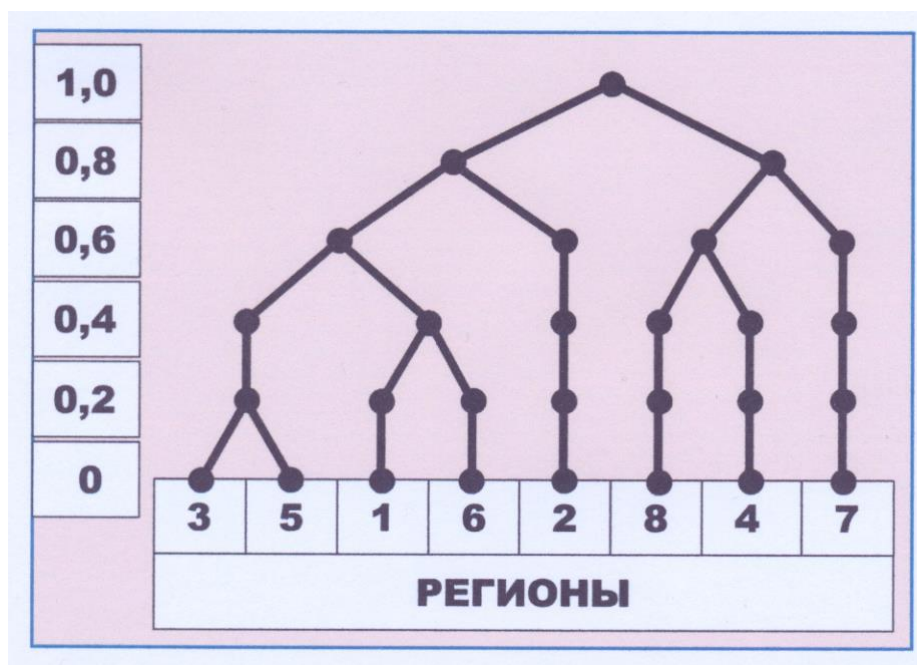


Рис. 7. Группировочное дерево, соответствующее дендрограмме условного примера (см. рис. 3)

При выборе оптимального варианта социально-экономического районирования происходит поиск наилучшего соотношения между требованиями максимальной однородности групп соседних регионов и минимальным числом таких групп [47, 48]. Если каждый регион является уникальным (наиболее распространенная ситуация при

региональном анализе [14]), то максимальная однородность групп регионов достигается только в исходной ситуации (число групп или кластеров соответствует количеству регионов). Последующее объединение регионов снижает однородность получаемых групп, но позволяет получить небольшое их количество, что удобно для социально-экономического управления [1, 13, 14]. Проблема в том, что данные требования противоречат друг другу: максимизация однородности приводит к максимизации числа групп или минимизация количества групп приводит к минимизации однородности. Отсюда возникает задача нахождения некоторого оптимума, при котором небольшое количество групп характеризуется приемлемой однородностью. Древовидный граф (группировочное дерево), отражающий последовательность объединения наибольшего количества регионов в наименьшее число групп (кластеров) за минимум шагов, по Ю.А. Шрейдеру [49] является наиболее сложным. Оценить сложность дерева можно с помощью следующей рекуррентной формулы [49]:

$$\sigma(x) = \alpha\gamma + \sum \sigma(y),$$

где  $\sigma(x)$  – сложность вершины  $x$ ;  $\alpha$  – количество ребер, выходящих вниз из вершины  $x$ ;  $\gamma$  – количество вершин дерева;  $\sum \sigma(y)$  – суммарная сложность вершин предыдущего яруса, соединенных ребрами с вершиной  $x$ . По этой формуле сложность всего группировочного дерева в нашем примере (см. рис. 7) составляет 182 единицы ( $2 \times 28 + 78 + 48$ ). Если имеет место не одно дерево, а их множество (лес), то сложность леса  $\sigma(D)$  соответствует суммарной сложности деревьев этого леса:

$$\sigma(D) = \sum_{x=1}^r \sigma(x),$$

где  $r$  – количество деревьев (кластеров). Так, в нашем примере сложность леса после первого шага (7 деревьев) составила 18 единиц ( $2 \times 3 + 2 + 2 + 2 + 2 + 2$ ).

Для определения относительной сложности леса  $C$  необходимо соотнести абсолютное значение с максимально возможной величиной сложности [14]:

$$C = \frac{\sigma(D)}{\sigma(D)_{max}} = \frac{\sigma(D)}{n(nh + k + 1)},$$

где  $C$  – степень сложности леса;  $h$  – количество ярусов леса;  $n$  – число регионов;  $k$  – коэффициент связи с числом ярусов леса, определяемый рекуррентным правилом ( $k = 0$  при  $h = 1$ ,  $k_i = k_{i-1} + h_i$  при  $h \geq 2$ ). По условному примеру для первого и последующих ярусов получились следующие значения  $C$ : 0,25; 0,24; 0,30; 0,38; 0,40. Отбросив последнее значение, наиболее сложным оказался четвертый ярус (объединение 8 регионов в 2 кластера).

Опираясь на представленный алгоритм, можно проверить обоснованность определения числа кластеров по дендрограммам из журнальных статей. Возьмем в качестве примера три дендрограммы кластеризации периферий Греции (см. рис. 6). Трансформируем их в три группировочных дерева, введя шаг объединения регионов в кластеры, равный 1 единице расстояния (минимальный шаг по дендрограмме). Тогда получится 25 шагов (ярусов). Напомним, что в соответствующей работе [23] выделялось по 5 кластеров. Для 1995 г. наибольшая степень сложности леса достигалась при 2-кластерном решении ( $C = 0,2275$ ), а 5-кластерное решение ( $0,2032$ ) уступало по сложности еще и варианту с тремя кластерами ( $0,2094$ ). В 2000 г. идентификация 5 кластеров ( $C = 0,2041$ ) имела меньшую сложность, чем выделение четырех ( $0,2192$ ), трех ( $0,2255$ ) и двух ( $0,2327$ ) кластеров. По 2007 г. 5-кластерное решение ( $C = 0,1857$ ) было хуже группировки регионов в четыре ( $0,1863$ ), восемь ( $0,1989$ ) и два ( $0,2056$ ) кластера. Таким образом, верификация трех дендрограмм с помощью альтернативного алгоритма показала, что 13 греческих регионов объединяются в два кластера, а не в пять групп.

Аналогичным образом можно проверить остальные дендрограммы. Например, 14 регионов Чехии по социально-экономическим характеристикам были сгруппированы в два кластера [50], но верификация показала, что такой вариант ( $C = 0,3986$ ) значительно уступает 3-кластерному решению (0,6667). По другому примеру, связанному с выявлением региональных тенденций в обеспечении занятости инвалидов в Италии [32], установлено, что выполненное авторами объединение 20 регионов в два кластера ( $C = 0,1910$ ) является оптимальным по сравнению с образованием трех (0,1763) или четырех (0,1734) групп. При анализе 16 регионов Польши по индикаторам рынка труда в 2005 и 2014 гг. выделено по 5 кластеров [51], но в обоих случаях оптимальной является группировка в 8 кластеров: со степенью сложности 0,2132 против 0,1946 (5 кластеров) в 2005 г. и 0,2059 против 0,1620 в 2014 г. Другие 8 кластеров ( $C = 0,3731$ ) соответствовали оптимальному варианту группировки 32 провинций Китая по внутренним инвестициям [52], а не 3-кластерное решение китайских исследователей (0,1198).

Приведенные примеры указывают на противоречия в выборе оптимального числа кластеров. Одна из причин неэффективности иерархического кластерного анализа – неиерархический результат, проявляющийся в выделение только одного вида таксонов. К примеру, в социально-экономическом районировании идентифицируются несколько иерархически упорядоченных таксонов – зона, подзона, провинция и округ [14], каждый из которых характеризуется увеличением степени сложности леса  $C$  по сравнению с предыдущим и последующим шагом группировки. Этот подход можно распространить на кластерный анализ в том случае, если при постепенном уменьшении числа кластеров наблюдается не один, а несколько «пиков» величины  $C$ . Количество таких пиков указывает на число таксонов. Тогда, например, при трех пиках, соответствующих увеличению степени сложности от первого к третьему пику, можно получить иерархический результат в виде кластеров (третий пик), подкластеров (второй пик) и групп (первый пик). «Кластером» целесообразно называть таксон с наибольшей степенью сложности, а часть кластера – «подкластером», «группой», «подгруппой». В таком случае объединение кластеров можно назвать «надкластером», «сплочением» и т.д. Альтернативой иерархическому результату может быть нечеткое кластерное решение, когда выделяются ядра кластеров, а остальные регионы с разной степенью (мерой) относятся к каждому ядру. В проанализированном массиве статей кластеризация регионов осуществлялась без учета иерархии таксонов и теории нечетких множеств.

## 4. Заключение

Обобщение мирового опыта использования дендрограмм для визуализации результатов регионального социально-экономического анализа, проведенное с учетом только журнальных статей за последние два десятилетия, позволило выявить следующие проблемы: (1) редкое применение (дендрограммы встречаются примерно в 1 из 45 статей; см. рис. 1 и 2); (2) использование только для отображения последовательности объединения регионов в кластеры и определения количества кластеров; (3) отсутствие дивизионной схемы получения кластеров (деление всей изучаемой территории на кластеры; проведение такого деления могло бы использоваться для верификации результатов объединения регионов в кластеры или агломеративной схемы); (4) не частое представление в виде радиальной дендрограммы (см. рис. 4; перечисление регионов по кругу позволяет зафиксировать значительно больше регионов, чем перечисление по диаметру, который является одной из сторон в обычных дендрограммах, но это преимущество не использовалось во многих исследованиях с большим количеством регионов, когда вместо последовательности объединения регионов в кластеры приводилась урезанная группировка некоторых множеств регионов в кластеры); (5) отсутствие способов идентификации безъядерных,



одно- и многоядерных кластеров; (6) недопонимание уровня социально-экономической сплоченности групп регионов; (7) игнорирование аномальных регионов («выбросов»); (8) субъективность визуального выбора оптимального числа кластеров; (9) недостаток альтернативных алгоритмов кластеризации для верификации дендрограмм; (10) неиерархичность результата иерархического кластерного анализа. В нашем исследовании первые четыре проблемы только обозначены, а по оставшимся шести проблемам намечены лишь контуры будущих решений.

## Благодарности

Исследование выполнено за счет средств государственного задания (номер регистрации темы АААА-А17-117041910166-3).

## Список литературы

1. Soltes V., Stofkova K.R., Kutaj M. (2016) Socio-economic analysis of development of regions. *Global Journal of Business, Economics and Management*, vol. 6, no. 2, pp. 171–178. doi: 10.18844/gjbem.v6i2.1382
2. Bross U., Walter G.H. (2000) *Socio-Economic Analysis of North Rhine-Westphalia*, Karlsruhe: Fraunhofer Institute for Systems and Innovation Research.
3. Zhang J., He X., Yuan X.-D. (2020) Research on the relationship between urban economic development and urban spatial structure – A case study of two Chinese cities. *PLoS ONE*, vol. 15, no. 7, e0235858. doi: 10.1371/journal.pone.0235858
4. Fura B., Wang Q. (2017) The level socioeconomic development of EU countries and the state of ISO 14001 certification. *Quality and Quantity*, vol. 51, no. 1, pp. 103–119. doi: 10.1007/s11135-015-0297-7
5. Lensen M., Li M., Malik A., Pomponi F., Sun Y.-Y., Wiedmann T., Faturay F., Fry J., Gallego B., Geschke A., Gómez-Paredes J., Kanemoto K., Kenway S., Nansai K., Prokopenko M., Wakiyama T. Wang Y., Yousefzadeh M. (2020) Global socio-economic losses and environmental gains from the Coronavirus pandemic. *PLoS ONE*, vol. 15, no. 7, e0235654. doi: 10.1371/journal.pone.0235654
6. Smith D.A. (2016) Online interactive thematic mapping: Applications and techniques for socio-economic research. *Computers, Environment and Urban Systems*, vol. 57, pp. 106–117. doi: 10.1016/j.compenvurbsys.2016.01.002
7. Sova M., Vukosav B. (2017) A contribution to measuring and visualizing regional disparities in the Czech Republic. *Kartografija i Geoinformacije*, vol. 16, no. 28, pp. 26–44.
8. Hajek P., Henriques R., Hajkova V. (2014) Visualizing components of regional innovation systems using self-organizing maps – Evidence from European regions. *Technological Forecasting and Social Change*, vol. 84, pp. 197–214. doi: 10.1016/j.techfore.2013.07.013
9. Гордеев С.С. (2016) Оценка устойчивости пространственного социо-эколого-экономического развития в среде геоинформатики. *Вестник Челябинского государственного университета. Экономические науки*, № 11, с. 37–50.
10. Schonlau M. (2004) Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. *Computational Statistics*, vol. 19, pp. 95–111. doi: 10.1007/BF02915278
11. Balash V., Balash O., Faizliev A., Chistopolskaya E. (2020) Economic growth patterns: Spatial econometric analysis for Russian regions. *Information*, vol. 11, e289. doi: 10.3390/info11060289
12. Блануца В.И. (2020) Кластеризация регионов Сибири и Дальнего Востока по перспективным экономическим специализациям. *Вестник Омского университета. Серия «Экономика»*, т. 12, № 2, с. 80–90. doi: 10.24147/1812-3988.2020.18(2).80-90

13. Mitchell W., Watts M. (2010) Identifying functional regions in Australia using hierarchical aggregation techniques. *Geographical Research*, vol. 48, no. 1, pp. 24–41. doi: 10.1111/j.1745-5871.2009.00631.x
14. Блануца В.И. (2018) *Социально-экономическое районирование в эпоху больших данных*, М.: ИНФРА-М, 194 с.
15. Schroeder M., Gilbert D., van Helden J., Noy P. (2001) Approaches to visualization in bioinformatics: From dendrograms to Space Explorer. *Information Science*, vol. 139, no. 1-2, pp. 19–57. doi: 10.1016/S0020-0255(01)00156-6
16. Arief V.N., DeLacy I.H., Basford K.E., Dieters M.J. (2017) Application of a dendrogram seriation algorithm to extract pattern from plant breeding data. *Euphytica*, vol. 213, e85. doi: 10.1007/s10681-017-1870-z
17. Petchey O.L., Gaston K.J. (2007) Dendrograms and measuring functional diversity. *Oikos*, vol. 116, no. 8, pp. 1422–1426. doi: 10.1111/j.0030-1299.2007.15894.x
18. Chehreghani M.H., Chehreghani M.H. (2020) Learning representations from dendrograms. *Machine Learning*, vol. 109, pp. 1779–1802. doi: 10.1007/s10994-020-05895-3
19. Блануца В.И. (2020) Региональные экономические исследования с использованием алгоритмов искусственного интеллекта: состояние и перспективы. *Вестник Забайкальского государственного университета*, т. 26, № 8, 2020, с. 100–111. doi: 10.21209/2227-9245-2020-26-8-100-111
20. Ward J. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244.
21. Блануца В.И. (2016) *Развертывание информационно-коммуникационной сети как географический процесс (на примере становления сетевой структуры сибирской почты)*, М.: ИНФРА-М, 246 с.
22. Воронов В.В. (2014) Конвергенция регионов Европейского союза: особенности и оценка. *Экономические и социальные перемены: факты, тенденции, прогноз*, № 6, с. 85–99. doi: 10.15838/esc.2014.6.36.7
23. Goletsis Y., Chletsos M. (2011) Measuring of development and regional disparities in Greek periphery: A multivariate approach. *Socio-Economic Planning Sciences*, vol. 45, no. 4, pp. 174–183. doi: 10.1016/j.seps.2011.06.002
24. Martín-López B., Palomo I., García-Llorente M., Imesta-Arandia I., Castro A.J., Del Amo D.G., Cómez-Baggethun E., Montes C. (2017) Delineating boundaries of social-ecological systems for landscape planning: A comprehensive spatial approach. *Land Use Policy*, vol. 66, pp. 90–104. doi: 10.1016/j.landusepol.2017.04.040.
25. Chivu L. (2019) Local entrepreneurship and social services in Romania. Territorial analysis. *European Research on Management and Business Economics*, vol. 25, no. 2, pp. 79–86. doi: 10.1016/j.iemeen.2019.04.001
26. Согачева О.В. (2016) Кластерный анализ как инструмент управления социально-экономическим развитием региона (на примере Центрального федерального округа). *Теория и практика сервиса: экономика, социальная сфера, технологии*, № 1, 2016, с. 43–46.
27. Жолудева И.И., Мельниченко Н.Ф., Козлов Г.Е. (2014) Применение кластерного анализа для оценки социально-экономического развития регионов на примере ЦФО и Ярославской области. *Экономика, Статистика и Информатика*, № 1, с. 144–148.
28. Филонова Е.С., Букреева Ю.В. (2013) Анализ состояния регионов Центрального федерального округа по показателям предприятий малого и среднего бизнеса. *Вестник Финансового университета*, № 6, с. 35–47.
29. Узденов У.А. (2014) Интеллектуальная система оценки кредитоспособности регионов. Часть 1. Многомерный статистический анализ. *Научный журнал КубГАУ*, № 104, с. 957–968.

30. Brida J.G., Garrido N., Mureddu F. (2014) Italian economic dualism and convergence clubs at regional level. *Quality and Quantity*, vol. 48, pp. 439–456. doi: 10.1007/s11135-012-9779-z
31. Ratigan K. (2017) Disaggregating the Developing Welfare State: Provincial social policy regimes in China. *World Development*, vol. 98, pp. 467–484. doi: 10.1016/j.worlddev.2017.05.010
32. Agovino M., Rapposelli A. (2017) Regional performance trends in providing employment for persons with disabilities: Evidence from Italy. *Social Indicators Research*, vol. 130, pp. 593–615. doi: 10.1007/s11205-015-1186-0
33. Dutta I., Das A. (2019) Exploring the dynamics of spatial inequality through the development of sub-city typologies in English Bazar urban agglomeration and its peri urban areas. *GeoJournal*, vol. 84, pp. 829–849. doi: 10.1007/s10708-018-9895-y
34. Chavent M., Kuentz-Simonet V., Labenne A., Saracco J. (2018) ClustGeo: An R package for hierarchical clustering with spatial constraints. *Computational Statistics*, vol. 33, pp. 1799–1822. doi: 10.1007/s00180-018-0791-1
35. Mookherjee D., White J. (2011) Urban-regional dualism in India: An exploration of developmental indicators across urban size classes. *Asian Geographer*, vol. 28, no. 1, pp. 21–31. doi: 10.1080/10225706.2011.577976
36. Кованова Е.С. (2013) Кластерный анализ в решении задач типологии регионов России по уровню и интенсивности внутренней трудовой миграции. *Вестник НГУЭУ*, № 4, с. 166–175.
37. Абаев В.А., Шахов А.В. (2010) Методический подход к определению типичных сельскохозяйственных районов Липецкой области. *Вестник ФГОУ ВПО МГАУ*, № 6, с. 94–99.
38. Cziráky D., Sambt J., Rován J., Puljiz J. (2006) Regional development assessment: A structural equation approach. *European Journal of Operational Research*, vol. 174, no. 1, pp. 427–442. doi: 10.1016/j.ejor.2005.03.012
39. Novo-Corti I., Barreiro-Gen M. (2015) Public policies based on social networks for the introduction of technology at home: Demographic and socioeconomic profiles of households. *Computers in Human Behavior*, vol. 51, pp. 1216–1228. doi: 10.1016/j.chb.2014.12.040
40. Flor M.A., Klarl T. (2017) On the cyclicity of regional house prices: New evidence for U.S. metropolitan statistical areas. *Journal of Economic Dynamics and Control*, vol. 77, pp. 134–156. doi: 10.1016/j.jedc.2017.02.001
41. Kyriakopoulos G.L., Arabatzis G., Tsialis P., Ioannou K. (2018) Electricity consumption and RES plants in Greece: Typology of regional units. *Renewable Energy*, vol. 127, pp. 134–144. doi: 10.1016/j.renene.2018.04.062
42. Krišto J., Dumičić K., Curković M. (2014) Banking business indicators in Croatian economic surroundings. *Economy of Eastern Croatia: Yesterday, Today, Tomorrow*, vol. 3, pp. 572–581.
43. Rahman S., Mohiuddin H., Kafy A.-A., Shell P.K., Di L. (2019) Classification of cities in Bangladesh based on remote sensing derived spatial characteristics. *Journal of Urban Management*, vol. 8, no. 2, pp. 206–224. doi: 10.1016/j.jum.2018.12.001
44. Литвинова Г.П., Лисицын А.Е. (2018) Оценка социально-экономического потенциала регионов Сибирского федерального округа. *Вестник Кемеровского государственного университета. Серия: Политические, социологические и экономические науки*, № 2, с. 114–121. doi:10.21603/2500-3372-2018-2-114-121
45. Jašková D., Havierníková K. (2020) The human resources as an important factor of regional development. *International Journal of Business and Society*, vol. 21, no. 3, pp. 1464–1478.
46. Andrejiova M., Kimakova Z. (2020) The comparison of transport infrastructures in individual Slovak regions by applying PCA and cluster analysis. *Acta Logistica*, vol. 7, no. 4, pp. 225–234. doi: 10.22306/al.v7i4.182

47. Berry B.J.L. (1961) A method for deriving multi-factor uniform regions. *Przegląd Geograficzny*, vol. 33, no. 2, pp. 263–282.
48. Cliff A.D., Haggett P. (1970) On the efficiency of alternative aggregation in region-building problems. *Environment and Planning A: Economy and Space*, vol. 2, no. 3, pp. 285–294. doi: 10.1068/a020285
49. Шрейдер Ю.А. (1971) *Равенство, сходство, порядок*, М.: Наука, 254 с.
50. Kvičalová J., Mazalová V., Široký J. (2014) Identification of the difference between the regions of the Czech Republic based on the economic characteristics. *Procedia – Economics and Finance*, vol. 12, pp. 343–352. doi: 10.1016/S2212-5671(14)00354-2
51. Tatarczak A., Boichuk O. (2018) The multivariate techniques in evaluation of unemployment analysis of Polish regions. *Oeconomica Copernicana*, vol. 9, no. 3, pp. 361–380. doi: 10.24136/oc.2018.018
52. Ma Y., Zhuang X., Li L. (2011) Research on the relationships of the domestic mutual investment of China based on the cross-shareholding networks of the listed companies. *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 4, pp. 749–759. doi: 10.1016/j.physa.2010.10

# Dendrograms in Regional Socio-Economic Analysis: Interpretation and Verification

V.I. Blanutsa<sup>1</sup>

V.B. Sochava Institute of Geography, Siberian Branch of the Russian Academy of Sciences

<sup>1</sup> ORCID: 0000-0003-3958-216X, [blanutsa@list.ru](mailto:blanutsa@list.ru)

## **Abstract**

Interest in regional socio-economic analysis has grown in the last decade. Various tools are used to visualize its results. One of them is the dendrogram. However, there is no generalization of the experience of using dendrograms to display the sequence of combining regions into socio-economic clusters in world science. Therefore, the goal of our study was to generalize such experience with an emphasis on the interpretation and verification of dendrograms. Based on eight bibliographic databases using a special semantic search algorithm, more than eighty journal articles published around the world in the last two decades have been identified. These articles contain one hundred and thirty dendrograms. Their analysis showed that the main purpose of tree diagrams is to fix the sequence of combining regions into clusters and to substantiate the number of clusters. Two new types of interpretation of dendrograms are proposed – the allocation of nuclei in clusters and the determination of the level of socio-economic cohesion of clusters. To test the validity of determining the number of clusters, the author's algorithm for identifying the optimal clustering option is proposed, based on the idea of the complexity of the tree in graph theory. The ten main problems of visualizing the results of regional socio-economic analysis using dendrograms are listed.

**Keywords:** regional economy, regional analysis, hierarchical cluster analysis, regional economic convergence, socio-economic regionalization, dendrogram, interpretation, verification.

## **References**

1. Soltes V., Stofkova K.R., Kutaj M. (2016) Socio-economic analysis of development of regions. *Global Journal of Business, Economics and Management*, vol. 6, no. 2, pp. 171–178. doi: [10.18844/gjbem.v6i2.1382](https://doi.org/10.18844/gjbem.v6i2.1382)
2. Bross U., Walter G.H. (2000) *Socio-Economic Analysis of North Rhine-Westphalia*, Karlsruhe: Fraunhofer Institute for Systems and Innovation Research.
3. Zhang J., He X., Yuan X.-D. (2020) Research on the relationship between urban economic development and urban spatial structure – A case study of two Chinese cities. *PLoS ONE*, vol. 15, no. 7, e0235858. doi: [10.1371/journal.pone.0235858](https://doi.org/10.1371/journal.pone.0235858)
4. Fura B., Wang Q. (2017) The level socioeconomic development of EU countries and the state of ISO 14001 certification. *Quality and Quantity*, vol. 51, no. 1, pp. 103–119. doi: [10.1007/s11135-015-0297-7](https://doi.org/10.1007/s11135-015-0297-7)
5. Lensen M., Li M., Malik A., Pomponi F., Sun Y.-Y., Wiedmann T., Faturay F., Fry J., Gallego B., Geschke A., Gómez-Paredes J., Kanemoto K., Kenway S., Nansai K., Prokopenko M., Wakiyama T. Wang Y., Yousefzadeh M. (2020) Global socio-economic losses and environmental gains from the Coronavirus pandemic. *PLoS ONE*, vol. 15, no. 7, e0235654. doi: [10.1371/journal.pone.0235654](https://doi.org/10.1371/journal.pone.0235654)
6. Smith D.A. (2016) Online interactive thematic mapping: Applications and techniques for socio-economic research. *Computers, Environment and Urban Systems*, vol. 57, pp. 106–117. doi: [10.1016/j.compenvurbsys.2016.01.002](https://doi.org/10.1016/j.compenvurbsys.2016.01.002)



7. Sova M., Vukosav B. (2017) A contribution to measuring and visualizing regional disparities in the Czech Republic. *Kartografija i Geoinformacije*, vol. 16, no. 28, pp. 26–44.
8. Hajek P., Henriques R., Hajkova V. (2014) Visualizing components of regional innovation systems using self-organizing maps – Evidence from European regions. *Technological Forecasting and Social Change*, vol. 84, pp. 197–214. doi: 10.1016/j.techfore.2013.07.013
9. Gordeev S.S. (2016) Ocenka ustojchivosti prostranstvennogo socio-ekologo-ekonomicheskogo razvitiya v srede geoinformatiki [Assessment of the sustainability of spatial socio-ecological-economic development in the environment of geoinformatics]. *Vestnik Chelyabinskogo gosudarstvennogo universiteta. Ekonomicheskie nauki*, no. 11, pp. 37–50 [in Russian].
10. Schonlau M. (2004) Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. *Computational Statistics*, vol. 19, pp. 95–111. doi: 10.1007/BF02915278
11. Balash V., Balash O., Faizliev A., Chistopolskaya E. (2020) Economic growth patterns: Spatial econometric analysis for Russian regions. *Information*, vol. 11, e289. doi: 10.3390/info11060289
12. Blanutsa V.I. (2020) Klasterizaciya regionov Sibiri i Dal'nego Vostoka po perspektivnym ekonomicheskim specializacijam [Clustering regions of Siberia and the Far East by promising economic specializations]. *Vestnik Omskogo universiteta. Seriya "Ekonomika"*, vol. 12, no. 2, pp. 80–90 [in Russian]. doi: 10.24147/1812-3988.2020.18(2).80-90
13. Mitchell W., Watts M. (2010) Identifying functional regions in Australia using hierarchical aggregation techniques. *Geographical Research*, vol. 48, no. 1, pp. 24–41. doi: 10.1111/j.1745-5871.2009.00631.x
14. Blanutsa V.I. (2018) *Social'no-ekonomicheskoe rajonirovanie v epohu bol'shij dannyh* [Socio-Economic Regionalization in the Era of Big Data], Moscow: INFRA-M Publ. [in Russian]. doi: 10.12737/monography\_59f81ac5ede918.09423566
15. Schroeder M., Gilbert D., van Helden J., Noy P. (2001) Approaches to visualization in bioinformatics: From dendrograms to Space Explorer. *Information Science*, vol. 139, no. 1-2, pp. 19–57. doi: 10.1016/S0020-0255(01)00156-6
16. Arief V.N., DeLacy I.H., Basford K.E., Dieters M.J. (2017) Application of a dendrogram seriation algorithm to extract pattern from plant breeding data. *Euphytica*, vol. 213, e85. doi: 10.1007/s10681-017-1870-z
17. Petchey O.L., Gaston K.J. (2007) Dendrograms and measuring functional diversity. *Oikos*, vol. 116, no. 8, pp. 1422–1426. doi: 10.1111/j.0030-1299.2007.15894.x
18. Chehreghani M.H., Chehreghani M.H. (2020) Learning representations from dendrograms. *Machine Learning*, vol. 109, pp. 1779–1802. doi: 10.1007/s10994-020-05895-3
19. Blanutsa V.I. (2020) Regional'nye ekonomicheskie issledovaniya s ispol'zovaniem algoritmov iskusstvennogo intellekta: sostoyanie i perspektivy [Regional economic research using artificial intelligence algorithms: state and prospects]. *Vestnik Zabajkal'skogo gosudarstvennogo universiteta*, vol. 26, no. 8, pp. 100–111 [in Russian]. doi: 10.21209/2227-9245-2020-26-8-100-111
20. Ward J. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244.
21. Blanutsa V.I. (2016) *Razvertyvanie informacionno-kommunikacionnoj seti kak geograficheskij process (na primere stanovleniya setevoy struktury sibirskoj pochty)* [Deployment of an Information and Communication Network as a Geographic Process (On the Example of the Formation of the Network Structure of the Siberian Post)], Moscow: INFRA-M Publ. [in Russian]
22. Voronov V.V. (2014) Konvergenciya regionov Evropejskogo soyuza: osobennosti i ocenka [Convergence of regions of the European Union: features and assessment].

*Ekonomicheskie i social'nye peremeny: fakty, tendencii, prognoz*, no. 6, pp. 85–99 [in Russian]. doi: 10.15838/esc.2014.6.36.7

23. Goletsis Y., Chletsos M. (2011) Measuring of development and regional disparities in Greek periphery: A multivariate approach. *Socio-Economic Planning Sciences*, vol. 45, no. 4, pp. 174–183. doi: 10.1016/j.seps.2011.06.002

24. Martín-López B., Palomo I., García-Llorente M., Imesta-Arandia I., Castro A.J., Del Amo D.G., Cómez-Baggethun E., Montes C. (2017) Delineating boundaries of social-ecological systems for landscape planning: A comprehensive spatial approach. *Land Use Policy*, vol. 66, pp. 90–104. doi: 10.1016/j.landusepol.2017.04.040

25. Chivu L. (2019) Local entrepreneurship and social services in Romania. Territorial analysis. *European Research on Management and Business Economics*, vol. 25, no. 2, pp. 79–86. doi: 10.1016/j.iedeen.2019.04.001

26. Sogacheva O.V. (2016) Klasternyj analiz kak instrument upravleniya social'no-ekonomicheskim razvitiem regiona (na primere Central'nogo federal'nogo okruga) [Cluster analysis as a tool for managing the socio-economic development of the region (On the example of the Central Federal District)]. *Teoriya i praktika servisa: ekonomika, social'naya sfera, tekhnologii*, no. 1, pp. 43–46 [in Russian].

27. Zholudeva I.I., Mel'nichenko N.F., Kozlov G.E. (2014) Primenenie klasternogo analiza dlya ocenki social'no-ekonomicheskogo razvitiya regionov na primere CFO i Yaroslavskoj oblasti [The use of cluster analysis to assess the socio-economic development of regions on the example of the Central Federal District and the Yaroslavl Region]. *Ekonomika, Statistika i Informatika*, no. 1, pp. 144–148 [in Russian].

28. Filonova E.S., Bukreeva Yu.V. (2013) Analiz sostoyaniya regionov Central'nogo federal'nogo okruga po pokazatelyam predpriyatij malogo i srednego biznesa [Analysis of the state of the regions of the Central Federal District by indicators of small and medium-sized businesses]. *Vestnik Finansovogo universiteta*, no. 6, pp. 35–47 [in Russian].

29. Uzdenov U.A. (2014) Intellekturnaya sistema ocenki kreditosposobnosti regionov. Chast' 1. Mnogomernyj statisticheskij analiz [Intelligent system for assessing the creditworthiness of regions. Part 1. Multivariate statistical analysis]. *Nauchnyj zhurnal KubGAU*, no. 104, pp. 957–968 [in Russian].

30. Brida J.G., Garrido N., Mureddu F. (2014) Italian economic dualism and convergence clubs at regional level. *Quality and Quantity*, vol. 48, pp. 439–456. doi: 10.1007/s11135-012-9779-z

31. Ratigan K. (2017) Disaggregating the Developing Welfare State: Provincial social policy regimes in China. *World Development*, vol. 98, pp. 467–484. doi: 10.1016/j.worlddev.2017.05.010

32. Agovino M., Rapposelli A. (2017) Regional performance trends in providing employment for persons with disabilities: Evidence from Italy. *Social Indicators Research*, vol. 130, pp. 593–615. doi: 10.1007/s11205-015-1186-0

33. Dutta I., Das A. (2019) Exploring the dynamics of spatial inequality through the development of sub-city typologies in English Bazar urban agglomeration and its peri urban areas. *GeoJournal*, vol. 84, pp. 829–849. doi: 10.1007/s10708-018-9895-y

34. Chavent M., Kuentz-Simonet V., Labenne A., Saracco J. (2018) ClustGeo: An R package for hierarchical clustering with spatial constraints. *Computational Statistics*, vol. 33, pp. 1799–1822. doi: 10.1007/s00180-018-0791-1

35. Mookherjee D., White J. (2011) Urban-regional dualism in India: An exploration of developmental indicators across urban size classes. *Asian Geographer*, vol. 28, no. 1, pp. 21–31. doi: 10.1080/10225706.2011.577976

36. Kovanova E.S. (2013) Klasternyj analiz v reshenii zadach tipologii regionov Rossii po urovnyu i intensivnosti vnutrennej trudovoj migracii [Cluster analysis in solving problems of the typology of Russian regions by the level and intensity of internal labor migration]. *Vestnik NGUEU*, no. 4, pp. 166–175 [in Russian].

37. Abaev V.A., Shahov A.V. (2010) Metodicheskiy podhod k opredeleniyu tipichnyh sel'skohozyajstvennyh rajonov Lipeckoj oblasti [Methodical approach to the definition of typical agricultural areas of the Lipetsk Region]. *Vestnik FGOU VPO MGAU*, no. 6, pp. 94–99 [in Russian].
38. Cziráky D., Sambt J., Rován J., Puljiz J. (2006) Regional development assessment: A structural equation approach. *European Journal of Operational Research*, vol. 174, no. 1, pp. 427–442. doi: 10.1016/j.ejor.2005.03.012
39. Novo-Corti I., Barreiro-Gen M. (2015) Public policies based on social networks for the introduction of technology at home: Demographic and socioeconomic profiles of households. *Computers in Human Behavior*, vol. 51, pp. 1216–1228. doi: 10.1016/j.chb.2014.12.040
40. Flor M.A., Klarl T. (2017) On the cyclicity of regional house prices: New evidence for U.S. metropolitan statistical areas. *Journal of Economic Dynamics and Control*, vol. 77, pp. 134–156. doi: 10.1016/j.jedc.2017.02.001
41. Kyriakopoulos G.L., Arabatzis G., Tsialis P., Ioannou K. (2018) Electricity consumption and RES plants in Greece: Typology of regional units. *Renewable Energy*, vol. 127, pp. 134–144. doi: 10.1016/j.renene.2018.04.062
42. Krišto J., Dumičić K., Curković M. (2014) Banking business indicators in Croatian economic surroundings. *Economy of Eastern Croatia: Yesterday, Today, Tomorrow*, vol. 3, pp. 572–581.
43. Rahman S., Mohiuddin H., Kafy A.-A., Shell P.K., Di L. (2019) Classification of cities in Bangladesh based on remote sensing derived spatial characteristics. *Journal of Urban Management*, vol. 8, no. 2, pp. 206–224. doi: 10.1016/j.jum.2018.12.001
44. Litvinova G.P., Lisicyn A.E. (2018) Ocenka social'no-ekonomicheskogo potenciala regionov Sibirskogo federal'nogo okruga [Assessment of the socio-economic potential of the regions of the Siberian Federal District]. *Vestnik Kemerovskogo gosudarstvennogo universiteta. Seriya: Politicheskie, sociologicheskie i ekonomicheskie nauki*, no. 2, pp. 114–121 [in Russian]. doi:10.21603/2500-3372-2018-2-114-121
45. Jašková D., Havierníková K. (2020) The human resources as an important factor of regional development. *International Journal of Business and Society*, vol. 21, no. 3, pp. 1464–1478.
46. Andrejiova M., Kimakova Z. (2020) The comparison of transport infrastructures in individual Slovak regions by applying PCA and cluster analysis. *Acta Logistica*, vol. 7, no. 4, pp. 225–234. doi: 10.22306/al.v7i4.182
47. Berry B.J.L. (1961) A method for deriving multi-factor uniform regions. *Przeglad Geograficzny*, vol. 33, no. 2, pp. 263–282.
48. Cliff A.D., Haggett P. (1970) On the efficiency of alternative aggregation in region-building problems. *Environment and Planning A: Economy and Space*, vol. 2, no. 3, pp. 285–294. doi: 10.1068/a020285
49. Shrader Yu.A. *Ravenstvo, skhodstvo, poryadok* [Equality, similarity, order], Moscow: Nauka, 1971 [in Russian].
50. Kvičalová J., Mazalová V., Široký J. (2014) Identification of the difference between the regions of the Czech Republic based on the economic characteristics. *Procedia – Economics and Finance*, vol. 12, pp. 343–352. doi: 10.1016/S2212-5671(14)00354-2
51. Tatarczak A., Boichuk O. (2018) The multivariate techniques in evaluation of unemployment analysis of Polish regions. *Oeconomica Copernicana*, vol. 9, no. 3, pp. 361–380. doi: 10.24136/oc.2018.018
52. Ma Y., Zhuang X., Li L. (2011) Research on the relationships of the domestic mutual investment of China based on the cross-shareholding networks of the listed companies. *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 4, pp. 749–759. doi: 10.1016/j.physa.2010.10