

ЛЕСОТОРГОВОЕ РАЙОНИРОВАНИЕ РОССИИ: РАЗРАБОТКА МЕТОДИКИ КЛАСТЕРИЗАЦИИ

© 2024 Н. А. Синицын^{1, *}, М. С. Елютин^{1, **}, А. С. Коротков^{2, ***},
А. О. Кушлевич^{1, ****}, Р. Р. Меняльщикова^{1, *****}, Д. С. Русаков^{1, *****}

¹Московский государственный университет имени М.В. Ломоносова, географический факультет, Москва, Россия

²Московский государственный университет имени М.В. Ломоносова, экономический факультет, Москва, Россия

*e-mail: nicksinus@yandex.ru

**e-mail: mikhail.elyutin@student.msu.ru

***e-mail: korotkov.andrey@geogr.msu.ru

****e-mail: artem.kushlevich@yandex.ru

*****e-mail: roman.menyalschikov@yandex.ru

*****e-mail: d_c_rusakov@mail.ru

Поступила в редакцию 27.06.2022 г.

После доработки 15.02.2024 г.

Принята к публикации 14.03.2024 г.

Для сетевого анализа лесной отрасли России использованы данные единой государственной автоматизированной информационной системы учета древесины по торговле лесоматериалами между отдельными предприятиями за 2020 г. Цель исследования — выделить и охарактеризовать кластеры лесной промышленности России. Разработана методика кластеризации графа, состоящая из двух шагов: сначала выполняется кластеризация алгоритмом Лейдена, затем каждый лейденский кластер кластеризируется еще раз с помощью укладки Фрюхтермана—Рейнгольда. Это позволяет решить проблему масштаба, свойственную алгоритмам на основе оптимизации модулярности. Предложен подход для снятия неопределенности результатов кластеризации недетерминированных алгоритмов. Выделенные кластеры имеют высокую замкнутость — в среднем 89% оборота их предприятий сосредоточено внутри каждого из них. Модулярность кластеризации высокая — 0.86. При нанесении кластеров на карту в разрезе населенных пунктов формируются ареалы поселений, принадлежащих одному и тому же кластеру — лесоторговые районы. Обычно они имеют хорошо читаемые границы, часто совпадающие с границами регионов. Выделены пять типов сетевых структур кластеров: вертикальная моноцентрическая, вертикальная полицентрическая, горизонтальная, дендритная и простая. Выделены три фактора формирования кластеров: производственная цепочка (лесопильная, фанерная, целлюлозная, плиточная), сбытовая цепочка (перераспределение в рамках кластера, аккумуляция для экспорта или для продажи в другой кластер, спекуляция), наличие общего собственника у группы предприятий кластера.

Ключевые слова: районирование, сетевой анализ, экономические кластеры, Россия, лесная промышленность, кластеризация графа

DOI: 10.31857/S2587556624020083, **EDN:** DSVJXG

ВВЕДЕНИЕ

С 2015 г. в России функционирует Единая государственная автоматизированная информационная система (ЕГАИС) “Учет древесины и сделок с ней”¹, из которой взята информация по торговым сделкам между компаниями лесной промышленности. Это новый источник данных — методики анализа не разработаны, потенциал не понятен. Наличие информации в разрезе отдельных предприятий позволяет анализировать экономические связи на максимально детальном уровне. Цель исследования —

понять, распадаются ли предприятия отрасли на отдельные группы, внутри которых связи более интенсивны, чем с внешним окружением. Если такие группы удастся выделить — нанести их на карту. Тем самым выявляется — согласуется ли близость в пространстве производственных связей с географической концентрацией на территории страны. При выполнении этого условия допустимо рассматривать группы предприятий как лесоторговые районы. По классификации Б.Б. Родмана (1999) это будут коннекционные районы коммуникационного типа, для выделения которых достаточно только анализа связей.

¹ <https://lesegais.ru/open-area/deal> (дата обращения 30.07.2024).

Важно проанализировать лесоторговые районы с точки зрения логики работы лесной промышленности — имеют ли они адекватное содержание, которое объясняется с точки зрения технологических, экономических и географических процессов. Важно выделить факторы формирования групп предприятий, провести их типологию. Группировка предприятий по данным о производственных связях на уровне отдельных компаний позволит более строго подойти к выделению экономических кластеров.

ОБЗОР ЛИТЕРАТУРЫ

В современной социально-экономической географии выделяются два представления о территориальных структурах в промышленности — территориально-производственные комплексы (ТПК) и кластеры (табл. 1). ТПК — “взаимообусловленное (соподчиненное) сочетание производственных предприятий и селитьбы (населенных мест) либо на ограниченной территории (локальные комплексы), либо на территории экономического района или подрайона (районные комплексы)” (Колосовский, 1969, с. 142; 2006). Кластер — “группа географически соседствующих взаимосвязанных компаний и связан-

ных с ними организаций, действующих в определенной сфере, характеризующихся общностью деятельности и взаимодополняющих друг друга” (Портер, 2005, с. 258).

Первые работы по экономической кластеризации в России — районирование хлебной торговли конца XIX — начала XX в. Детальные данные позволяли выделять районы по связям между отдельными предприятиями, станциями и пристанями. Такой подход ближе к нашему исследованию, чем более поздние советские районирования, когда авторы часто не имели подробной информации. В.И. Чаславский (1873), вероятно, первый разделил хлебопроизводящие районы Центральной России и Черноземья на три части по принципу основного направления вывоза зерна — тяготеющие к Москве и Санкт-Петербургу, к Риге, к Нижнему Новгороду и Рыбинску. Д.И. Рихтер (1903) выделил на территории Российской империи для Совета съездов русских мукомолов 16 районов. Каждый из них объединяет мельницы с одинаковым рынком сырья и сбыта.

Ключевая работа по промышленному районированию в дореволюционной России — “Торговля и промышленность Европейской России по районам” В.П. Семенова Тянь-Шанского. Тер-

Таблица 1. Сравнительная характеристика ТПК и кластеров

Характеристика	ТПК	Кластер
Экономическая система	Плановая	Рыночная
Структурная основа	“Решетка” района	Форма сети района
Появление	Целенаправленно	Стихийно
Цель/эффект от существования	Комплексность районного хозяйства, рациональное сочетание производительных сил	Конкурентная борьба, повышение производительности, определение направлений и интенсивности организации нового бизнеса, внедрение инноваций
Конкуренция	Нет	Есть
Кооперация	Есть	Есть
Особенности отраслевого состава	Преобладание промышленности	Произвольный
Размер предприятий	Произвольный	Произвольный
Территориальный размер	До экономического района	От одного города или региона до страны или нескольких стран
Роль связей между предприятиями в сравнении с географической близостью	Определяющая	Определяющая, но допускаются оговорки
Роль государственной политики	ТПК — результат реализации советских экономических планов, но, часто побочный	Правительство должно способствовать разрастанию имеющихся и возникающих кластеров, а не пытаться создавать совершенно новые

ритория разделялась на 1065 районов по принципу тяготения к крупным торговым и промышленным пунктам с оборотом не менее 50 тыс. рублей. В одной зоне тяготения может быть несколько районов, если есть различия в структуре экономики разных частей зоны (Торговля ..., 1900).

А.Е. Пробст (1939) нанес на карту районы потребления угля в СССР. Исследуя его перевозки по железной дороге, он выделил на территории страны ареалы, куда поставляется уголь из одного и того же бассейна. И.И. Белоусов (1958) занимался визуализацией графов перевозок важнейших железнодорожных грузов, в том числе круглого леса и пиломатериалов вручную на бумаге. Однако, вершинами в его графах были не отдельные предприятия или железнодорожные станции, а энерго-производственные районы Н.Н. Колосовского.

Сегодня инструменты анализа данных позволяют более строго и точно подойти к узловому районированию по торговым связям. Этому способствуют большие массивы открытых экономических данных. За рубежом есть несколько современных работ, где применяется кластеризация графа экономических связей.

П. Валерио с соавторами (Valerio et al., 2020) на примере продажи живых животных (козы, овцы, ослы) между странами Западной Африки составили граф, где узлы – скотные рынки, ребра – факт продажи живых животных, объем продаж – вес ребра. Источник данных – CILSS (2013–2017 гг.). С помощью алгоритма жадной

оптимизации модулярности были выделены два типа торговых кластеров: внутривосточные и межвосточные.

Я. Хи с соавторами (He et al., 2021) изучали распространение некачественных продуктов в Китае с 2014 по 2019 г. Использовалась база данных результатов проверок безопасности пищевых продуктов, в которых у проверяемых партий указана точка отправления и пункт назначения. Это близкая к случайной выборка из пищевой промышленности Китая. Вершины графа – города, ребра – наличие торговли между производителем и дистрибьютором некачественных продуктов. Выделено 13 кластеров, большинство из которых формируется вокруг городов (Чжэнчжоу, Пекин, Чунцин и др.), имеющих наиболее высокую центральность в графе. Кластеры часто находятся в пределах одной или нескольких провинций, причем их границы совпадают с административными рубежами.

Кластеризацию графов используют не только в экономической, но и в социальной географии (табл. 2). Диапазон тем для исследования весьма широк – виртуальное общение, телефонные разговоры, миграции, научные связи, интеллектуальная собственность.

Исходные детальные данные обычно агрегируют. Часто рассматривают графы, где вершины – города, но, например, не отдельные пользователи соцсети. Это позволяет снизить размер данных – вершин и ребер в графе становится намного меньше, расчеты выполняются

Таблица 2. Примеры кластеризации графов в социальной географии

Источник данных	Вершина графа	Вес ребра в графе	Алгоритм кластеризации	Ссылка
Венгерская соцсеть iWiW	Города	Число дружеских связей между пользователями двух городов	Лувена	Lengyel et al., 2015
Обзор пригородных маятниковых миграций (ACS)	Переписной район	Число поездок на работу между двумя переписными районами	Визуально, Комбо	Dash and Rae, 2016
Сообщества социальной сети “ВКонтакте”	Крупные села Тверской области, Москва, Санкт-Петербург, Тверь	Доля подписчиков в виртуальном сообществе села из Твери, Москвы или Санкт-Петербурга	Визуально	Смирнов и др., 2019
Домашние телефонные звонки в Великобритании	Центры ячеек 9.5 × 9.5 км, на которые разбита страна	Продолжительность переговоров в часах между абонентами из пары ячеек	Оптимизация модулярности	Ratti et al., 2010
Базы научного цитирования	Страны и/или города	Число совместных публикаций	Визуально	Leydesdorf, 2013; Olechinka et al., 2018
База данных WIPO	Города	Число совместных патентов	Визуально	World ..., 2019

быстрее, меньше требуется вычислительных ресурсов. Зачастую агрегирование требует постановка задач исследования. Если граф небольшой или его можно отразить на карте, то кластеры иногда выделяют визуально. При машинной кластеризации чаще всего используются алгоритмы на основе оптимизации модулярности. Самый популярный из них — алгоритм Лувена. Многие исследователи выявляют, насколько полученные кластеры соответствуют местной сетке административно-территориального деления. Очень близки к границам регионов и медье кластеры дружбы пользователей, выделенные в венгерской соцсети iWiW. Похожи на ячейки уровня NUTS2 кластеры телефонных переговоров Великобритании. Кластеры маятниковых миграций в США существенно отличаются от штатов и графств. Выделяется два типа сетевых структур графа: вертикальные и горизонтальные. Например, в США, Германии, Республике Корея, Англии, Индии внутренние связи патентной активности распределены равномерно между основными центрами НИОКР. Во Франции, Испании, Японии вертикальная структура связей — они концентрируются в столицах государств.

Граф — удобная абстракция для исследования сетевых структур. В настоящей работе также используется кластерный анализ графа.

ДАННЫЕ И МЕТОДИКА ИССЛЕДОВАНИЙ

Важнейший источник данных настоящего исследования — ЕГАИС учета древесины, в которую обязаны предоставлять информацию о сделках юридические лица и индивидуальные предприниматели, торгующие древесиной. Часть данных открыта: номер сделки, ИНН и название продавца и покупателя, дата и объем сделки в м³. Вид торгуемой продукции в открытом доступе отсутствует. Подлежат декларированию лесоматериалы (необработанные бревна) и пиломатериалы (доски, брусья). Не подлежат декларированию фанера, древесные плиты, целлюлоза и бумага, щепа и стружка, пеллеты и брикеты, клееные деревянные конструкции.

Мы использовали информацию о сделках за 2020 г. и, вероятно, за предыдущие годы (точно сказать нельзя). Существуют сделки с открытой датой, в которые можно приплюсовывать новые объемы торговли, чтобы каждый раз не оформлять новую декларацию². Такие сделки мы называли плавающими — при поставках новых партий их объем складывается с уже накопленным, а дата обновляется, сделки постоянно обновляют свою дату, как бы “молодеют”. Плавающие сделки

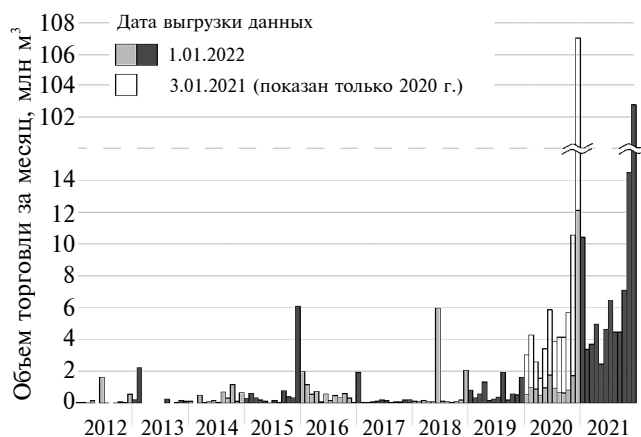


Рис. 1. Динамика торговли древесиной в России по данным ЕГАИС учета древесины.

актуальны для компаний, имеющих стабильное долгосрочное сотрудничество. В исследовании использована выгрузка на 3 января 2021 г. Были отобраны те сделки, у которых дата относилась к 2020 г., среди них есть плавающие сделки из предыдущих лет, поэтому нельзя обозначить временные рамки исследования строго 2020 г. Для кластеризации это неплохо — долговременные постоянные торговые связи будут дополнительно выявлены. Долю объемов плавающих сделок можно оценить на рис. 1 — порядка 3/4 всех объемов “уплыли” за год из 2020 в 2021 г. Пики в декабре, возможно, связаны с тем, что компании предпочитают по каким-то причинам ограничивать плавание сделки в конце года и в новом году открывать новую. Схожая причина у некоторых пиков в конце полугодия или квартала.

В ЕГАИС учета древесины у сделки указано два значения — задекларированный объем со стороны продавца и покупателя, так как по закону³ подавать декларацию в ЕГАИС обязаны обе стороны. Мы пользовались объемами, указанными продавцом — они более полные (сумма по ним больше, чем по объемам, указанным покупателем). Иногда покупатели не декларируют объем сделки, хотя бывает и наоборот, но намного реже. Много сделок (порядка половины), где задекларированы нулевые объемы. Судя по всему, это сделки, которые не были завершены, хотя обе стороны собирались это сделать.

Для каждого предприятия из данных ЕГАИС учета древесины по ИНН были получены данные ЕГРЮЛ (реестр юридических лиц) или ЕГРИП (реестр индивидуальных предпринимателей) с сайта ФНС⁴ в виде pdf-документов. Из pdf-выписок были взяты адрес, код ОКВЭД основного вида деятельности, а также собствен-

² <https://lesegaits.ru/pomosh-i-podderjka> (дата обращения 07.04.2022).

³ Статья 50.6 Лесного кодекса Российской Федерации.

⁴ <https://egrul.nalog.ru/> (дата обращения 07.04.2022).

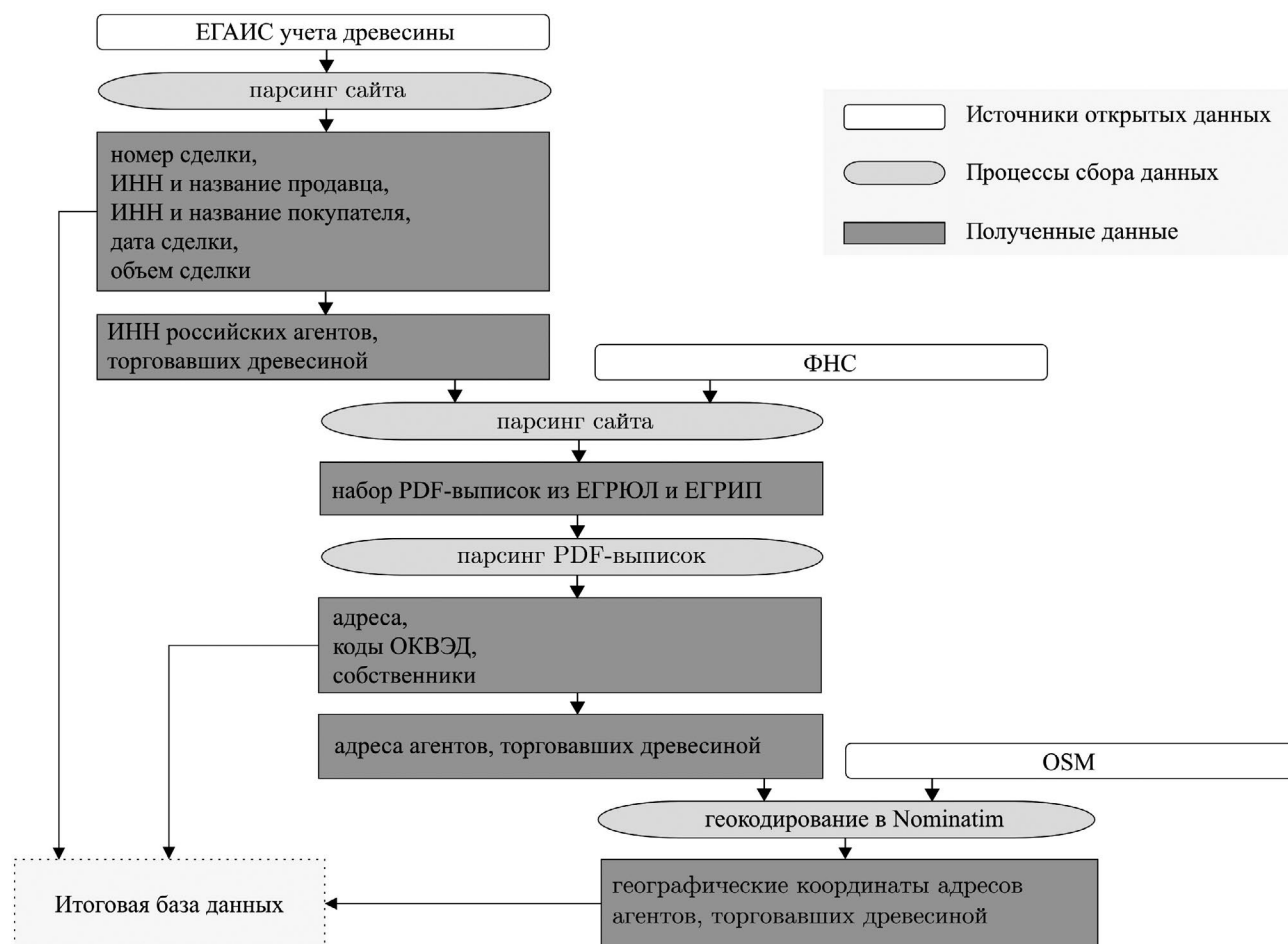


Рис. 2. Схема сбора данных.

ники предприятий. Для всех адресов было выполнено геокодирование⁵ в OSM Nominatim⁶. Общая схема сбора информации показана на рис. 2. Парсинг⁷ сайтов выполнен с помощью инструмента для автоматизации браузера Selenium через обращение к элементам html-разметки сайтов на языке XPath. Для чтения pdf-выписок был написан скрипт, собирающий нужную информацию в тексте по определенным маркерам.

Таблица сделок ЕГАИС учета древесины — это граф, заданный в табличной форме, где каждая строка — ребро. Чтобы не было нескольких ребер между двумя одинаковыми вершинами, взята сумма всех сделок между одной и той же парой компаний, если они задекларировали несколько сделок (не все пользуются плавающими сделками). Вершины в графе — это продавцы и покупатели древесины. Вес ребра — объем сделки в м³. Для понимания логики представления таблицы в виде графа можно сравнить табл. 3 и рис. 3. В реальном графе для идентификаторов вершин использовались не названия

предприятий, а ИНН (поскольку названия часто заполняются с ошибками). Были исключены сделки с нулевыми объемами, а также экспортные и импортные. В итоге сформирован граф, где примерно 16 тыс. вершин (агентов) и около 26 тыс. ребер (сделок).

Было опробовано более 20 разных вариантов методики кластеризации. Для сравнения использовались три критерия: замкнутость кластеров (доля внутрикластерного оборота предприятий кластера от общероссийского оборота), модулярность кластеризации (это показатель корректности деления граф на предлагаемый вариант кластеризации), соответствие содержания кластеров экономическим процессам лесной промышленности. Итоговый вариант — комбинация из нескольких алгоритмов, ключевые из которых — алгоритм Лейдена (Traag et al., 2019) и укладка Фрюхтермана–Рейнгольда (Fruchterman and Reingold, 1991) (рис. 4).

Алгоритм Лейдена — способ кластеризации графа на основе оптимизации функции модуляр-

⁵ Геокодирование — процесс перевода текстового адреса в географические координаты.

⁶ <https://nominatim.openstreetmap.org/> (дата обращения 30.07.2024).

⁷ Парсинг — автоматизированный сбор данных.

Таблица 3. Граф в табличной форме

Продавец	Покупатель	Объем сделки, м ³
ООО “Толстое бревно”	ЗАО “Длинная доска”	1000
ООО “Толстое бревно”	ПАО “Честный купец”	100
ЗАО “Длинная доска”	ИП Дуболомов	200
ЗАО “Длинная доска”	ФГБУ “Уютный барак”	800

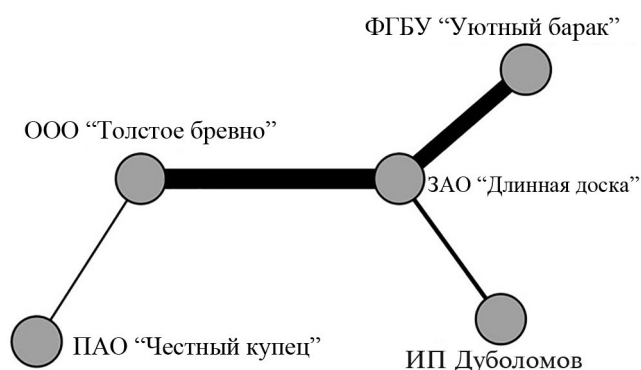


Рис. 3. Визуализация графа табл. 3 (толщина ребер пропорциональна объему сделки).

ности. Он разработан в 2019 г. на базе алгоритма Лувена (2008 г.), который, в свою очередь, потомок алгоритма быстрой жадной кластеризации (2004 г.). Рассмотрим работу алгоритма на основе оптимизации модулярности на примере алгоритма Лувена — он весьма похож на алгоритм Лейдена и проще для изложения.

Модулярность двух вершин в невзвешенном графе — разность реального числа ребер между парой вершин и ожидаемого числа ребер между ними в случайном графе с таким же распределением степеней вершин. Модулярность графа — сумма модулярностей каждой пары вершин (1), между которыми есть ребра.

$$Q = \frac{1}{2m} \cdot \sum_{vw} \left(A_{vw} - \frac{k_v \cdot k_w}{2m} \right), \quad (1)$$

где m — число ребер в графе, vw — пара вершин, связанных ребром, A_{vw} — число ребер между парой вершин (как правило, равно 1), k_v — число ребер, связанных с вершиной v , k_w — число ребер, связанных с вершиной w . Вычитаемое в формуле (1) — вероятность появления ребра между рассматриваемыми вершинами.

Модулярность можно рассчитывать и для нескольких кластеров, состоящих из множества вершин. Кластеры рассматриваются как вершины с ребрами-петлями, чей вес равен удвоенной сумме числа внутрикластерных ребер. Вес соединяющего ребра равен числу ребер, соединяющих вершины двух кластеров. В случае взвешенного графа число ребер заменяется на сумму их весов.

Модулярность используют для оценки уровня выраженности кластеров всей кластеризации графа. Она меняется от -1 до 1 . Если модулярность ближе к 1 , то кластеры хорошо выражены, ребра графа концентрируются внутри кластеров.

На первом шаге работы алгоритма Лувена каждая вершина является самостоятельным кластером. Каждый шаг состоит из двух этапов. На первом этапе каждую пару вершин, соединенных ребром, пробуем объединить в один кластер — рассчитываем приращение модулярности от объединения. Среди всех вариантов объединения выбираем для вершины тот, где приращение модулярности будет максимальным и положительным. Если положительных приращений нет, то объединения не происходит. Наборы вершин, которые нужно объединить, становятся предварительными кластерами. На втором этапе выполняем пересборку графа: превращаем предварительные кластеры в вершины с ребрами-петлями, чей вес равен удвоенной сумме числа внутрикластерных ребер. Новые вершины соединяем ребрами с весом, равным числу межкластерных ребер. Алгоритм заканчивает работу на том шаге, когда все рассчитанные приращения модулярности получаются отрицательные.

Кластеризация невзвешенного графа с помощью алгоритма Лувена показана на рис. 5, где при расчете приращения модулярности разность не умножается на $1/2m$ — это не влияет на относительные различия значений и не приводит к смене знаков.

Работа алгоритма Лейдена несколько сложнее. Во-первых, в нем есть третий этап — уточнение промежуточных кластеров. Во-вторых, в нем оптимизирован порядок расчета приращения модулярности, и алгоритм Лейдена работает быстрее. При его использовании выделенные кластеры являются в определенном смысле оптимальными, а также внутренне связанными.

Применение алгоритма Лейдена, как и других алгоритмов на основе оптимизации модулярности, связано с проблемой разрешающей способности: некоторые мелкие и средние кластеры он объединяет в более крупные. Для решения этого вопроса каждый лейденский кластер мы кластеризовали еще раз с помощью силовых алгоритмов укладки графов. Работа этих алгоритмов основана на имитации действия физических сил. Представим, что вершины гра-

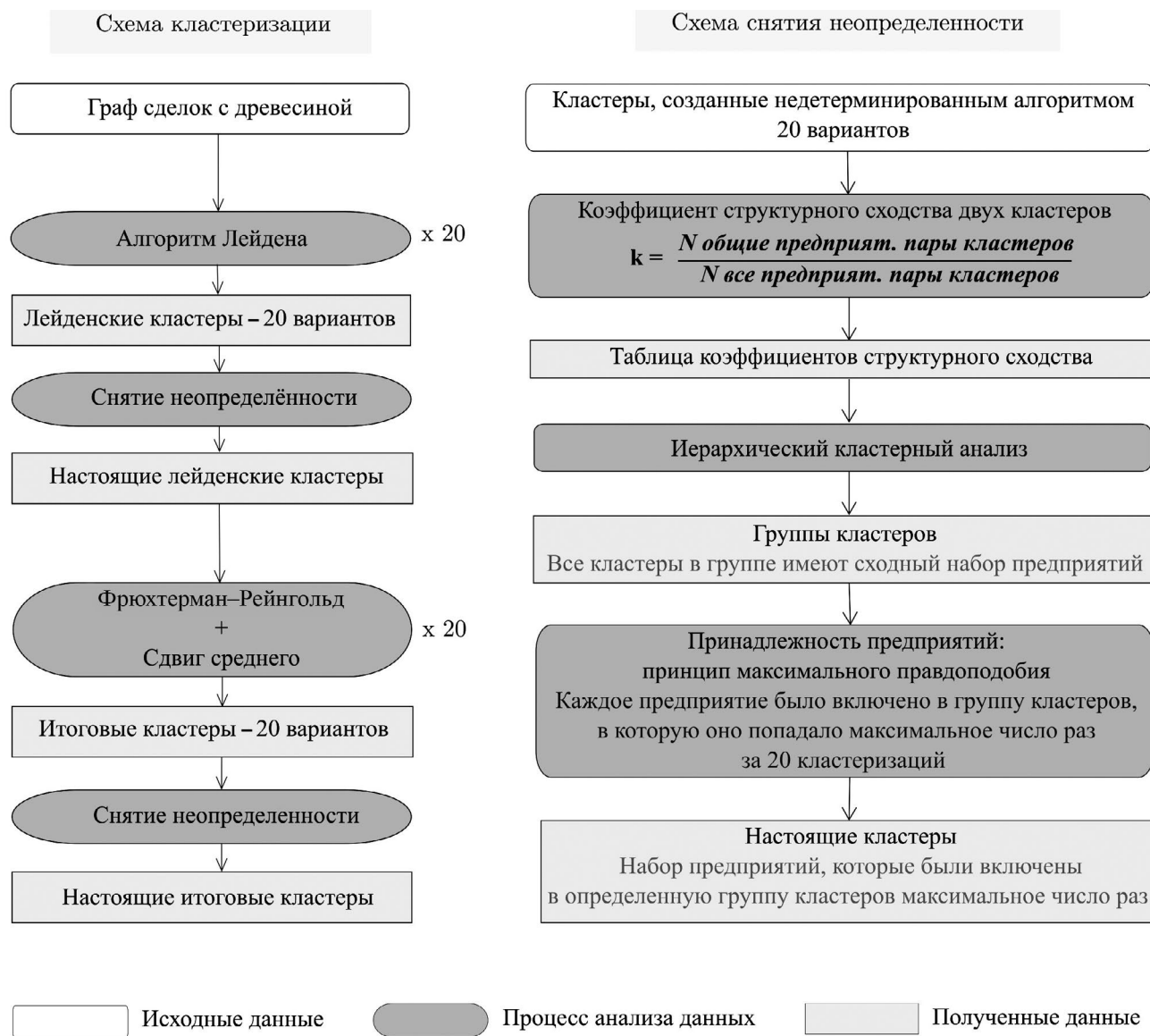


Рис. 4. Методика кластеризации.

фа — это стальные шары с одинаковыми электрическими зарядами одного знака. Те шары-вершины, между которыми есть ребра, соединяются пружинами. Затем все шары раскладывают на какой-нибудь ограниченной плоскости в случайном порядке, при этом положение каждого шара жестко фиксируется. После того, как все шары разложены, их отпускают, вся система сразу приходит в движение. Вершины, которые связаны ребрами-пружинами, будут притягиваться друг к другу и образовывать скопления — кластеры. Несвязанные вершины станут отталкиваться за счет электростатических сил, которым нечего противопоставить, когда нет ребра.

Для извлечения кластеров с укладки мы использовали довольно простой алгоритм плотностной кластеризации — сдвиг среднего (Fukunaga and Hosteler, 1975). Обычно укладка

графа применяется для визуализации, но вместе со сдвигом среднего это и неплохой инструмент выделения кластеров.

Алгоритм Лейдена и укладка Фрюхтермана–Рейнгольда — недетерминированные — при запуске алгоритма с теми же настройками на тех же данных результаты могут немного отличаться в деталях. В нашем случае это приводило к тому, что разные кластеры то объединялись, то делились на части. Например, иногда Смоленский и Брянско-Калужский кластеры сливались друг с другом, а Костромской, наоборот, порой делился на две части. За истинный вариант мы принимали наиболее часто встречающийся. Например, если из 10 запусков Костромской кластер делится на части 3 раза, но остается целым 7 раз — он все-таки будет считаться целым. Для снятия

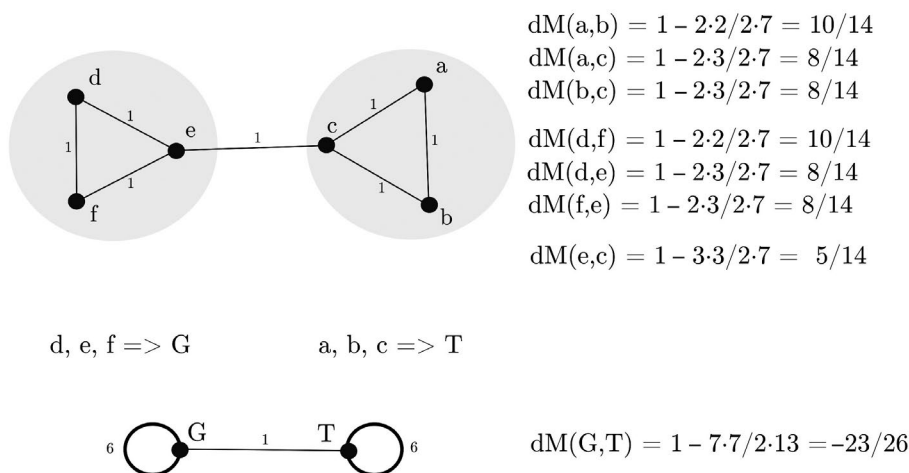


Рис. 5. Алгоритм Лувена.

этой неопределенности был разработан специальный подход:

— запустить алгоритм кластеризации несколько десятков раз;

— посчитать коэффициент структурного сходства между всеми кластерами всех запусков (рис. 6);

— выделить с помощью иерархической кластеризации группы кластеров со сходным составом;

— отнести каждую вершину графа к той группе кластеров, куда она входила максимальное число раз.

В качестве коэффициента структурного сходства между двумя кластерами мы использовали расстояние Жаккара — отношение числа общих предприятий для обоих кластеров ко всему числу предприятий в этих кластерах.

Алгоритм Лейдена запускался 20 раз подряд, кластеризация лейденских кластеров — 50 раз подряд. Лучше запустить как можно больше раз

подряд, но матрицы расстояний получаются слишком большими для иерархического кластерного анализа. По этой же причине каждый раз брались 200 крупнейших леденских кластеров. Остальные несколько сотен кластеров — это небольшие группы из нескольких предприятий, обычно полностью изолированных от основной части компаний. После 20 запусков Лейденского алгоритма и снятия его неопределенности взяты 100 крупнейших кластеров. Для каждого из этих кластеров 50 раз запустили укладку и сдвиг среднего с последующим снятием неопределенности, в итоге получилось 246 кластеров. Мы анализировали в своей работе 100 крупнейших из них — 95% всей суммы исходных сделок.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

При работе с кластерами удобно пользоваться двумя видами оборотов. Внутрикластерный оборот кластера — сумма сделок между предприятиями внутри кластера. Внутривосский оборот кластера — сумма внутрикластерных и межкластерных сделок, в которых участвовали предприятия кластера. Важнейший формальный показатель уровня кластеризации с точки зрения торговли — замкнутость кластера, т.е. отношение внутрикластерного оборота к внутривосскому. Чем выше замкнутость, тем лучше выражены и очерчены выделенные кластеры, ведь сложно назвать кластером группу предприятий с замкнутостью в 50–60% или менее. Ключевой показатель уровня кластеризации с точки зрения геометрической структуры сети — модулярность, которая показывает, насколько ребра концентрируются внутри кластеров.

Весь исходный объем сделок (153.3 млн м³) составляет примерно 62% от всей заготовленной древесины (219 млн м³) и произведенных пи-

		Кластеризация 1		Кластеризация 2	
		Кластер 1	Кластер 2	Кластер 1	Кластер 2
Кластеризация 1	Кластер 1	1	0	0.9	0.1
	Кластер 2		1	0.2	0.8
Кластеризация 2	Кластер 1			1	0
	Кластер 2				1

Рис. 6. Пример таблицы коэффициента структурного сходства (две кластеризации, в каждой два кластера).

Таблица 4. Основные показатели кластеризации

Показатель	Значение, млн м ³ (%)
Вся исходная сумма сделок	153.3
Вся исходная сумма сделок без экспорта	111.3 (100%)
Внутрироссийский оборот 264 кластеров	108 (97%)
Внутрироссийский оборот 100 крупнейших кластеров	105.9 (95%)
Внутрикластерный оборот 100 крупнейших кластеров	94.8

ломатериалов (29 млн м³) за 2020 г., хотя такое сравнение не вполне корректно из-за наличия плавающих сделок прошлых лет. Реальная доля торгуемой древесины от всех лесоматериалов и пиломатериалов, произведенных за год, вероятно, составляет первые десятки процентов (до половины).

В процессе снятия неопределенности при кластеризации из 111.3 млн м³ сделок были потеряны 3% оборотов, при отборе 100 крупнейших кластеров из 264 получившихся — еще 2%. В итоге детально проанализировано 95% исходных данных, на которых создавался граф (табл. 4).

Средняя замкнутость кластеров высока — 89%, почти все торгуемые объемы замкнуты внутри выделенных групп, среди сотни крупнейших нет ни одного, чья замкнутость была бы ниже 70%. Модулярность тоже имеет высокое значение — 0.86.

Всего выделено 246 кластеров. Самый крупный имеет размер — 9.6 млн м³. Средний размер среди 100 крупнейших — 1.05 млн м³, медиана — 0.26 млн м³ (указаны внутрироссийские обороты).

Территориальная структура лесоторговых кластеров России состоит из трех крупных ареалов — Европейского, Сибирского и Дальневосточного (рис. 7а). Крупнейшие по обороту кластеры расположены на северо-западе России, у них самая низкая замкнутость. Все кластеры Восточной Сибири имеют крайне высокую замкнутость, по размеру несколько меньше европейских. Дальневосточный ареал самый маленький. Есть еще совсем небольшой ареал на северном Кавказе, он в статье не рассматривается.

Кластеры на карте действительно образуют довольно хорошо очерченные ареалы из населенных пунктов — лесоторговые районы (рис. 7б). Их границы часто совпадают с границами регио-

нов, хотя и не всегда точно. Один кластер может объединять два или три региона вместе, а бывает, наоборот, регион состоит из нескольких кластеров.

В Карелии представлено 2 кластера. На севере расположен кластер крупнейшего лесопромышленного холдинга России — компании “Сегежа Групп” (ключевое предприятие — “Сегежский ЦБК”). На юге республики — группа между Ладожским и Онежским озерами, сложившаяся вокруг нескольких крупных лесопилок (“Соломенский лесозавод”, “Сортавальский лесозавод”, “Шуялес”, “Сетлес”) и их лесозаготовителей.

Важнейший кластер в границах Архангельской области образуют в первую очередь предприятия холдинга “ПКП Титан” (основные активы — “Архангельский ЦБК” и “Лесозавод 25”).

Лесная промышленность Ленинградской и Вологодской областей тесно связаны. Крупнейший кластер Санкт-Петербурга (рыжий⁸) относится к нему формально — в нем зарегистрирована «Группа “Илим”». Реально ключевой кластер Санкт-Петербурга и всего северо-запада — синий. Это очень плотно связанная горизонтальная сетевая структура, где важнейшую роль играют огромные оптовые торговцы древесиной (“ЛесТрейд-Экспорт”, “Рослеском”, “Стрит”, “Лесная Рапсодия”, “ЮПМ-Кюммене”), собирающие лес со всего северо-запада России на экспорт и для крупных российских производств. На карте заметна концентрация мелких и средних лесозаготовительных и торговых предприятий этого кластера к крупнейшим железным дорогам. Важнейший кластер Ленинградской области сформировался вокруг компании “Интернешнл Пейпер” (Светогорский ЦБК). Этот кластер имеет “филиал” в Вологодской области, самые большие поставщики балансов Светогорского ЦБК находятся в этом регионе: «Группа компаний “Белая Сосна”», “Череповецлес”, “Вологодский лес”. В восточной половине Вологодской области расположен кластер “Группы Илим”. Сибирские ЦБК этой компании практически не закупают на стороне древесину. ЦБК в Коряжме имеет обширные связи, собирая древесину у сотен мелких и средних компаний, расположенных на стыке Вологодской, Архангельской, Кировской областей и в Республике Коми. Собственно вологодский кластер занимает совсем небольшую площадь, хотя оборот у него велик. Важнейшую роль в нем играют компании двух комплексных холдингов: “Вологодские лесопромышленники” и “Альтернатива”.

Кластер Республики Коми состоит из двух частей. На территории региона находится основная часть, сформированная вокруг ЦБК “Монди СЛПК” и крупных сыктывкарских

⁸ Здесь и далее — цвета пунсонов на рис. 7.

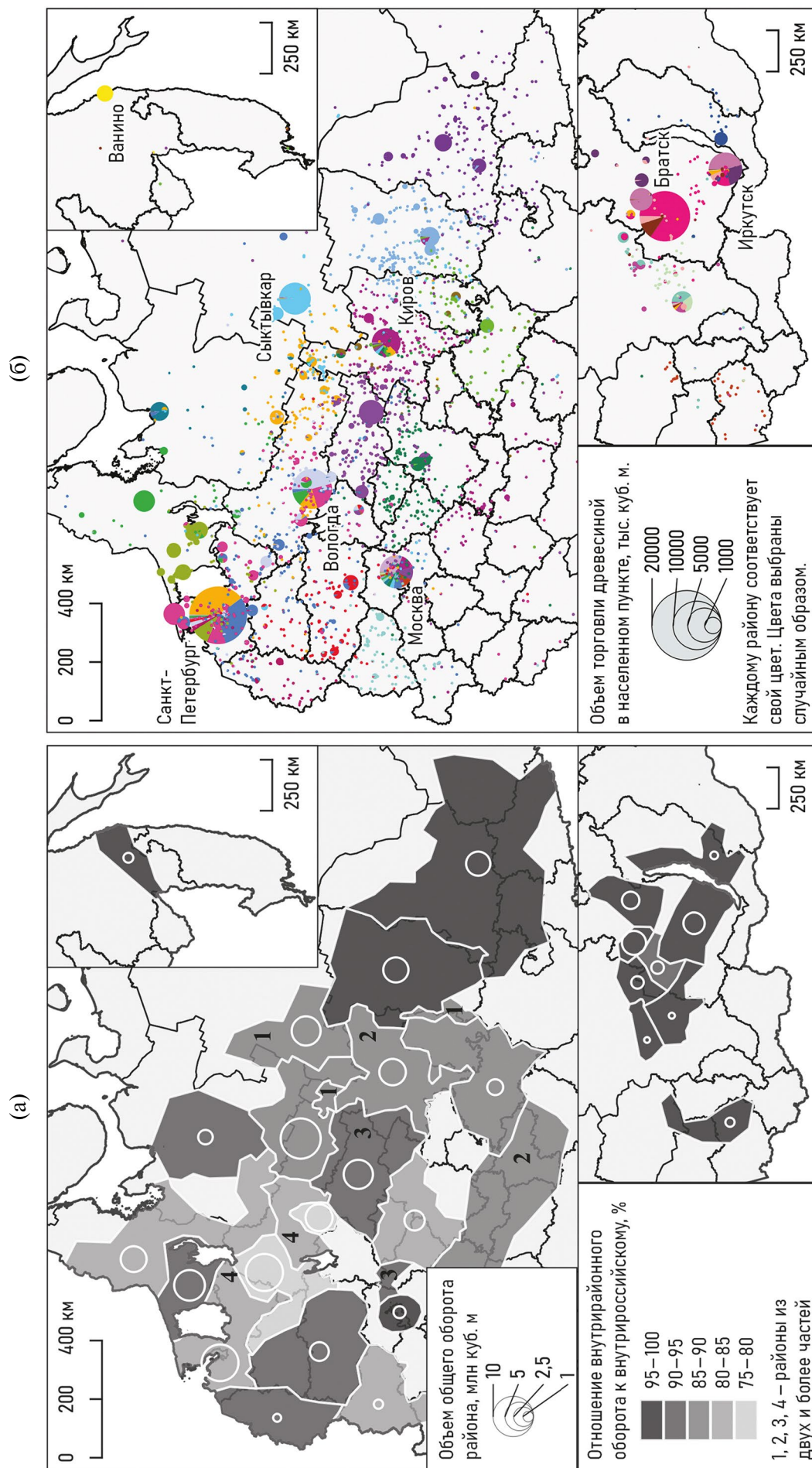


Рис. 7. Размер и замкнутость крупнейших лесоторговых кластеров в России (а), структура оборота древесины населенных пунктов по крупнейшим лесоторговым кластерам России (б), 2020 г.

лесопилок. Другая его часть, расположенная в Подосиновском районе Кировской области, состоит из множества небольших компаний, среди которых можно выделить “Демьяновский завод ДВП”. Обе части соединяет Жешартский ЛПК — крупный производитель фанеры и древесных плит. Интересно, что обе части относительно изолированы как на графе связей, так и на карте.

В южной половине Псковской области расположен кластер холдинга «ИФК “Нева”». Это проект комплексного освоения лесов (заготовка, лесопиление, трейдер), организованный компанией GS Group (владелец — А. Ткаченко) в рамках диверсификации бизнеса. Ключевые активы GS Group расположены в Гусеве (Калининградская область) — производят микроэлектронику. А. Ткаченко — вероятный собственник компании “Триколор-ТВ” (крупнейший оператор спутникового ТВ в России).

Это максимально нестандартный пример диверсификации бизнеса, когда из микроэлектроники ушли в лесную промышленность, причем очень успешно — тот, кто делает жесткие диски и печатные платы теперь структурирует лесную промышленность целого региона.

В кластере Тверской области большую роль играет лесозаготовительный холдинг “Новая транспортная компания”. Важный узел — компания “СТОД” — самый большой производитель LVL-бруса в России.

Смоленский кластер имеет два выдающихся центра — “Эггер Древпродукт Гагарин” (древесные плиты) и “Смоленский ДОК” (фанера). Производство фанеры и плит — основная отрасль специализации кластера.

Кластер Костромской области можно разделить на две части. Предприятия восточной части региона тяготеют к фанерному комбинату “Свежа Мантурово” и особенно к огромному плиточному заводу “Свисс Крона” (швейцарский капитал) в Шарье. Компании западной части больше связаны с фанерным комбинатом “Свежа Кострома” и десятком других относительно крупных фирм.

Нижегородская, Ивановская и Владимирская области входят в один большой кластер. Нижегородская его часть — вертикальная конструкция, где доминирует ЦБК “Волга”, расположенный в Балахне. Владимирско-Ивановская часть — горизонтальная, основное ее предприятие — “Эггер Древпродукт Шуя” (производство древесных плит), а также несколько крупных лесопилок.

Кластер Кировской области — очень плотно связанная горизонтальная структура. Отраслевая структура сбалансированная, ярко выраженных центров нет. Среди ключевых компаний — “Алмис” (лесопиление), “Красный якорь” (фане-

ра), “Сибирь-Лес” (заготовка и воспроизводство леса), “НЛК” (древесные плиты), несколько крупных торговых предприятий, перераспределяющих древесину внутри кластера (“Терминал Вятка”, “Регион Лес”, “Абарис” и др.). На графе и на карте на периферии кластера выделяется группа предприятий вокруг фанерных заводов в Пензенской области (“Власть труда”) и Республике Мордовия (“Плайтерра”), эти заводы закупают существенную часть сырья в Кировской области.

Кластер Татарстана и Удмуртии весьма необычен. Доминирует здесь завод древесных плит “Кастамону Интегрейтед Вуд Индастри” (турецкий капитал). Почти все остальные предприятия кластера — это районные лесничества Республики Татарстан, обеспечивающие сырьем завод “Кастамону” и не имеющие никаких связей с другими предприятиями страны. Такую же функцию выполняет “Удмуртлес”.

Пермский край, наоборот, похож на Кировскую область — горизонтальная плотная структура, где представлены почти все направления лесной промышленности. Ключевые предприятия (ЦБК “Соликамскбумпром”, фанерный комбинат “Свежа Уральский”) формируют два слабо заметных сгущения в рамках кластера.

Свердловская область тоже имеет горизонтальную сетевую структуру. На общем фоне выделяется фанерный комбинат “Свежа Верхняя Синячиха”.

Томский кластер формируется вокруг производителя древесных плит “Латат” и включает его поставщиков древесины из Алтайского края.

Кластеры Красноярского края и Иркутской области похожи друг на друга. Там мало горизонтальных и вертикальных конструкций, почти все кластеры — это малочисленные цепочки из отдельных компаний или их небольших групп, часто древовидные. Такие сетевые структуры мы назвали дендритными. Отраслевой состав их похожий — основу составляют одна или несколько крупных лесопилок, много чистых лесозаготовителей (поставщики сырья), широко распространена оптовая торговля (сбыт круглого леса и пиломатериалов в Китай и Европу).

Кластер республики Бурятия собран вокруг Селенгинского ЦКК.

Единственный рассматриваемый дальневосточный кластер состоит из одной компании и ее нескольких спутников. Такие элементарные сетевые структуры мы назвали простыми. “Нью Форест Про” — самый масштабный лесопильный комплекс Дальнего Востока.

Один простой кластер состоит из двух дочек “РЖД”, зарегистрированных в Москве. “ТВС” — производитель шпал и брусьев для стрелок и мостов. Всю продукцию продает «ТД “РЖД”».

ВЫВОДЫ

Разработанная методика кластеризации графа состоит из двух этапов: первичная кластеризация исходного графа с помощью алгоритма Лейдена и вторичная кластеризация каждого лейденского кластера с использованием укладки Фрюхтермана—Рейнгольда. Вторичная кластеризация позволяет решить проблему разрешающей способности, которая свойственна алгоритмам на основе оптимизации модулярности (к ним относится и алгоритм Лейдена). Разработан подход, позволяющий снимать неопределенность кластеризации графа. Применяемые алгоритмы не являются детерминированными, что важно для точности результатов и их воспроизводимости. Кластеры хорошо выражены (средняя замкнутость — 89%), кластеризация имеет высокую модулярность — 0.86. Предприятия на карте образуют ареалы, где преобладает какой-нибудь кластер — лесоторговые районы. Главный недостаток нашей методики кластеризации — потребность в детальных данных по каждому предприятию в масштабе всей страны.

При визуализации графа кластера на плоскости с помощью укладки, можно выделить пять типов сетевых структур кластеров. Вертикальный моноцентрический кластер состоит из одного большого предприятия и множества мелких и средних фирм. Вертикальный полицентрический кластер состоит из нескольких (обычно 2 или 3) больших предприятий и множества мелких и средних фирм. Все связи в вертикальных кластерах реализуются между мелкими или средними фирмами и большим предприятием. Мелкие и средние фирмы почти не связаны между собой. В вертикальных кластерах гораздо меньше конкуренции, чем в горизонтальных, они больше похожи на ТПК Н.Н. Колосовского (см. табл. 1). Горизонтальные кластеры — хорошо связанные группы предприятий разного размера, плотность связей в них часто очень высока, иногда при визуализации заметны подкластеры. Как правило, в таких кластерах присутствует несколько больших предприятий одной и той же отрасли, промышленные компании обычно дополняются торговыми. Конкуренция здесь выше, покупатели могут выбирать продавцов и наоборот. Горизонтальные кластеры похожи на кластеры М. Портера (см. табл. 1). Дендритные кластеры (последовательные цепочки из предприятий или их небольших групп) широко распространены в Сибири и на Дальнем Востоке. Во многих случаях они приурочены к территориям сравнительно нового освоения. Возможно, это ядра будущих горизонтальных и вертикальных кластеров. Простые сетевые структуры состоят из нескольких предприятий,

объединенных в произвольную геометрическую фигуру. Их природа может быть любой.

Есть три важнейших фактора формирования лесоторговых кластеров в России. Первый фактор — производственные цепочки, т.е. прямые продажи и покупки сырья производственными компаниями друг другу без участия торговых посредников. Выделяются четыре основные цепочки: лесопильная, целлюлозная, фанерная и плиточная. Они отличаются видом используемого сырья: для целлюлозы — балансы (тонкая древесина), для пиломатериалов — пиловочник (средняя и толстая древесина), для фанеры — фанкряж (толстая древесина), для плит — балансы и техсырье (древесина любой толщины с дефектами). Второй фактор — сбытовые цепочки (продажи и покупки сырья производственными компаниями посредством торговых фирм). Представлены три основные цепочки: перераспределение древесины внутри кластера, аккумуляция древесины для поставки за пределы кластера или на экспорт, перепродажа друг другу. Третий фактор — общий собственник группы предприятий. Есть кластеры, которые сформировались по другим причинам. Несколько кластеров можно назвать этническими — азербайджанский в республике Коми и условно-армянский в Адыгее (в статье не рассматриваются).

СПИСОК ЛИТЕРАТУРЫ

- Белоусов И.И. Основы межрайонных связей и перевозок. Автореф. дисс. ... д-ра геогр. наук. М.: Географический факультет МГУ, 1958. 32 с.
- Колосовский Н.Н. Теория экономического районирования. М.: Мысль, 1969. 335 с.
- Колосовский Н.Н. Избранные труды. Смоленск: Ойкумена, 2006. 336 с.
- Портер М. Конкуренция. М.: Вильямс, 2005. 608 с.
- Пробст А.Е. Основные проблемы географического размещения топливного хозяйства СССР. М.—Л.: Изд-во Акад. наук СССР, 1939. 404 с.
- Рихтер Д.И. Проект разделения России на районы товарного мукомолья (Сост. по поручению Совета съездов мукомолов). СПб.: Типо-литография И. Лурье и К°, 1908. 88 с.
- Родоман Б.Б. Территориальные ареалы и сети. Смоленск: Ойкумена, 1999. 256 с.
- Смирнов И.П., Виноградов Д.М., Алексеев А.И. К Москве или к Санкт-Петербургу? Тяготение населения Тверской области по данным сети “ВКонтакте” // Изв. РГО. 2019. Т. 151. № 6. С. 69–80. <https://doi.org/10.31857/S0869-6071151669-80>
- Торговля и промышленности Европейской России по районам. Общая часть и приложения / сост. В.П. Семенов-Тян-Шанский, Н.М. Штрупп. СПб.: Типография В.Ф. Киришаума, 1911. 218 с.

- Чаславский В.И. Хлебная торговля в Центральном районе России. Ч. I. Торговля в Примосковском районе. СПб.: Типография В. Безобразова и К°, 1873. 338 с.
- Dash N.G., Rae A. An Economic Geography of the United States: From Commutes to Megaregions // PLOS One. 2016. № 11 (11). Art. e0166083. <https://doi.org/10.1371/journal.pone.0166083>
- Fruchterman T.M.J., Reingold E.M. Graph Drawing by Force-directed Placement // Softw. Pract. Exp. 1991. № 21. P. 1129–1164.
- Fukunaga K., Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition // IEEE Transactions on Information Theory. 1975. Vol. 2. № 1. P. 32–40. <https://doi.org/10.1109/TIT.1975.1055330>
- He Y., Jiang J., Li S. The circulation analysis of substandard foods in China based on GIS and social network analysis // PLOS One. 2021. № 16 (3). Art. e0248037. <https://doi.org/10.1371/journal.pone.0248037>
- Lengyel B., Varga A., Ságvári B., Jakobi Á., Kertész J. Geographies of an Online Social Network // PLOS One. 2015. № 10 (9). Art. e0137248. <https://doi.org/10.1371/journal.pone.0137248>
- Leydesdorf L., Wagner C., Park H., Adams J. Colaboración internacional en ciencia: mapa global y red // Profesional de la Informacion. 2013. Vol. 22. № 1. P. 87–94. <https://doi.org/10.3145/epi.2013.ene.12>
- Olechnicka A., Płoszaj A., Celińska-Janowicz D. The Geography of Scientific Collaboration. NY: Routledge, 2019. 237 p.
- Ratti C., Sobolevsky S., Calabrese F., Andris C., Reades J., Martino M., Rob C., Strogatz S.H. Redrawing the Map of Great Britain from a Network of Human Interactions // PLOS One. 2010. № 5 (12). Art. e14248. <https://doi.org/10.1371/journal.pone.0014248>
- Traag V.A., Waltman L., van Eck N.J. From Louvain to Leiden: guaranteeing well-connected communities // Nature. 2019. № 9. Art. 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- Valerie C.V., Olivier J.W., Marjatta E., Brahim C., Rachata M., Gregory A.K. Network analysis of regional livestock trade in West Africa // PLOS One. 2020. № 15 (5). Art. e0232681. <https://doi.org/10.1371/journal.pone.0232681>
- World Intellectual Property Report. The Geography of Innovation: Local Hotspots, Global Networks. Geneva: World Intellectual Property Organization, 2019. 134 p.

Wood Trade Regions of Russia: Clustering Approach Development

N. A. Sinitsyn^{a,*}, M. S. Elutin^{a,**}, A. S. Korotkov^{a,***}, A. O. Kushlevich^{a,****},
R. R. Menyalshikov^{a,*****}, and D. S. Rusakov^{a,*****}

^aLomonosov Moscow State University, Moscow, Russia

*e-mail: nicksinus@yandex.ru

**e-mail: mikhail.elyutin@student.msu.ru

***e-mail: korotkov.andrey@geogr.msu.ru

****e-mail: artem.kushlevich@yandex.ru

*****e-mail: roman.menyalschikov@yandex.ru

*****e-mail: d_c_rusakov@mail.ru

Network analysis of the Russian timber industry was based on the information about timber transactions between Russian companies in 2020. The data were collected from the Unified State Automated Information System *Accounting Timber and Transactions with It*. The method of graph clustering was developed. The first step is clustering of the hole graph by Leiden algorithm. The second step is drawing of each Leiden created cluster in abstract space using Fruchterman–Reingold layout and extraction of clusters from layout using meanshift algorithm. Such approach helped us to divide many Leiden-created clusters into components. This algorithm has a problem of resolution limit. It tends to combine some medium and small clusters into large ones. Also, a special method was developed to limit the uncertainty of clustering – both Leiden algorithm and Fruchterman–Reingold layout are non-deterministic algorithms. Wood trade clusters vary in size and content. They are rather autonomous. The average share of intra-cluster trade is 89%. The modularity score of our partition is 0.863. Each cluster has been described using open sources from the Internet. The content of the clusters does not contradict the logic of the wood industry processes. The clusters create rather compact areas on the map – wood trade regions. Their borders often overlap with existing administrative boundaries. Five types of network structure of clusters were defined: vertical monocentric, vertical polycentric, horizontal, dendritic, simple. There are three factors of clustering in wood industry of Russia: production chains (4 types – pulp, plywood, lumber, chipboard chains), demand chains (redistribution of raw material between manufacturers within the cluster, accumulation of wood volumes

for some enterprises outside the cluster, wood trade between traders within the cluster), common holder. We have shown that existing approaches to cluster detection based on location quotients often misrepresent economic networks.

Keywords: economic clusters, Russia, wood industry, trade regions, graph clustering

REFERENCES

- Belousov I.I. The fundamentals of inter-district communications and transportation. *Extended Abstract of Doct. Sci. (Geogr.) Dissertation*. Moscow: MSU, 1958. 32 p.
- Chaslavskii V.I. *Khlebnaya trgovlya v Tsentral'nom raione Rossii. Chast' I. Torgovlya v Primoskovskom raione* [Grain Trade in the Central Region of Russia. Part 1. Trade in the Moscow Region]. St. Petersburg: Tipogr. V. Bezobrazova i Ko., 1873. 338 p.
- Dash N.G., Rae A. An economic geography of the United States: From commutes to megaregions. *PLoS ONE*, 2016, vol. 11, no. 11, art. e0166083. <https://doi.org/10.1371/journal.pone.0166083>
- Fruchterman T.M.J., Reingold E.M. Graph drawing by force-directed placement. *Softw. Pract. Exp.*, 1991, no. 21, pp. 1129–1164.
- Fukunaga K., Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory*, 1975, vol. 21, no. 1, pp. 32–40. <https://doi.org/10.1109/TIT.1975.1055330>
- He Y., Jiang J., Li S. The circulation analysis of substandard foods in China based on GIS and social network analysis. *PLoS One*, 2021, vol. 16, no. 3, art. e0248037. <https://doi.org/10.1371/journal.pone.0248037>
- Kolosowsky N.N. *Izbrannye trudy* [Selected Works]. Smolensk: Oikumena Publ., 2006. 336 p.
- Kolosowsky N.N. *Teoriya ekonomicheskogo raionirovaniya* [Theory of Economic Zoning]. Moscow: Mysl' Publ., 1969. 335 p.
- Lengyel B., Varga A., Ságvári B., Jakobi Á., Kertész J. Geographies of an online social network. *PLoS One*, 2015, vol. 10, no. 9, art. e0137248. <https://doi.org/10.1371/journal.pone.0137248>
- Leydesdorf L., Wagner C., Park H., Adams J. Colaboración internacional en ciencia: mapa global y red. *Prof. Inf.*, 2013, vol. 22, no. 1, pp. 87–94. (In Spanish). <https://doi.org/10.3145/epi.2013.ene.12>
- Olechnicka A., Ploszaj A., Celińska-Janowicz D. *The Geography of Scientific Collaboration*. New York: Routledge, 2019.
- Porter M. *Konkurentsia* [On Competition]. Moscow: Vil'yams Publ., 2005. 608 p.
- Probst A.E. *Osnovnye problemy geograficheskogo razmeshcheniya toplivnogo khozyaistva SSSR* [The Main Problems of the Fuel Industry Localization in USSR]. Moscow, Leningrad: Izd-vo Akad. Nauk SSSR, 1939. 404 p.
- Ratti C., Sobolevsky S., Calabrese F., Andris C., Reades J., Martino M., Rob C., Strogatz S.H. Redrawing the map of Great Britain from a network of human interactions. *PLOS One*, 2010, vol. 5, no. 12, art. e14248. <https://doi.org/10.1371/journal.pone.0014248>
- Richter D.I. *Proekt razdeleniya Rossii na raiony tovarnogo mukomol'ya (Sostavlen po porucheniyu Soveta s'ezdov mukomolov)* [The Draft of the Division of Russia into Commercial Flour-milling Areas (Compiled on behalf of the Millers Congresses Council)]. St. Petersburg: Tipogr. I. Lur'e i Ko., 1908. 88 p.
- Rodoman B.B. *Arealy i seti* [Areas and Networks]. Smolensk: Oikumena Publ., 1999. 256 p.
- Semenov Tyan-Shanskii V.P. *Torgovlya i promyshlennosti Evropeiskoi Rossii po raionam. Obshchaya chast' i prilozheniya* [Trade and Industry in European Russia by Region. General Part and Applications]. St. Petersburg: Tipogr. V.F. Kirshbauma, 1911. 218 p.
- Smirnov I.P., Vinogradov D.M., Alekseev A.I. To Moscow or to Saint Petersburg? Population gravity of the Tver Region according to the data of “VKontakte” online network. *Izv. RGO*, 2019, vol. 5, no. 6, pp. 69–80. (In Russ.). <https://doi.org/10.31857/S0869-6071151669-80>
- Traag V.A., Waltman L., van Eck N.J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.*, 2019, no. 9, art. 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- Valerie C.V., Olivier J.W., Marjatta E., Brahima C., Rachata M., Gregory A. K. Network analysis of regional livestock trade in West Africa. *PLOS One*, 2020, vol. 15, no. 5, art. e0232681. <https://doi.org/10.1371/journal.pone.0232681>
- World Intellectual Property Report. *The Geography of Innovation: Local Hotspots, Global Networks*. Geneva: World Intellectual Property Organization, 2019.