Организация Объединенных Наций ECE/CES/GE.41/2023/10

Экономический и Социальный Совет

Distr.: General
27 July 2023
Russian
Original: English

Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по переписям населения и жилищного фонда

Двадцать пятое совещание
Женева, 20–22 сентября 2023 года
Пункт 2 предварительной повестки дня
Уроки, извлеченные из переписей раунда 2020 года

Эстонская перепись населения 2021 года: определение домохозяйств и жилищ по регистрам с использованием графов

Записка Статистического управления Эстонии*[1]

Резюме

Перепись 2021 года стала первой в Эстонии, в которой все обязательные для ЕС переписные характеристики были рассчитаны на основе административных данных. Основной проблемой была низкая, всего 80 %, точность данных о месте жительства в регистре населения и ее влияние на домохозяйства.

При регистровой переписи домохозяйство определяется как совокупность людей, проживающих в одном жилище. При использовании для определения домохозяйств и семей данных о месте жительства из Регистра населения получаемая статистика завышает число одиноких родителей и занижает число супружеских пар.

Для улучшения статистических данных о домохозяйствах и семьях мы разработали графовый метод, использующий данные из административных источников. Мы рассматриваем людей и адреса как узлы графа. Связи между двумя людьми (например, брак, родительство) или между человеком и местом (например, владение недвижимостью) образуют ребра графа. Домохозяйство рассматривается как подграф, содержащий членов домохозяйства и их жилище. Тогда определение домохозяйств и их жилищ эквивалентно нахождению подграфов сильной связности, или, другими словами, обнаружению сообществ.

* Автор: Хелле Виск.
Примечание: Обозначения, используемые в настоящем документе, не подразумевают выражения со стороны Секретариата Организации Объединенных Наций какого бы то ни было мнения в отношении правового статуса той или иной страны, территории, города или района, или их органов власти, или делимитации ее границ.

[1] Настоящий документ представлен с опозданием в связи с задержкой его представления Статистическим управлением Эстонии.

Для поиска связей между людьми или людьми и местами мы использовали данные из 17 регистров. Каждому ребру графа присваивался вес, характеризующий вероятность совместного проживания людей или проживания лица по тому или иному адресу. Вероятностные модели были определены на основе данных о домохозяйствах, полученных в ходе существующих обследований.

Этот новый подход использовался для определения домохозяйств и места жительства в рамках переписи.

I. Введение

1. Перепись 2021 года стала первой переписью в Эстонии, в которой все обязательные для ЕС переменные были получены из административных данных. Система регистров Эстонии хорошо оснащена для проведения регистровой переписи. Регистры охватывают широкий спектр признаков переписи. Кроме того, увязка источников не вызывает затруднений, поскольку мы располагаем уникальными идентификаторами людей, адресов и предприятий.

2. Хотя качество данных в регистрах в целом высокое, имеются некоторые исключения. Например, данные о месте жительства в Регистре населения (РН) точны в случае лишь примерно 80 % людей (Gortfelder & Puur, 2021). Среди причин, по которым люди не обновляют информацию о себе в регистре, можно назвать, в частности, то, что они считают регистрацию ненужной, пользуются услугами и льготами определенного муниципалитета, воспринимают нынешнее жилье как временное (Gortfelder & Puur, 2021; Äär, 2017).

3. В случае регистровой переписи домохозяйство состоит из людей, проживающих по одному адресу, независимо от наличия или отсутствия общего бюджета. Семья определяется в узком смысле как семейная ячейка. Это либо сожительствующая или состоящая в браке пара с детьми или без них, либо одинокий родитель с одним или несколькими детьми. Семья состоит из людей, живущих в одном домохозяйстве. Поскольку место жительства является основой для деления населения на домохозяйства, его неточность сказывается на статистике домохозяйств и семей.

4. Статистическое управление Эстонии провело пробную перепись населения в 2016 году, домохозяйства определялись на основе данных о месте жительства из РН. Статистика домохозяйств и семей существенно отличается от переписи 2011 года. Например, число одиноких родителей по сравнению с переписью 2011 года увеличилось на 67 %, число партнеров уменьшилось на 26 %.

5. Завышение числа одиноких родителей характерно для ситуации, когда члены семьи зарегистрированы по разным адресам. Например, рассмотрим семью из четырех человек: мать, отец, дочь и сын. Если отец и дочь зарегистрированы по разным адресам, то в РН они будут фигурировать как две неполные семьи (мать-сын, отец-дочь).

6. Поскольку статистика домохозяйств и семей, полученная на основе РН, страдает сильным смещением, данные о месте жительства не могут быть использованы для переписи как таковой. Для получения более точной статистики необходимо было найти метод воссоединения семей, которые в регистрах оказались неполными.

II. Методы

A. Домохозяйства или жилища

7. Реконструкция семей — это не просто выявление связей между возможными членами семьи. Рассмотрим пример из пункта 5. Воссоединение этой семьи означает,

что отец и дочь «переезжают» обратно в жилище матери и сына. Отметим, что при этом отец и дочь должны быть отнесены к жилищу, отличному от того, которое указано в РН. Кроме того, состав домохозяйства будет отличаться от РН по всем членам семьи. Таким образом, воссоединение семей предполагает изменение состава домохозяйств и жилищ.

8. Информация, позволяющая определять домохозяйства и их жилища, может быть получена из административных источников. Например, данные о родительстве и браках помогают идентифицировать людей из одной семьи. Для поиска жилищ людей могут быть полезны данные о собственности и договорах энергоснабжения.

9. Неясно, что лучше — начать с определения домохозяйств, а затем закрепить за каждым домохозяйством жилье, или сначала найти жилье для каждого жителя, а затем сформировать домохозяйства из людей, проживающих по одному адресу. Если начать с домохозяйств, то мы не сможем найти семьи, не имеющие прямых связей, например сожителей, совместно владеющих квартирой. С другой стороны, если увязывать людей с адресами, игнорируя данные, которые связывают людей (например, брак), то вряд ли получаемые семьи будут намного более реалистичными, чем в РН.

B. Сетка людей и мест

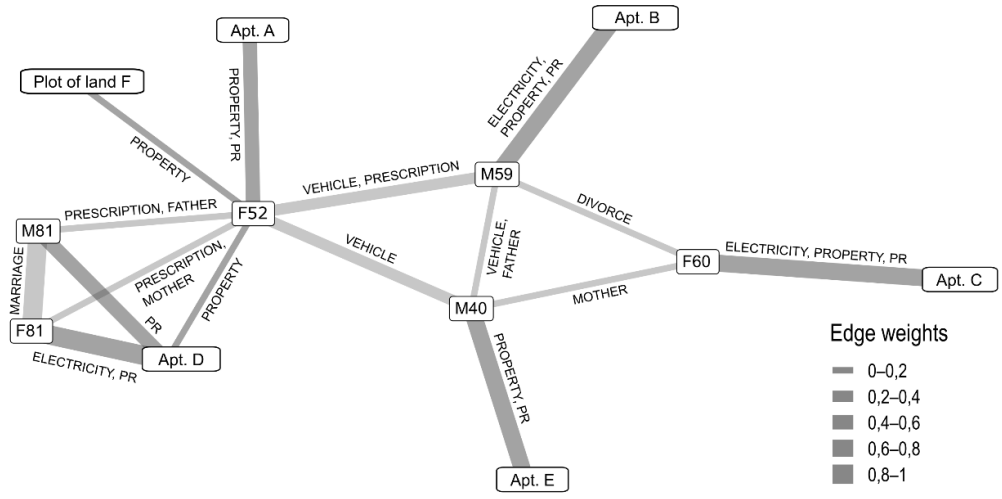
10. Наша идея заключается в одновременном построении домохозяйств и присвоении им жилищ. Для этого мы рассматриваем людей и жилища как узлы графа. Ребра этого графа задаются:

a) связями между людьми (например, брак, родительство, покупка для кого-то лекарства по рецепту, совместное использование автомобиля); или

b) людьми и местами (например, место жительства, собственность, договор энергоснабжения).

Пример такого графа приведен на рис. 1.

Рис. 1
Фрагмент графа людей и мест



Примечание: Узлами являются люди (метка указывает пол и возраст) и места (квартиры и один земельный участок). Ребра соединяют людей (более светло-серый цвет) или людей с местами (более темный серый цвет). Веса ребер показывают вероятность того, что по данному адресу проживает одно и то же домохозяйство или одно и то же лицо. Метки ребер указывают тип связи. ЭЛЕКТРИЧЕСТВО: Лицо имеет договор энергоснабжения по данному адресу, АВТОМОБИЛЬ: люди связаны с одним и тем же транспортным средством как владельцы или пользователи, РЕЦЕПТ: Лицо приобрело лекарство по рецепту для другого, РН: место жительства в регистре населения. Рисунок был опубликован ранее (Tiit et al., 2021).

11. Нас интересует поиск разбиения графа всех людей и жилищ на подграфы, каждый из которых содержит членов одного домохозяйства и их жилище. Естественно предположить, что люди, проживающие в одном домохозяйстве, сильно связаны друг с другом, и люди имеют сильную связь со своим жилищем.

12. В теории графов сообщество определяется как «группа узлов, относительно тесно связанных друг с другом, но слабо связанных с другими группами сильной связности сетки» (Porter et al., 2009). Мы заключаем, что поиск домохозяйств и жилищ подобен поиску сообществ. Выявление сообществ является распространенной задачей при анализе сеток, и существует множество алгоритмов на различных языках программирования.

13. Связи между людьми, или людьми и жилищами, различаются по силе. Например, почти все несовершеннолетние дети проживают с родителями, но это происходит реже, когда ребенок становится совершеннолетним. Следовательно, имеет смысл присваивать ребрам веса.

C. Рабочий процесс

14. Первый шаг — сбор данных. Мы называем различные типы связей «человек-человек» или « человек-место» признаками. Данные о признаках были получены из 17 регистров (таблица 1, таблица 2). Кроме того, в качестве обучающих данных для моделирования связей между знаками и реальными закономерностями были использованы данные крупных ежегодных обследований домашних хозяйств — Эстонского социального обследования/Обследования доходов и условий жизни населения в ЕС (ОДУЖ-ЕС) 2021 года и Эстонского обследования рабочей силы (ОРС) 2021 года.

Таблица 1
Признаки «человек-человек» из регистров

Регистр	*Признаки «человек-человек»*
Э-файл	Лица, находящиеся на одной стороне в споре об алиментах
	Лица, находящиеся по разные стороны в споре об алиментах
Информационная система медицинского страхования	Один человек ухаживал за другим человеком в течение года, предшествующего переписи
Транспортный регистр	Лица связаны с одним и тем же транспортным средством (например, пользователь и владелец автомобиля)
Реестр налогооблагаемых лиц	Лица подали совместную заявку на получение ипотечного кредита
	Один из супругов передал другому супругу необлагаемый налогом доход
	Использование льготы по подоходному налогу на двух и более детей (связь устанавливается между ребенком и лицом, подающим декларацию)
	Лицо, использующее льготу по подоходному налогу для оплаты расходов на образование другого лица
Эстонский центр медицинских рецептов	Лицо приобрело лекарства по рецепту другого лица

Регистр	*Признаки «человек-человек»*
Регистр населения	Лица состоят в браке
	Лица находятся в разводе
	Взрослый выступает в качестве опекуна другого взрослого
	Человек является матерью другого лица
	Человек является отцом другого лица
	Лицо имеет полную опеку над ребенком
	Лицо имеет ограниченную опеку над ребенком
	Ребенок разлучен с родителем
Регистр социальных услуг и пособий	Лица, получавшие пособие по обеспечению прожиточного минимума в одном домохозяйстве
Информационная система социального обеспечения	Лицо получает семейное пособие на ребенка
	Лицо получает родительское пособие на ребенка
	Взрослый получает дополнительный отпуск по уходу за нетрудоспособным взрослым

Таблица 2
Признаки «человек-место» из регистров

	Регистр	*Признаки «человек-место»*
Потенциальные жилища	Элеринг (оператор системы энергоснабжения)	Человек имеет договор энергоснабжения по данному адресу
	Регистр лиц, зарегистрированных в качестве безработных или ищущих работу, и предоставление услуг на рынке труда	Место жительства лица
		Почтовый адрес лица
	Регистр заключенных	Место жительства условно осужденных
	Земельный кадастр	Недвижимое имущество, принадлежащее лицу
	Регистр населения	Зарегистрированное место жительства лица
		Дополнительный адрес лица
		Предыдущие места жительства лица
		Место временного проживания лица (например, общежитие)
	Перепись населения и жилищного фонда 2011 года	Адреса лица и его/ее матери

	Регистр	*Признаки «человек-место»*
	Регистр социальных услуг и пособий	Место жительства лица
	Реестр налогооблагаемых лиц	Недвижимость, приобретенная за счет жилищного кредита лица
Уровень муниципалитета	Эстонская информационная система образования	Детский сад ребенка
		Студент университета или профессионального училища
		Школа ученика в системе общего образования
		Место работы учителя
	Информационная система медицинского страхования	Стоматологическое учреждение, которое посетило лицо
		Медицинское учреждение, которое посетило лицо
		Семейный врач человека (ОП)
	База данных документов, удостоверяющих личность	Место получения документа, удостоверяющего личность
	Регистр обязательных накопительных пенсий	Адрес лица, вступившего в накопительную пенсионную систему
	Эстонский центр медицинских рецептов	Аптека, в которой лицо приобрело лекарственные средства
	Регистр занятых	Место работы лица

15. Веса граней моделировались как:

a) вероятность того, что лица проживают в одном домохозяйстве (была подобрана модель логистической регрессии); или

b) вероятность того, что человек проживает в жилище (алгоритм случайного леса). Последняя модель также включала данные на уровне муниципалитета (например, наличие врача общей практики в определенном муниципалитете), а также расстояния до детского сада, школы и работы.

Модели были подобраны на основе данных обследования и затем применены ко всей совокупности.

16. Всего граф включал 5,2 млн узлов и 7,8 млн ребер. Выявление сообществ проводилось в два этапа:

a) сначала был использован Лувенский метод (Blondel et al., 2008) для разбиения исходного графа на подграфы до 5000 узлов;

b) к каждому из этих подграфов рекурсивно применялся алгоритм Infomap (Rosvall & Bergstrom, 2008) до тех пор, пока сообщества не становились достаточно малыми или модульность не улучшалась существенно.

17. Полученные сообщества были похожи на домохозяйства, а статистика семей улучшилась по сравнению с РН (таблица 3, рис. 2). Однако численность многосемейных домохозяйств и семей со взрослыми детьми была завышена, а численность домохозяйств, состоящих из одного человека, — занижена. Кроме того, небольшое число детей оказалось в домохозяйствах без взрослых.

Таблица 3
Распределение семейного положения людей в обучающих данных

Семейное положение	*Обучающие данные*	*PH*	*Кластеры*
Партнеры, состоящие в браке	34,6	27,6	34,6
Партнеры, совместно проживающие	16,3	10,4	14,5
Одинокие родители	4,2	9,6	4,8
Ребенок, не одинокого родителя	24,2	19,6	27,3
Ребенок, одинокого родителя	5,6	13,5	5,9
Не в семейной ячейке	15,1	19,3	12,9

Примечание: обучающие данные состоят из людей из ОДУЖ-ЕС и ОРС 2021 года.

В таблицу включены данные 32 802 лиц, присутствовавших во всех источниках (исключены нерезиденты, члены институциональных домохозяйств на момент переписи — 00 ч 00 мин 31 декабря 2021 года, лица, родившиеся после этого момента, и умершие ранее). Данные являются невзвешенными.

18. В ходе постобработки дети, живущие одни, добавлялись в домохозяйства родителей или других взрослых родственников. С использованием данных обследования были разработаны эвристические правила, позволяющие разбить некоторые наименее связанные сообщества на более мелкие части (например, если в каком-либо домохозяйстве было несколько семей, мы рассматривали возможность выделения наименее связанной семьи в отдельное домохозяйство).

19. На следующем этапе каждому домохозяйству присваивалось жилище. Эта задача тривиальна, если в сообществе имеется только одно жилище. Однако в некоторых сообществах было несколько жилищ на выбор, а в некоторых — ни одного. Кроме того, после постобработки, описанной в пункте 18, были обнаружены сообщества, в которых несколько домохозяйств конкурировали за имеющееся(щиеся) жилище(а). Как правило, каждое сообщество включало $m \geq 0$ жилища и $n \geq 1$ домохозяйство.

a) В каждом сообществе сила связи между домохозяйствами и жилищами определялась на основе весов «человек-жилище». Мы отдавали предпочтение комбинациям «домохозяйство-жилище», которые были наиболее сильно связаны между собой. В случае связей мы отдавали предпочтение жилищам с более высоким потреблением электроэнергии и более просторным жилищам. На этом этапе 96 % домохозяйств были присвоены жилища:

i) для остальных домохозяйств был выбран наиболее вероятный муниципалитет. Кроме того, мы вычислили точку привязки для каждого домохозяйства в их выбранном муниципалитете, основываясь на географических координатах мест, с которыми члены домохозяйства имели связи;

ii) для этого этапа отбирались жилища, оставшиеся незанятыми. Мы рассматривали места, с которыми члены домохозяйства связаны напрямую или через других людей. Примером потенциального жилища может быть квартира матери члена домохозяйства. Для каждого домохозяйства мы рассматривали только те жилища, которые находились в их отобранном муниципалитете;

iii) между домохозяйствами и жилищами существует связь «многие-ко-многим». Для некоторых домохозяйствах было несколько приемлемых кандидатов на жилище. С другой стороны, некоторые жилища были кандидатами для разных домохозяйств. В этом случае существует множество способов сопряжения домохозяйств и жилищ. Мы рассматривали домохозяйства и жилища как двудольный граф и выбирали устойчивое соответствие. При стабильном сопряжении нет ни одной комбинации «домохозяйство-жилище», которые бы предпочли бы друг друга своим сопряженным компаньонам (Gale & Shapley, 1962);

iv) необходимым условием для вычисления стабильного сопряжения является своего рода ранжирование: какие домохозяйства предпочитают те или иные жилища и наоборот. Мы заявили, что домохозяйства предпочитают жилища, которые: 1) находятся близко к точке привязки и 2) больше по площади. Жилища предпочитают: 1) более крупные домохозяйства, 2) домохозяйства с более тесными связями;

После этого этапа 99,4 % домохозяйств имели жилища;

b) другим домохозяйствам было предоставлено случайно отобранное незанятое жилище, расположенное недалеко от их точки привязки.

20. Отдельно рассматривались лица, находящиеся в институциональных домохозяйствах, и бездомные. Списки бездомных были предоставлены муниципалитетами; члены институциональных домохозяйств были известны из регистров.

III. Результаты

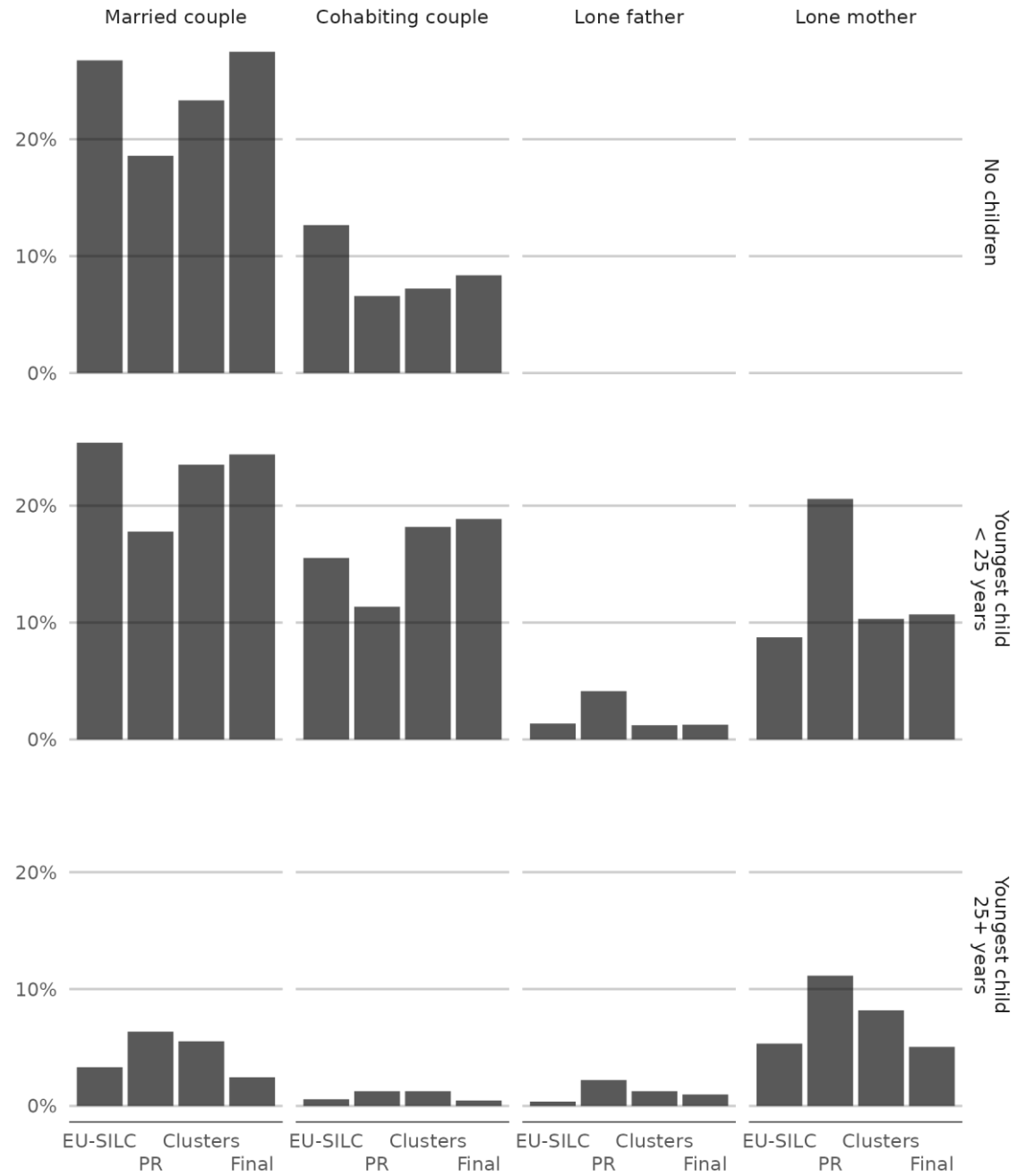
21. Статистика семей и домохозяйств в рамках переписи населения рассчитывалась по новой методике. Она также заменила РН в качестве основы для географической разбивки ежегодной статистики населения, начиная с 1 января 2022 года.

22. Несмотря на использование целого спектра источников информации о жилищах, 3 лица из 4 сохранили свое место жительства в том виде, в котором оно было зарегистрировано в РН. Поскольку некоторым лицам были присвоены другие жилища, численность населения муниципалитетов изменилась по сравнению с РН. Из 79 муниципалитетов 35 остались примерно такими же по численности населения, как и в РН (±2 %), 23 уменьшились не менее чем на 2 %, а 21 увеличили численность населения более чем на 2%. Наибольшие потери понесли малые острова (Рухну –29 %, Вормси –23 %, Кихну –19 % и т. д.) и другие популярные места для летних домиков (Алутагузе –8 %, Нарва-Йыэсуу –6 %). В лидерах оказались русскоязычные города северной Эстонии (Локса +7 %, Маарду +6 %, Кохтла-Ярве +4 %).

23. Мотивацией для применения графового подхода послужило улучшение статистики домохозяйств и семей. Рис. 2 иллюстрирует распределение типа семейной ячейки по различным источникам. В качестве эталона мы используем статистику Европейского союза по доходам и условиям жизни за 2022 год (ОДУЖ-ЕС 2022), собранную через 1–5 месяцев после критического момента переписи населения. Эти данные служат базой для ежегодной статистики домохозяйств и семей. В данных РН указываются семьи, определенные по зарегистрированному месту жительства. Как и в случае пробной переписи, в РН занижена доля семей партнеров и завышена доля семей одиноких родителей. Если применить метод выявления сообществ, то число пар каждого типа возрастет, хотя в случае сожительствующих пар без детей рост скромнее. Наконец, после постобработки мы получаем распределение, которое хорошо согласуется с данными ОДУЖ-ЕС. Мы по-прежнему наблюдаем несоответствие среди сожительствующих пар: подход, основанный на графах, по-видимому, занижает долю пар без детей (12,7 % ОДУЖ-ЕС против 8,4 % с графиками) и завышает долю семей с детьми младшего возраста (15,5 % против 18,9 %). Тем не менее мы считаем результаты, полученные на основе графов, значительным улучшением по сравнению с исходными семьями РН.

24. Начиная с пандемии Covid-19 способ опроса в ОДУЖ-ЕС и обследовании рабочей силы (ОРС) изменился с личного опроса на телефонный и через Интернет. Это может снизить надежность данных о месте жительства в обследованиях. Еще одним недостатком использования данных опроса является то, что мы наблюдаем домохозяйства как экономические единицы, а не домохозяйства с привязкой к адресам. Хотя обычно они совпадают, следует признать, что данные обследований не идеально воспроизводят домохозяйства с привязкой к адресам.

Рис. 2
Распределение типов семейных ячеек по различным источникам



Источники: ОДУЖ-ЕС 2022, РН — регистр населения, Кластеры — графовый подход, после выделения сообществ, Конечные результаты графовый подход, после постобработки.

IV. Заключение

25. Неточность указания места жительства в эстонском РН оказывает влияние на состав домохозяйств. Семьи, построенные на основе данных РН, страдают смещением в сторону семей с одним родителем.

26. Смещение статистики по домохозяйствам и семьям может быть снижено за счет использования других источников административных данных. Мы рассматриваем людей и жилища как узлы графа. ребра — это связи, найденные в регистрах (например, брак соединяет супругов, собственность соединяет квартиру с ее владельцем). Тогда домохозяйство и его жилище можно рассматривать как подграф сильной связности,

или, другими словами, сообщество. Для их поиска мы применили метод обнаружения сообществ.

27. Результирующая статистика хорошо согласуется с оценками семей по данным ОДУЖ-ЕС и демонстрирует заметный прогресс по сравнению со статистикой на основе РН.

Источники

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Gale, D., & Shapley, L. S. (1962). College Admissions and the Stability of Marriage. The American Mathematical Monthly, 69(1), 9. https://doi.org/10.2307/2312726.

Gortfelder, M., & Puur, A. (2021). Tegeliku ja registripõhise elukoha lahknevus ning selle põhjused: 2020. Aasta Eesti tööjõu-uuringu analüüs (lk 31). https://sisu.ut.ee/sites/default/files/mobiilneelu/files/tp1_tlu_tegeliku_ja_registripohise_elukoha_kattuvus_ning_selle_pohjused_etu2020_analuus_gortfelderpuut2021_0.pdf.

Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communitites in Networks. Notices of the American Mathematical Society, 56(9), 1082–1097.

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105.

Tiit, E.-M., Visk, H., Maasing, E., Levenko, V., & Lehto, K. (2021). Järjekordne rahva ja eluruumide loendus: Milleks ja kuidas? Akadeemia, 2021(11), 2009–2064.

Äär, H. (2017). Coincidence of actual place of residence with Population Register records. Quarterly Bulletin of Statistics Estonia, 1, 80–83.