

DOI: 10.51790/2712-9942-2023-4-1-08

О НЕЯВНЫХ СООБЩЕСТВАХ НА ГРАФЕ ВЗАИМОДЕЙСТВУЮЩИХ ОБЪЕКТОВ**А. А. Чеповский***Национальный исследовательский университет «Высшая школа экономики», г. Москва,
Российская Федерация**ORCID: <https://orcid.org/0000-0002-0938-3517>,  aachepovsky@hse.ru*

Аннотация: в работе рассматриваются текущие и перспективные задачи анализа графов, представляющих сложные сети взаимодействующих объектов. Описаны предметные области, для которых актуальны данные модели, рассматриваются практика, проблемы анализа графов социальных сетей и выделения на них неявных сообществ. Обсуждаются основные алгоритмы выделения сообществ на графах и проблема тестирования качества получаемых разбиений. Автором предложены перспективные, на его взгляд, направления развития данной тематики для графов, полученных при импорте данных из реальных социальных сетей. Показана актуальность подходов, предложенных в предыдущих работах автора.

Ключевые слова: сложные сети, граф взаимодействующих объектов, выделение неявных сообществ, анализ социальных сетей.

Для цитирования: Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов. *Успехи кибернетики*. 2023;4(1):56–64. DOI: 10.51790/2712-9942-2023-4-1-08.

*Поступила в редакцию: 05.12.2022.**В окончательном варианте: 17.02.2023.***IMPLICIT COMMUNITIES DEFINED ON THE GRAPH FOR INTERACTING OBJECTS****A. A. Chepovskiy***National Research University Higher School of Economics, Moscow, Russian Federation**ORCID: <https://orcid.org/0000-0002-0938-3517>,  aachepovsky@hse.ru*

Abstract: the paper considers the problems of analyzing graphs that represent complex networks of interacting objects. The relevant domains are described, social network graph analysis and implicit community detection are considered. The key graph-based community detection algorithms and the quality assessment of their results are discussed. The author proposed promising development trends using the graphs representing actual social networks. The relevance of the approaches proposed in the author's previous works is also highlighted.

Keywords: complex networks, graph for interacting objects, implicit community detection, social networks analysis.

Cite this article: Chepovskiy A. A. Implicit Communities Defined on the Graph for Interacting Objects. *Russian Journal of Cybernetics*. 2023;4(1):56–64. DOI: 10.51790/2712-9942-2023-4-1-08.

*Original article submitted: 05.12.2022.**Revision submitted: 17.02.2023.***Введение**

Многие системы, состоящие из большого числа объектов, можно изучать, представляя их как сетевые структуры, образованные взаимодействием их элементов между собой. Речь идет, например, о биологических, технологических, телекоммуникационных, социальных сетях, которые принято объединять под термином «сложные сети» («complex networks»). Анализ построенных на их основе графов взаимодействующих объектов является актуальной на сегодняшний день проблемой, в ходе исследования которой возникают задачи построения алгоритмов и методов выделения ключевой информации, разработки прикладного программного обеспечения для обработки данных.

Графы взаимодействующих объектов в разных предметных областях

Существенный интерес к методам анализа графов взаимодействующих объектов наблюдается в предметных областях, связанных с вопросами информационной безопасности, криминальных исследований, борьбы с экстремизмом и терроризмом. Также интерес к таким методам имеется и для задач, возникающих в биологии, экономике, социологии, маркетинге и многих других дисциплинах.

Проблемы из таких разнообразных предметных областей объединяет то, что в реальных задачах исследуемые модели сводятся к описанию сетевого взаимодействия, требующего анализа устройства достаточно больших по размеру и сложных по структуре графов [1]. Для графов, представляющих реальные сети, часто можно выделить подграфы с высокой плотностью ребер внутри них и сравнительно низкой плотностью ребер между такими подграфами. Разбиение графа на такие подграфы и называется выделением неявных сообществ [1–4].

Телекоммуникационные сети могут быть представлены как граф взаимодействующих объектов. Сеть контактов абонентов мобильной связи, которая обычно обозначается термином «биллинг», может представлять из себя как «классический» биллинг, включающий звонки и SMS-сообщения абонентов, так и соединения в рамках мобильного Интернета. Анализ биллинга является актуальной задачей информационных технологий и задач информационной безопасности [5–7]. Задачи обеспечения информационной безопасности информационно-коммуникационных сетей рассматриваются в [8, 9].

Одной из задач анализа биллинга является использование его результатов для раскрытия преступлений [6]. Целью такого анализа соединений между абонентами в следственной ситуации, когда фигуранты дела неизвестны, является как геопозиционирование абонентов, так и очерчивание круга абонентов, которые потенциально могут иметь отношение к совершенному преступлению. Для решения таких задач полезно выделять неявные сообщества в графе телекоммуникационного взаимодействия.

Биологические сетевые структуры могут представлять собой, например, пищевые сети, метаболические сети, белок-белковые сети взаимодействия [10]. Анализ биомолекулярных сетей актуален для понимания молекулярного механизма биологических систем, в диагностике, лечении и разработке лекарств для сложных заболеваний или расстройств [11].

Важно также отметить применение сетевого анализа как набора инструментов для решения задач реинжиниринга программного обеспечения при проектировании информационных систем [12, 13].

Встречаются и иные задачи, для которых применяется анализ структуры графа взаимодействующих объектов, в том числе включающий в себя выделение неявных сообществ. Например, для оценки социально-экономических моделей городского хозяйства с целью инфраструктурного и транспортного планирования, разработки политики в области недвижимости и социально-экономического развития [14].

В работе [15] сетевые методы анализа применяются для описания геофизических транспортных процессов (например, океанических или атмосферных циркуляций), продемонстрирована их эффективность в оценке транспортировки и смешивания течений в геофизических расчетах.

В качестве примера использования сетей для маркетинговых целей можно привести работу [16], в которой рассматриваются подходы по проектированию сети клиентов на основе рейтинговых данных участников электронной коммерции.

Сетевые структуры могут быть построены на основе данных с сайтов, позволяющих своим пользователям взаимодействовать друг с другом, например, по профессиональной направленности. В таких случаях выделение неявных сообществ на основе взаимодействия пользователей актуально для достижения разных задач. Так, в [17] приведен пример анализа части сети LinkedIn и выявления сообществ среди практиков в сфере социальной работы. Часто сетевые структуры с последующим выделением в них сообществ строятся на основании данных с сайтов размещения вакансий и поиска работников [18].

Всемирную информационно-телекоммуникационную сеть Интернет можно интерпретировать как граф, вершинами которого являются сайты, а ребрами — гиперссылки между ними. Моделирование сети Интернет в общем случае связано с описанием взаимодействия с помощью разреженных графов, что позволяет разным авторам рассматривать, в первую очередь, модели случайных графов для описания такого взаимодействия [1, 19]. Возможно проводить исследование графов, построенных на основе отдельных тематических фрагментов сети Интернет. Например, в работах [20, 21] рассматриваются сайты академических учреждений РАН и их разбиение на сообщества.

Еще одной задачей, возникающей при анализе сетевых структур, является поиск часто встречающихся подграфов в заданном графе, что представляет теоретический интерес для задач обнаружения групп атак в социальных сетях [22, 23]. Также актуальна и постановка схожей задачи поиска часто встречающихся вхождений шаблонов в базах данных, содержащих много небольших графов.

Задачи анализа графов социальных сетей

Одним из самых актуальных направлений исследований сетевых систем является анализ социальных сетей, в том числе социальных онлайн-сетей [3, 4, 24–33]. При построении графов тут в качестве вершин обычно рассматриваются аккаунты пользователей. Ребра и вес на них (при необходимости) определяются, как правило, действиями пользователей каждой конкретной социальной сети: отношениями «дружбы», «подписки», простановкой «лайков» или иными подобными средствами коммуникации. В результате получается граф взаимодействующих объектов. Структура построенных на основе реальных данных графов сложных сетей, к которым относятся и графы социальных сетей, имеет свои особенности. Для таких графов характерны в том числе следующие свойства: маленький диаметр графа (эффект «малого мира»), высокие значения кластерного коэффициента (эффект «транзитивности»), медленно спадающее распределение вершин по их степеням («большой хвост распределения»), структура неявных сообществ [1–3, 33].

Замечено, что асимптотическая зависимость распределения вершин для больших степеней («в хвосте») близка к степенному закону. Т.е. у небольшого числа вершин в таких сетях высокие показатели их степени. Поэтому иногда в литературе такие сети называются безмасштабными (scale-free networks), так как распределение сохраняется и в локальных фрагментах — при «изменении масштаба». Это, в свою очередь, связано с тем, что степенная функция является единственным решением соответствующего функционального уравнения [33, 34].

Среди сложных сетей, в свою очередь, обычно выделяют социальные, биологические, технологические [35]. Графы социальных сетей обладают содержательной структурой неявных сообществ, представляющих группы пользователей, объединенных либо активной коммуникацией друг с другом, либо по набору дополнительных признаков (например, это могут быть однокурсники, подписчики определенных тематических сайтов). Тут возникают задачи выделения неявных сообществ, выявления лидеров мнений, исследований в части распространения информации по сети.

Построение моделей распространения информации и информационного влияния в социальных сетях рассматривалось в [26, 27]. В этих работах предложены методы описания информационных потоков и психологических операций для задач информационного противоборства. Обзор инструментария для анализа распространения информации и информационного влияния в социальных сетях, включая выделение неявных сообществ, можно найти в работах [9, 26, 27].

Задачи выявления путей распространения информации, повышения уровня защищенности от информационно-психологического воздействия в социальных сетях, в том числе выявления источников и ретрансляторов информации, освещены в [36]. В данной работе авторы приводят методику построения графа распространения информации, но далее построенный так граф предлагается рассматривать методами визуального анализа, что накладывает определенные ограничения на размер обрабатываемых так графов.

Для социологов важной задачей является мониторинг общественных изменений посредством анализа структуры явных и неявных групп пользователей в социальных сетях. Как пример решения таких задач в [37] рассматривается метод, названный его авторами «алгоритмом зерновой кластеризации». Данный метод соединяет в себе этапы экспертного определения целевых групп и их дополнения на основе данных из социальной сети. Выделение сообществ в графах взаимодействующих объектов онлайн социальных сетей дает возможность обнаруживать лидеров мнений и экспертов, осуществлять управление и контроль деятельности групп, представляющих угрозу для национальной безопасности [9].

В научной литературе рассматривается множество классических мер центральности вершины графа [2, 3, 38, 39], характеризующих степень ее влияния на другие вершины (центральность по посредничеству, центральность по Кацу, Pagerank и т.д.). Возможности применения таких характеристик для анализа графов социальных сетей представлены в [40] на примере алгоритмов ранжирования вершин. При этом классические меры центральности учитывают различные локальные или глобальные свойства вершины, но не учитывают структуру сообществ графа. Между тем, для целей сетевого анализа часто важно учитывать именно структуру сообществ, поэтому в последнее время внимание исследователей обращено на основанные на сообществах меры центральности (community-aware centrality measures) [41]. Это означает, что центральности вершин считаются после выделения сообществ и учитывают отдельно инцидентные этой вершине ребра внутри сообщества и между сообществами.

В обзоре [41] приведены исследования 7 таких мер центральности, правда, стоит отметить, что все рассмотренные меры центральности и эксперименты приведены для невзвешенных графов.

Методы выделения сообществ на графах

Как уже было сказано ранее, под выделением неявных сообществ на графе подразумевается разбиение графа на подграфы, такое что плотность связей внутри этих подграфов сильно выше плотности связей между ними. Понятие сообщества на графах встречается во множестве работ. При этом можно рассмотреть и задачу выделения на графе пересекающихся сообществ. В этом случае подразумевается наличие общих вершин, принадлежащих сразу двум или более сообществам. Алгоритмы поиска неявных сообществ в указанном смысле являются актуальной темой публикаций последних 20 лет [42–45].

Один из иных интуитивно понятных способов выделения сообществ на графе — поиск такого его разбиения, при котором максимизируется/минимизируется некоторый функционал, характеризующий качество этого разбиения. Такой функционал, называемый модулярностью, может сравнивать разбиение графа с некоторой «нулевой гипотезой», заключающейся в том, что ребра распределены случайно, но сохранены некоторые свойства исходного графа. Значение этого функционала будет зависеть от выбранной «нулевой гипотезы», такая была предложена в работе [46–48], а модулярность, рассматриваемая в ней, носит название авторов (Ньюмана–Гирвана). Данная модель сохраняет исходные степени вершин графа, и при этом предполагает случайное распределение ребер между ними, т.е. нет закономерностей в распределении плотности ребер внутри и между выделенными сообществами. Такая модулярность может быть использована для взвешенных графов и является одной из наиболее популярных характеристик качества разбиения графа на сообщества, за последние 20 лет ей было посвящено множество работ [49].

Подходы по оптимизации данного функционала качества за эти годы были предложены разные, это и жадный алгоритм, и алгоритм симуляция отжига. Одним из быстрых алгоритмов, основанных на активном использовании модулярности Ньюмана–Гирвана, является агрегативный иерархический алгоритм Louvain быстрой оптимизации модулярности [50]. Данный алгоритм находит разбиения больших графов с высокой степенью модулярности за короткое время, кроме того, предоставляет информацию о полной иерархической структуре сообществ, тем самым давая доступ к различным расширениям выделенных сообществ.

Другой популярный подход к выделению сообществ заключается в имитации условного динамического процесса на графе. С точки зрения структуры сообществ малая плотность связей между ними влечет за собой с большей вероятностью протекание процесса внутри сообществ, чем переход между ними. Самый популярный на сегодня алгоритм, реализующий данный подход [51–53], предусматривает сжатие информации о динамическом процессе, проходящем в графе, а именно о случайном блуждании. Этот динамический алгоритм сводит задачу нахождения наилучшего структурного разбиения графа на сообщества к задаче оптимального сжатия информации о структуре графа. Для вычисления показателя качества заданного разбиения используется энтропия, описывающая среднюю длину кодового слова, взятого для кодирования вершины. Показатель качества полученного разбиения, выраженный через энтропию, может быть легко подсчитан для любого разбиения. Обновление и пересчет этого показателя являются быстрой операцией. В дальнейшем были предложены пути развития этого алгоритма, в том числе и для пересекающихся сообществ [54–56].

Встречаются и иные подходы к выделению сообществ. Например, теоретико-игровой метод в работе [57], где рассматривается выявление академических сообществ российских ученых в сети Интернет. А работа [45] посвящена алгоритму выделения пересекающихся сообществ с помощью рассмотрения полных подграфов заданного размера и его применению в биологических сетях.

В недавней статье [49] исследователей, стоящих у истоков текущей теории сложных сетей, в том числе в вопросах выделения сообществ, были указаны некоторые актуальные, по мнению авторов, аспекты. Как одна из задач указана проблема предела разрешения («resolution limit») [58] — ограничения снизу по размеру выделяемого в большой сети сообщества, что мешает находить маленькие сообщества на крупных графах.

Решение такой проблемы видится в реализации корректного итерационного перехода от анализа исходного графа к отдельным его подграфам.

О тестировании алгоритмов выделения сообществ

Отсутствие в общем случае строгого определения для сообщества и однозначного решения задачи выделения как пересекающихся, так и непересекающихся сообществ ставит проблему оценки корректности и эффективности работы алгоритмов и методов выделения сообществ на графах взаимодействующих объектов.

Стандартным тут может быть подход по генерации случайных графов для проведения на них тестирования алгоритмов и оценки полученного разбиения на сообщества. Первая предложенная в свое время для анализа сетей модель — модель Эрдеша–Реньи, плохо соответствует реальным сложным сетям, особенно социальным (онлайн) сетям [1–3, 35]. Модель Эрдеша–Реньи симулирует эффект «малого мира», но остальные свойства реальных сложных сетей в этой модели не реализуются, например, «эффект транзитивности», а биномиальное (пуассоновское для больших графов) распределение степеней вершин нехарактерно для графов реальных сетей.

Для описания растущих сетей, которые развиваются со временем и которые нельзя описать моделью Эрдеша–Реньи, рассматриваются различные варианты модели Барабаши–Альберт. Модель Барабаши–Альберт при добавлении каждой новой вершины реализует метод предпочтительного присоединения между новой вершиной и старыми. Он заключается в построении ребер между новой вершиной и старыми с вероятностями, пропорциональными степеням старых вершин. В графах, сгенерированных по модели Барабаши–Альберт, больше вершин с малой степенью, чем в случайных графах; более вероятны вершины с высокой степенью. Т.е. распределение степеней вершин ближе к степенному закону. При этом модель симулирует и эффект «малого мира». Однако данная модель не реализует «эффект транзитивности»: при увеличении размера генерируемого графа его кластерный коэффициент существенно падает [33]. Это противоречит свойству реальных сложных сетей.

Возникает запрос на генерацию графов, подходящих под условия сложных сетей. На первых этапах работы над алгоритмами нахождения сообществ на графах небольших размеров использовались составленные простые тесты. Наиболее известный тест был предложен Гирваном и Ньюманом (тест GN) [42] и содержал графы размером 128 вершин, которые разбиваются на сообщества одинаковых размеров. Но вершины у таких сгенерированных графов имели примерно одинаковые степени, а сообщества состояли из одинакового числа вершин. Таким образом, данная модель также была далека от реальных сложных сетей.

Многие разработчики применяют LFR-модели [59] генерации случайных графов, обладающих структурой сообществ. Получаемые в рамках этих моделей графы обладают распределением степеней вершин по степенному закону. При этом размеры встроенных при построении сообществ также реализуются моделью по степенному закону, что нельзя однозначно отнести к положительным или отрицательным свойствам данных моделей.

Логичным развитием разработки моделей генерации графов для тестирования алгоритмов является выбор меры качества для сравнения получаемых алгоритмами разбиений сгенерированного графа на сообщества и исходно встроенного в него разбиения на сообщества. Можно выделить три наиболее распространенных подхода для решения этой задачи [3]. Первый из них основан на сравнении полученных сообществ для пар вершин, т.е. подсчете числа таких пар вершин, попавших в единое сообщество при обоих разбиениях, и попавших в разные сообщества в обоих разбиениях, и попавших в разные сообщества только при одном из разбиений. Наиболее известным из первой группы является индекс Жаккара. Для такого сравнения характерен недостаток, который заключается в том, что расхождения на локальном фрагменте графа могут дать существенное снижение общего результата для всего графа.

Второй возможный подход для сравнения двух разбиений состоит в поиске наибольших пересечений пар сообществ из разных разбиений. Но у этого подхода имеются свои недостатки, связанные с небольшими пересечениями сообществ из разных разбиений.

Третий вариант решения данной задачи основан на методах из теории информации и заключается в предположении о том, что для близких разбиений достаточно сообщить небольшое количество информации, чтобы получить из одного другое. Одним из наиболее популярных тут является подсчет NMI — нормированной взаимной информации (Normalized Mutual Information) [60]. Во многих работах последних лет сравнение результатов алгоритмов произведено с помощью NMI, между тем выявлены и недостатки у этой меры качества, связанные со случайными разбиениями на большое число маленьких сообществ. Поэтому ряд авторов предлагают свои корректировки для NMI [61].

Как было ранее отмечено, реальные сложные сети обладают рядом свойств, которые не во всех моделях генерируемых графов реализуются. Более того, как показывают разные исследования, сложные сети тоже нельзя считать однородными, ибо имеются характерные для того или иного типа сетей (социальные, биологические, технологические) топологические свойства [1, 2, 35]. Кроме этого, стоит отметить, что во многих работах последних 20 лет детально рассматриваются случаи невзвешенных графов, тогда как для социальных сетей важную роль играют атрибуты ребер и вершин [62]. Например, интенсивность взаимодействия между пользователями удобно описывается взвешенными графами. Поэтому важным видится исследование моделей импорта данных из реальных сетей и построение на их основе взвешенных графов взаимодействующих объектов [63-66]. Распределение степеней вершин в таких графах слабо соответствует существующим на текущий момент искусственным моделям. В работе [49] также поднят как актуальный вопрос об оценке качества выделения сообществ на сгенерированных графах. Авторами отмечено, что сравнение с заранее заданным разбиением на искусственных сетях дает сбой для разреженных графов, что как раз характерно и для социальных сетей.

Для оценки качества выделения сообществ на графе взаимодействующих объектов для социальных онлайн-сетей целесообразным видится анализ получаемых сообществ на основе тематических, социально-политических факторов, обработки текстов и иных характеристик и атрибутов пользователей. Это поможет решать проблему оценки качества разбиения при импорте данных из телекоммуникационных сетей с помощью исследования атрибутивных данных вершин и ребер выделенных сообществ. Пример такого использования дан в работе [67], где проводится психолингвистический анализ текстов и их характеристик у получившихся сообществ для предлагаемого в статье метода их выделения.

Заключение

За прошедшие годы текущего века многие ученые по всему миру развили теорию сложных сетей: были построены различные их модели, исследованы свойства таких сетей, разработаны алгоритмы выделения неявных сообществ. Между тем стоит отметить, что для реальных сетей во многих случаях ключевыми, оказывающими влияние на их топологию, являются свойства объектов и связей между ними. Поэтому одним из актуальных направлений дальнейших исследований видится эффективный импорт данных из реальных сетей с последующим построением взвешенных графов взаимодействующих объектов, выделением на них неявных сообществ, в том числе с использованием механик, основанных на переходе от анализа всего исходного графа к его подграфам. Примеры такого подхода, названные «комбинированный метод», «метод ядра», «метод Галактик», были предложены в работах автора [67-69].

С точки зрения актуальных прикладных задач, в том числе задачи выявления лидеров мнений, информационного воздействия и распространения информации в социальных сетях целесообразно применять сочетание алгоритмического выделения сообществ с психолингвистическим анализом полученных в них текстов. Также перспективным видится поиск основанных на сообществах мер центральности для взвешенных графов.

ЛИТЕРАТУРА

1. Евин И. А. Введение в теорию сложных сетей. *Компьютерные исследования и моделирование*. 2010;2(2):121-141. DOI: 10.20537/2076-7633-2010-2-2-121-141.
2. Newman M. E. J. *Networks: An Introduction*. Oxford University Press; 2010. 784 p.
3. Fortunato S. Community Detection in Graphs. *Physics Reports*. 2010;486(3):75-174.
4. Aggarwal C. *Social Network Data Analytics*. NY:Springer New York; 2011. 502 p. DOI: 10.1007/978-1-4419-8462-3.
5. Рабинович Б. И. Кластерный анализ детализаций телефонных переговоров. *Системы и средства информатики*. 2007;17:52-78.
6. Себякин А. Г. Анализ информации о соединениях между абонентами, использование его результатов в раскрытии и расследовании преступлений. *Полицейская и следственная деятельность*. 2018;4:29-38. DOI: 10.25136/2409-7810.2018.4.27992.
7. Семенищев И. А., Синадский А. Н., Синадский Н. И., Сушков П. В. Синтез массивов биллинговой информации на основе статистико-событийной модели взаимодействия абонентов сетей сотовой связи. *Вестник УРФО. Безопасность в информационной сфере*. 2018;1(27):47-56.

8. Еремеев И. Ю., Татарка М. В., Шуваев Ф. Л., Цыганов А. С. Анализ мер центральности вершин сетей на основе метода главных компонент. *Труды СПИИРАН*. 2020;19(6):1307-1331. DOI: 10.15622/ia.2020.19.6.7.
9. Кириченко Л., Радивилова Т., Барановский А. Обнаружение киберугроз с помощью анализа социальных сетей. *International Journal "Information Technologies & Knowledge"*. 2017;11(1):23-48.
10. Rahiminejad S., Maurya M. R., Subramaniam S. Topological and Functional Comparison of Community Detection Algorithms in Biological Networks. *BMC Bioinformatics*. 2019;20:212.
11. Wu F., Chen L., Wang J., Li M., Wang H. Biomolecular Networks for Complex Diseases. *Complexity*. 2018;4210160.
12. Šubelj L., Bajec M. Ubiquitousness of Link-Density and Link-Pattern Communities in Real-World Networks. *The European Physical Journal B*. 2012;85(1):32.
13. Šubelj L., Bajec M. Group Detection in Complex Networks: an Algorithm and Comparison of the State of the Art. *Physica A: Statistical Mechanics and Its Applications*. 2014;397:144-156.
14. Landsman D., Kats P., Nenko A., Sobolevsky S. Zoning of St. Petersburg Through the Prism of Social Activity Networks. *Procedia Computer Science*. 2020;178:125-133.
15. Ser-Giacomi E., Legrand T., Hernández-Carrasco I., Rossi V. Explicit and Implicit Network Connectivity: Analytical Formulation and Application to Transport Processes. *Physical Review E*. 2021;103(4).
16. Banerjee S. Designing and Connectivity Checking of Implicit Social Networks from the User-Item Rating Data. *Multimedia Tools and Applications*. 2021;80(17):26615-26635.
17. Castillo-de Mesa J., Gómez-Jacinto L. Connectedness, Engagement, and Learning through Social Work Communities on LinkedIn. *Psychosocial Intervention*. 2020;29(2):103-112.
18. Skobtsov Y. A., Obolensky D. M., Shevchenko V. I., Chengar O. V. Building And Analysing A Skills Graph Using Data From Job Portals. *Economic and Social Trends for Sustainability of Modern Society (ICEST-III 2022). European Proceedings of Social and Behavioural Sciences*. European Publisher. 2022;127:147-162. DOI: 10.15405/epsbs.2022.08.17.
19. Райгородский А. М. *Модели Интернета: Учебное пособие*. Долгопрудный: Интеллект; 2019. 64 с.
20. Ермолин Н. А., Мазалов В. В., Печников А. А. Теоретико-игровые методы нахождения сообществ в академическом Вебе. *Труды СПИИРАН*. 2017;6(55):237-254. DOI 10.15622/sp.55.10.
21. Мазалов В. В., Никитина Н. Н., Печников А. А. О сообществах в коммуникационных графах. *Вероятностные методы в дискретной математике: расширенные тезисы докладов X Международной Петрозаводской конференции*. 2019. Петрозаводск: Федеральное государственное бюджетное учреждение науки Федеральный исследовательский центр «Карельский научный центр Российской академии наук»; 2019:99-100.
22. Черемисинов Д. И., Черемисинова Л. Д. Поиск часто встречающихся подграфов. *BIG DATA Advanced Analytics: Collection of Materials of the Fourth International Scientific and Practical Conference*, Minsk, Belarus, May 3-4, 2018. Minsk: BSUIR. 2018:171-176.
23. Rehman S. U., Kexing Liu K., Tariq Ali T., Nawaz A., Fong S. J. A Graph Mining Approach for Ranking and Discovering the Interesting Frequent Subgraph Patterns. *International Journal of Computational Intelligence Systems*. 2021;14(152). DOI: 10.1007/s44196-021-00001-4.
24. Базенков Н. И., Губанов Д. А. Обзор информационных систем анализа социальных сетей. *Управление большими системами: сб. трудов*. 2013:357-394.
25. Батура Т. В. Методы анализа компьютерных социальных сетей. *Вестник НГУ. Серия: Информационные технологии*. 2012;10(4):13-28.
26. Батура Т. В., Копылова Н. С., Мурзин Ф. А., Проскуряков А. В. Методы анализа данных из социальных сетей. *Вестник НГУ. Серия: Информационные технологии*. 2013;11(3):5-21.
27. Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. *Социальные сети: модели информационного влияния, управления и противоборства*. М.: Физматлит; 2010. 228 с.
28. Borgatti S. P., Everett M. G., Johnson J. C. *Analyzing Social Networks*. SAGE Publications Limited; 2013. 304 p.
29. Coscia M., Rossetti G., Giannotti F., Pedreschi D. Demon: a Local-First Discovery Method for Overlapping Communities. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2012:615-623.
30. Gaisbauer F., Pournaki A., Banisch S., Olbrich E. Ideological Differences in Engagement in Public Debate

- on Twitter. *Plos One*. 2021;16(3):e0249241. DOI: <https://doi.org/10.1371/journal.pone.0249241>.
31. Kanavos A., Voutos Y., Grivokostopoulou F., Mylonas P. Evaluating Methods for Efficient Community Detection in Social Networks. *Information*. 2022;13(5), 209.
 32. Yang J., Leskovec J. Defining and Evaluating Network Communities Based on Ground-Truth. *Knowledge and Information Systems*. 2015;42(1):181–213.
 33. Гусарова Н. Ф. *Анализ социальных сетей. Основные понятия и метрики*. СПб.: Университет ИТМО; 2016. 67 с.
 34. Newman M. E. J. The Structure and Function of Complex Networks. *SIAM Review*. 2003;45(10):167–256.
 35. Евин И. А., Хабибуллин Т. Ф. Социальные сети. *Компьютерные исследования и моделирование*. 2012;4(2):423–430. DOI: 10.20537/2076-7633-2012-4-2-423-430.
 36. Проноза А. А., Виткова Л. А., Чечулин А. А., Котенко И. В., Сахаров Д. В. Методика выявления каналов распространения информации в социальных сетях. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*. 2018;14(4):362–377. DOI: 10.21638/11702/spbu10.2018.409.
 37. Градосельская Г. В., Щеглова Т. Е., Карпов И. А. Картирование политически активных групп в Фейсбуке¹: динамика 2013–2018 гг. *Вопросы кибербезопасности*. 2019;32(4):94–104. DOI: 10.21681/2311-3456-2019-4-94-104.
 38. Borgatti S. P. Centrality and Network Flow. *Social Networks*. 2005;27(1):55–71. DOI: 10.1016/j.socnet.2004.11.008.
 39. Щербакова Н. Г. Меры центральности в сетях. *Проблемы информатики*. 2015;2(27):18–30.
 40. Печенкин В. В., Королев М. С., Димитров Л. В. Прикладные аспекты использования алгоритмов ранжирования для ориентированных взвешенных графов (на примере графов социальных сетей). *Труды СПИИРАН*. 2018;6(61):94–118. DOI: 10.15622/sp.61.4.
 41. Rajeh S., Savonnet M., Leclercq E. et al. Comparative Evaluation of Community-Aware Centrality Measures. *Qual Quant*. 2022. DOI: 10.1007/s11135-022-01416-7.
 42. Girvan M., Newman M. E. J. Community Structure in Social and Biological Networks. *Proc. Natl. Acad. Sci. USA*. 2002;99(12):7821–7827.
 43. Clauset A., Newman M. E. J., Moore C. Finding Community Structure in Very Large Networks. *Physical Review E*. 2004;70:066111.
 44. Radicchi F., Castellano C., Loreto V., Cecconi F., Parisi D. Defining and Identifying Communities in Networks. *Proc. Natl. Acad. Sci. USA*. 2004;101(9):2658–2663.
 45. Palla G., Derenyi I., Farkas I., Vicsek T. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature*. 2005;435:814–818.
 46. Newman M. E. J., Girvan M. Finding and Evaluating Community Structure in Networks. *Physical Review E*. 2004;69:026113.
 47. Newman M. E. J. Fast Algorithm for Detecting Community Structure in Networks. *Physical Review E*. 2004;69:066133.
 48. Newman M. E. J. Modularity and Community Structure in Networks. *Proc. Natl. Acad. Sci. USA*. 2006;103(23):8577–8582.
 49. Fortunato S., Newman M. E. J. 20 years of Network Community Detection. *Nat. Phys*. 2022;18:848–850.
 50. Blondel V. D., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;10:P10008.
 51. Rosvall M., Bergstrom C. T. An Information-Theoretic Framework for Resolving Community Structure in Complex Networks. *Proc. Natl. Acad. Sci. USA*. 2007;104(18):7327–7331.
 52. Rosvall M., Bergstrom C. T. Maps of Information Flow Reveal Community Structure in Complex Networks. *Proc. Natl. Acad. Sci. USA*. 2008;105(4):1118–1123.
 53. Rosvall M., Bergstrom C. T., Axelsson D. The Map Equation. *The European Physical Journal Special Topics*. 2009;178(1):13–23.
 54. Esquivel A., Rosvall M. Compression of Flow Can Reveal Overlapping Modular Organization in Networks. *Physical Review X*. 2011;1:021025.
 55. Rosvall M., Esquivel A., Lancichinetti A., West J., Lambiotte R. Memory in Network Flows and Its

¹ Meta Platforms Inc. признана экстремистской организацией на территории РФ. — Прим. ред.

- Effects on Spreading Dynamics and Community. *Nature Communications*. 2014;5:4630.
56. Domenico M., Lancichinetti A., Arenas A., Rosvall M. Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems. *Physical Review X*. 2015;5:011027.
 57. Мазалов В. В., Никитина Н. Н. Метод максимального правдоподобия для выделения сообществ в коммуникационных сетях. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*. 2018;14(3):200–214.
 58. Fortunato S., Barthélemy M. Resolution Limit in Community Detection. *Proc. Natl. Acad. Sci. USA*. 2007;104:36–41.
 59. Lancichinetti A., Fortunato S., Radicchi F. Benchmark Graphs for Testing Community Detection Algorithms. *Physical Review E*. 2008;78:046110.
 60. Danon L., Díaz-Guilera A., Duch J., Arenas A. Comparing Community Structure Identification. *J. Stat. Mech.* 2005;P09008.
 61. Amelio A., Pizzuti C. Is Normalized Mutual Information a Fair Measure for Comparing Community Detection Methods? *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. Association for Computing; 2015.
 62. Коломейченко М. И., Поляков И. В., Чеповский А. А., Чеповский А. М. *Методы визуального анализа графов*. М.: Национальный открытый университет «ИНТУИТ»; 2016. 167 с.
 63. Лещев Д. А., Сучков Д. В., Хайкова С. П., Чеповский А. А. Алгоритмы выделения групп общения. *Вопросы кибербезопасности*. 2019;32(4):61–71. DOI: 10.21681/2311-3456-2019-4-61-71.
 64. Попов В. А., Чеповский А. А. Модели импорта данных из Твиттера. *Вестник НГУ. Серия: Информационные технологии*. 2021;19(2):76–91. DOI: 10.25205/1818-7900-2021-19-2-76-91.
 65. Воронин А. Н., Ковалева Ю. В., Чеповский А. А. Взаимосвязь сетевых характеристик и субъектности сетевых сообществ в социальной сети Твиттер. *Вопросы кибербезопасности*. 2020;37(3):40–57. DOI: 10.21681/2311-3456-2020-03-40-57.
 66. Попов В. А., Чеповский А. А. Модели импорта данных из мессенджера Telegram. *Вестник Новосибирского государственного университета. Серия: Информационные технологии*. 2022;20(2):60–71. DOI: 10.25205/1818-7900-2022-20-2-60-71.
 67. Попов В. А., Чеповский А. А. Выделение неявных сообществ на графе взаимодействия Telegram-каналов с помощью «метода Галактик». *Труды ИСА РАН*. 2022;72(4):39-50. DOI: 10.14357/20790279220405.
 68. Чеповский А. А., Лобанова С. Ю. Комбинированный алгоритм выделения сообществ в графах взаимодействующих объектов. *Бизнес-информатика*. 2017;42(4):64–73. DOI: 10.17323/1998-0663.2017.4.64.73.
 69. Chepovskiy A. A., Leshchev D. A., Khaykova S. P. Core Method for Community Detection. *Complex Networks & Their Applications IX. Volume 1: Proceedings of the Ninth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2020*. Springer. 2021:38–50.