

ДМИТРИЙ СЕРГЕЕВИЧ СЕМЕНЬКО

аспирант школы по социологическим наукам,
стажер-исследователь Международной лаборатории
прикладного сетевого анализа Национального
исследовательского университета «Высшая школа экономики»,
Москва, Россия;
e-mail: dssemenko@hse.ru



Обзор исследований в области сетевого анализа сообществ на платформе *GitHub*

УДК: 316.77

DOI: 10.24412/2079-0910-2025-1-174-192

В современном мире важную роль занимает разработка программного обеспечения. В связи с этим исследователи все чаще обращаются к изучению сообществ программистов, а также коммуникации в этих сообществах. Важнейшей частью социального пространства сообщества программистов являются платформы для коллаборации. Это системы, которые предназначены для хостинга и управления проектами. Крупнейшая из таких систем — платформа *GitHub*. На ней более 100 миллионов разработчиков размещают свои проекты, принимают участие в разработке проектов других программистов или просто наблюдают за происходящим в интересующей их области. *GitHub* также активно используется в современном академическом сообществе для преподавания или публикации программного обеспечения, используемого в научных целях. Отсюда актуальность изучения этой платформы для социологии науки и техники. Взаимодействие разработчиков на платформе *GitHub* представляет собой один из наиболее масштабных и сложных видов социального взаимодействия в интернете. Статья посвящена подходам, которые предлагаются в сетевых исследованиях сообществ этой платформы. Эти исследования связаны прежде всего с разработкой открытого (open source) программного обеспечения. В рамках обзора существующих исследований по этой теме рассматриваются виды сетей, которые существуют на платформе *GitHub*, их структурные особенности, а также каналы коммуникации, которые позволяют разработчикам взаимодействовать. Анализируются подходы, связанные с изучением географического и темпорального аспектов сетевого взаимодействия. Рассмотрение существующих в данной области работ позволит, во-первых, подготовить переход от изучения однородных сетей к изучению гетерогенных сетей (то есть таких сетей, которые включают в себя различные сущности и их связи, — пользователей, репозитории, форки репозитория, подписки на пользователей и репозитории), а во-вторых, осмыслить потенциал использования сетевого анализа сообществ *GitHub* в рамках социологии науки и технологии, а также социологии труда.

Ключевые слова: социология программирования, программисты, разработчики, платформы, *GitHub*, сети, сетевой анализ, коммуникация, интернет-сообщества, открытое программное обеспечение.

© Семенов Д.С., 2025

Благодарность

Исследование выполнено в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

Введение

GitHub представляет собой платформу для хостинга и командной разработки IT-проектов. Первые академические исследования сообществ этой платформы начали появляться более 10 лет назад, когда *GitHub* стал получать все более широкую известность в качестве площадки для взаимодействия и сотрудничества программистов, а в дальнейшем также ученых, преподавателей и студентов. Исследования в этой области с самого начала были связаны с использованием социального сетевого анализа (social network analysis). Сетевые исследования *GitHub* рассредоточены по различным направлениям, которые занимаются изучением онлайн-сообществ. Значительное число публикаций посвящено изучению взаимодействия в командах разработчиков открытого программного обеспечения (open source software) и влиянию социальных факторов на разработку ПО. Эти исследования, как правило, осуществляются в области компьютерных наук или менеджмента. Исследователи в области социологии труда относительно недавно начали обращать внимание на *GitHub*. Следует отметить, что социология IT-сообществ только начинает формироваться, поэтому сетевые исследования взаимодействия разработчиков ПО могли бы быть полезны для формирования и развития этого направления. В связи с этим для исследователей в области социологии науки и технологий важно обратиться к наработкам в этой области, сделанным представителями других направлений.

Устройство платформы подразумевает возможность появления сетей разного вида, и для социологического изучения этих сетей необходимо осмыслить уже существующие подходы. Эти подходы могут варьироваться по степени задействования концептуального аппарата сетевого анализа и той роли, которую сетевой анализ играет в исследовании. В некоторых публикациях методы сетевого анализа используются для проверки гипотез, связанных с конкретной областью. В других публикациях предпринимается более творческое использование сетевого анализа, в частности, к понятийному аппарату сетевого анализа обращаются для концептуализации тех социальных явлений, которые обнаруживаются в ходе изучения сообществ *GitHub*. Ряд публикаций посвящен попыткам систематизации того множества различных сетей, которое обнаруживается в рамках взаимодействия на платформе.

Сети взаимодействия на *GitHub*

GitHub отличается большим разнообразием инструментов, предлагаемых для разработчиков. Отсюда разнообразие сетей, которые возникают в ходе взаимодействия пользователей платформы. Одна из ранних попыток систематизации сетей *GitHub* дана в статье, посвященной изучению влияния в сообществах платформы

[Bana, Arora, 2018]. Авторы обнаруживают несколько разновидностей сетей, возникающих на *GitHub*.

Структуру этих сетей можно описать с помощью математического понятия «граф». Граф представляет собой множество узлов (также называемых вершинами); некоторые из них имеют между собой связи. Эти связи могут быть направленными (в одну сторону или в обе), и тогда граф называется направленным. Если связи между узлами графа не имеют направления, то такой граф называется ненаправленным.

1. Сеть «разработчик — разработчик». Эта сеть представляет собой граф, в котором разработчики представлены узлами, а связь между двумя узлами присутствует в том случае, если разработчики участвуют хотя бы в одном общем проекте. Такую сеть можно назвать сетью коллабораций, или сетью сотрудничества.
2. Сеть «репозиторий — репозиторий». Узлами такого графа являются репозитории, а связь между ними присутствует, если у двух репозиторияев есть хотя бы один общий разработчик.
3. Сеть «разработчик — репозиторий». Это сеть участия пользователя в разработке каких-то проектов с помощью пул-реквестов (pull-requests, запросов на внесение изменений в репозиторий).
4. Сеть «разработчик — подписчик». Эта сеть строится на основании подписок на разработчиков. Подписка на *GitHub*-аккаунт разработчика может означать как личное знакомство с человеком, так и интерес к его деятельности.
5. Сеть звезд. Пользователи ставят звезды интересующим их репозиториям. Сеть звезд, как и сеть «разработчик — подписчик», может использоваться для построения графов интересов пользователя.

Стоит упомянуть, что эта систематизация не учитывает сети форков (forks), то есть ответвлений от основных проектов. Пользователи создают форки для того, чтобы развивать проект в том направлении, в котором пожелают. Изучение взаимодействия разработчиков в рамках сетей форков осуществляется уже давно, поскольку этот аспект взаимодействия является одним из ключевых как для платформы *GitHub*, так и вообще для разработки открытого программного обеспечения. Отсюда необходимость для изучения сетей форков в социологическом аспекте, поскольку эти сети позволяют социологам науки и технологий изучать не только сотрудничество в рамках профессиональных онлайн-сообществ, но и инновации, связанные с размежеванием и ответвлением от первоначального проекта. Многие известные проекты (в том числе связанные с наукой) выстраиваются на основе уже существующих.

Развитие программного обеспечения через форки представляет собой его обособление от родительского проекта (основная ветка, mainline) и дальнейшее превращение в самостоятельный проект, от которого, в свою очередь, могут отделяться новые ветви. Все изменения в кодовой базе проектов *GitHub* зафиксированы, и по ним можно последовательно проследить всю хронологию развития проекта. Каждый коммит (фиксация изменений в проекте), сделанный пользователем, хранится в истории репозитория, и это позволяет оценить индивидуальный вклад конкретного разработчика в развитие проекта. Таким образом, благодаря своему устройству *GitHub* представляет собой довольно удобный для социологического изучения объект, который позволяет исследовать деятельность специалистов в области технологий на микроуровне.

Для изучения развития независимых форк-проектов исследователями была введена концепция *iCommits* (interesting commits, «интересные коммиты») [Biazzini, Baudry, 2014]. Это те коммиты, которые характерны именно для рассматриваемого репозитория, но не для его ветки или других форков ветки. *iCommits* — это своего рода инновации, которые производятся в рамках форка. Авторы исследования проводят анализ форков различных репозиторий и на основании этого предлагают концепцию «видообразования» программного обеспечения, когда возникают одна или несколько ветвей, которые используются для создания альтернативных решений, имеющих мало общего с другими форками. В рамках социологии технологий для лучшего понимания осуществления технологических инноваций в разработке ПО имеет смысл обратиться к изучению тех сообществ, которые связаны с таким «видообразованием» на платформе *GitHub*.

Другим важным исследованием сети форков является статья, авторы которой рассматривают более 1,8 млн форков на *GitHub* и в дополнение к этому проводят опрос разработчиков [Jiang et al., 2016]. В исследовании предпринимается попытка выявления тех факторов, которые способствуют разветвлению проектов на *GitHub*. Как выяснили исследователи, одним из ключевых факторов, определяющих потенциал возникновения сети форков вокруг репозитория, является репутация разработчика, которую оценивают по различным показателям (участие в проектах, количество подписчиков, дата регистрации и т. д.). Исследования такого рода закладывают основы теории социального капитала *GitHub*. Развитие этой теории могло бы стать важным шагом в рамках социологии науки и технологий, поскольку помогло бы лучше описывать и анализировать взаимодействие как разработчиков ПО, так и ученых в связанных с разработкой областях. Такая теория должна выстраиваться на основе существующих наработок в области изучения социального капитала, а также исследований, затрагивающих специфику коммуникации в интернет-пространстве.

Одним из важных направлений в сетевом анализе *GitHub* является анализ различных метрик взаимодействия разработчиков. Исследователи [Batista et al, 2017] приходят к выводу, что большинство *open source*-разработчиков активно работают в нескольких репозиториях, а количество связей между этими разработчиками варьируется в зависимости от используемого языка программирования, хотя лишь немногие пары разработчиков взаимодействуют более чем в одном репозитории.

Сеть сотрудничества представлена исследователями как взвешенный граф, то есть граф, связи (ребра) которого имеют определенное значение. Значение (или вес) связей показывает силу сотрудничества между разработчиками. В случае с разработчиками на *GitHub* это количество репозиторий, в которых приняли участие оба программиста. Стоит отметить, что возможен более конкретизированный подход, при котором могут рассматриваться также пул-реквесты с изменением кода проекта (или одобрение пул-реквестов одного разработчика другим).

Изучение влияния в сетях *GitHub*

Отдельная область в рамках сетевых исследований *GitHub* — это изучение влияния в сообществах платформы. Авторы одного из исследований [Badashian, Stroulia, 2016] задаются вопросом: что такое «влияние» в социальных сетях разработчиков ПО? Они сравнивают три показателя влияния:

- количество подписчиков;
- количество форков (ответвлений), созданных на основе проектов пользователя, — это показатель ценности кода разработчика в глазах других пользователей (ценности, выражаемой в повторном использовании);
- количество подписчиков или наблюдателей (*watchers*) у проектов пользователя.

В результате исследователям удалось обнаружить, что влияние разработчика редко распространяется более чем на один язык программирования, хотя среди опытных программистов часто встречается хорошее знание нескольких языков программирования. Таким образом, влияние в контексте *GitHub* является преимущественно узкоспециализированным и связанным с каким-то конкретным инструментом.

Следует отметить, что для понимания взаимодействия разработчиков между собой довольно важно анализировать технологический стек, который они используют (прежде всего, это язык программирования и дополнительные инструменты, применяемые для разработки). Это та область, в которой социологическое изучение сообществ разработчиков связано с социологией языков программирования, которая пока что не оформилась в качестве самостоятельной сферы исследований. Такие исследования в рамках социологии науки и технологий имеют большой потенциал, причем эти исследования должны включать не только изучение использования языков программирования, но также изучение создания и развития языков программирования сообществами разработчиков.

Помимо влиятельности разработчиков в сетевых исследованиях *GitHub* также затрагивается тема различной степени влиятельности отдельных репозиторий. Это могут быть как репозитории каких-то программ, так и репозитории языков программирования или фреймворков, связанных с языками программирования. В одном из исследований изучается сеть звезд, то есть сеть подписок пользователей на репозитории проектов [Hu et al., 2016]. Для анализа этой сети используется алгоритм HITS (Hyperlink Induced Topic Search). На основе анализа ежемесячных изменений исследователи описывают динамику изменения влиятельности репозитория в сообществе *GitHub*.

Авторы другого исследования обращаются к изучению пользовательской популярности и мотивации отслеживания (то есть подписки на пользователя или проект) [Blincoe et al., 2016]. Это исследование представляет собой анализ сетей подписок, выполненный с помощью смешанных методов: был проведен опрос 800 пользователей *GitHub*, а также количественный анализ активности на сайте. Исследование показало, что на *GitHub* существует несколько типов лидерства и влияния — один связан с вкладом в проекты, а другой — с популярностью. Второй тип влияния встречается в случае, когда пользователь является одним из разработчиков какого-то популярного продукта (языка программирования, фреймворка, библиотеки и т. д.). Авторы обнаруживают, что популярные пользователи привлекают подписчиков к новым проектам. В рамках социологии науки и технологий возможны исследования участия популярных пользователей в разработке ПО с целью изучения того, как провоцирует ли их вовлеченность возникновение качественных инноваций в проектах.

Один из распространенных типов социальной структуры сообществ на *GitHub* — это так называемая пирамидальная меритократия — централизация без авторитар-

ного лидерства. Новые участники проекта находятся в нижнем эшелоне пирамиды, а выше — их коллеги по степени влиятельности. Наверху располагается «совет старейшин», которые чаще всего не имеют абсолютной власти над проектом, но к их мнению прислушивается большинство участников. Голос «старейшин» может быть решающим в случае каких-то споров. В качестве такого «старейшины» можно вспомнить автора языка *Python* Гвидо ван Россума, который в течение длительного времени (до 2018 г.) был «великодушным пожизненным диктатором» *Python* и решал возникающие споры относительно будущего проекта.

Одним из ключевых методов сетевого анализа является анализ эгоцентрических сетей, то есть сетей вокруг конкретного человека. В ходе изучения эго-сетей *GitHub* исследователи пришли к выводу, что на платформе можно выделить два типа пользователей [Mohan et al., 2021]. Первый тип — это новые пользователи или те пользователи, которые редко заходят на сайт или редко вносят изменения в репозитории (что возможно без посещения сайта — через интерфейс командной строки). Они склонны формировать вокруг себя простые сети (4–6 участников). Как отмечают авторы статьи, такие пользователи слабо влияют на общую культуру платформы *GitHub*. Второй тип пользователей — это влиятельные разработчики, которые имеют большое количество подписчиков, тем самым образуя эго-сети, которые могут включать десятки тысяч пользователей.

В рамках изучения сетевого взаимодействия на *GitHub* было обнаружено, что репутация участников сообщества привлекает сторонних разработчиков, однако мотивация участия в проектах этим не ограничивается [Celińska, 2018]. Существует также мотивация взаимности: разработчики участвуют в проектах друг друга (автор статьи называет это экономикой дара в рамках системы *GitHub*). Отдельный аспект взаимодействия сообществ на платформе, затронутый в упомянутом исследовании, заключается в наличии технологической стандартизации: проекты на популярных языках программирования обычно имеют больше соавторов (соразработчиков). Соответственно, следование авторов проекта технологическим стандартам (связанным не только с языками программирования, но и с определенными практиками) может способствовать вовлечению новых участников в сетевые структуры взаимодействия. Особенности влияния выбранного технологического инструментария на разработку проекта могут представлять интерес для социологии технологий.

Структурные особенности сетей *GitHub*

Сети в рамках *GitHub* могут иметь различную структуру. Авторы одного из исследований [Mezouar et al., 2019] предлагают для изучения социальной структуры команд разработчиков взять за основу процесс проверки пул-реквестов, то есть проверки запросов на включение изменений в проект. Когда пользователь *GitHub* предлагает внести изменения в код репозитория, то ответственным за проект необходимо оценить эти изменения и проверить их безопасность. Эффективность этой проверки зависит как от технических, так и от социальных факторов. В рамках исследования на платформе было изучено 7 850 наиболее популярных проектов, созданных на десяти языках программирования. В результате было выявлено несколько командных структур — около трети проектов имеют структуру единоличного управления, а команда этих проектов слабо связана. Слабая связность и единолич-

ность управления могут быть связаны с высоким уровнем компетенции создателя (владельца) репозитория. Если в сообществе вокруг проекта есть несколько кластеров пользователей, то создатель проекта или кто-то другой является связующим звеном между этими кластерами, своего рода мостом, закрывающим структурную дыру (structural hole) в сети взаимодействия.

В одном из исследований [Li et al., 2017] обнаруживается, что если в проекте есть структурные дыры и владелец проекта представляет собой посредника между несколькими частями сети (например, разработчики проекта могут являться представителями различных областей программирования или вообще иметь иную специализацию), то решения о принятии изменений в проект, как правило, происходят быстрее. Авторы исследования связывают это с большим уровнем профессионализма у создателей таких репозиториях — у них обычно больше опыт и они знакомы с разными практиками и стилями программирования, поэтому могут быстрее реагировать на предлагаемые изменения в коде проекта.

Тема анализа структурных дыр в сообществах и посредников между кластерами пользователей изучается не только в контексте структуры сети, но и в поведенческом контексте [Fu et al., 2021]. Авторы исследования обращаются к истории участия разработчика в проектах (как уже было отмечено ранее, *GitHub* позволяет смотреть вклад всех участников в разработку проекта), что дает возможность для изучения взаимодействий, в которые пользователь вступает в определенные периоды. Благодаря этому социолог может рассматривать сети взаимодействия в темпоральном аспекте. Темпоральный (временной) сетевой анализ является, пожалуй, наиболее перспективным в сфере сетевых исследований *GitHub*, поскольку отличительной особенностью разработки ПО на платформе *GitHub* является фиксация конкретных изменений, сделанных пользователем. То есть сети *GitHub*, в отличие, например, от сетей академического соавторства, позволяют рассмотреть вклад конкретного разработчика в проект. Также этот темпоральный аспект возможно рассматривать в контексте личностных траекторий, связанных с карьерой и интересами, что позволяет рассматривать данные об участии пользователей в проектах в контексте направления *social sequence analysis* (анализ социальных последовательностей).

Особенности взаимодействия пользователей в сетевых сообществах *GitHub*

Взаимодействие в сообществах, связанных с технологическими инновациями, часто описывалось посредством выявления неких идеальных типов. Сфера разработки ПО не является исключением. Однако в рамках исследований сетевых сообществ на *GitHub* было обнаружено, что привычные подходы в типологии моделей взаимодействия в командах разработчиков не всегда актуальны для этой платформы. Так, например, уже ставшая классической типология Эрика Реймонда уже не в полной мере описывает положение дел в *open source*-разработке. Известный разработчик и идеолог *open source*-движения Реймонд выдвинул следующее противопоставление: иерархическая модель (собор, cathedral) противопоставляется децентрализованной или модульной (базар, bazaar) [Raymond, 1999]. Как показывают исследователи [Zöller et al., 2020], для *GitHub* характерно несколько способов взаимо-

действия и стратегий поведения разработчиков. Более подходящей авторы считают следующую типологию:

- группы, имеющие четкую иерархию. Как правило, это сотрудничество под руководством корпораций (например, *Microsoft* или *Google*);
- меритократические группы. Чаще всего это проекты под эгидой некоммерческих организаций;
- группы, выстраивающиеся вокруг одного или нескольких разработчиков;
- горизонтальные группы, состоящие из разработчиков-добровольцев.

Авторы также отмечают, что существует два основных вида участия в совместной разработке — это не только создание программного кода, но также и обсуждение (например, принятие решений относительно включения или невключения каких-то изменений в код проекта). В исследовании также отмечается, что в работе команд на *GitHub* возникает гораздо меньшая информационная асимметрия (как в *open source*-разработке, так и в работе над проектами, недоступными внешним разработчикам) по сравнению с традиционной бюрократией. Исследователи говорят о том, что доступность информации в рамках проектов на *GitHub* отражается на топологии группы.

В одном из исследований [Newton et al., 2018] взаимодействие в командах разработчиков на *GitHub* рассматривается в контексте возникающих взаимозависимостей. При работе в команде может возникнуть несколько видов взаимозависимостей — это объединенная взаимозависимость (когда результат команды представляет собой сумму результатов разработчиков), последовательная взаимозависимость (когда для продвижения работы одного разработчика требуется результат работы другого), взаимная взаимозависимость (в случае, когда прогресс одного разработчика зависит от вклада второго, а второго — от вклада первого), интенсивная взаимозависимость (когда команда сотрудничает в диагностике и решении проблем). Исследователи предлагают рассматривать координацию разработчиков в двух видах — явном и неявном. Неявная координация бывает в том случае, когда некая проблема решается без спланированного и/или централизованного решения: разработчик видит, что нечто работает некорректно или может в дальнейшем вызвать проблемы, и исправляет это. Этот аспект взаимодействия связан с открытым характером *GitHub*.

Исследования открытости разработки ПО и роли внешних разработчиков в проектах *GitHub* связывают открытость с инновациями и производительностью. В одном из исследований [Lee et al., 2020] в качестве исследовательского фреймворка используется теория организационного обучения и в качестве показателя производительности рассматривается количество коммитов в репозитории. Авторы отмечают, что этот критерий должен быть усовершенствован и во внимание должно приниматься также качество коммитов. В результате исследования авторы приходят к выводу о взаимосвязи производительности проекта и открытости для участия в нем сторонних разработчиков.

Также важно отметить исследования, посвященные выявлению отличий в сетевом взаимодействии пользователей *GitHub* по сравнению с пользователями других социальных сетей. Поскольку платформа *GitHub* — это не просто хостинг для проектов и совместной их разработки, но также и социальная сеть, то к нему могут быть применимы те методы изучения, которые используются в отношении социальных сетей. Исследователи [Strzalkowski et al., 2019] задаются рядом вопросов: какие особенности имеет *GitHub* как социальная сеть? что отличает его от других социальных

сетей? В рамках своего исследования авторы предлагают изучать взаимодействие пользователей в социальных сетях через использование концепции информационного потока (в случае с другими социальными сетями примером информационного потока будет лента новостей). Однако сложность заключается в том, что на *GitHub* нет таких элементов информационного потока, как посты и твиты, отображающиеся в ленте. В связи с этим авторы предлагают считать единицей информационного потока *GitHub* отдельный репозиторий — то, что публикуется (информация о создании репозитория отображается в ленте новостей) и распространяется (посредством упоминания об участии в разработке проекта в ленте у подписчиков того или иного разработчика). Репозитории также включают в себя обсуждения развития проекта (разделы *Issues* и *Discussions*). На основании пул-реквестов или участия в обсуждении проекта авторы исследования предлагают выстраивать цепи событий, которые являются частью сетей диалогов *GitHub*. Такие сети, основанные не только на вкладе пользователя в код проекта, но и на его участии в обсуждениях, расширяют концепцию сетей пул-реквестов.

Поскольку одним из базовых аспектов *GitHub* является коммуникация — рабочая или связанная с хобби, — то в рамках изучения *GitHub* актуален вопрос того, какие каналы коммуникации существуют на платформе и как они функционируют. Одно из исследований [*Tantisuwankul et al.*, 2019] посвящено топологическому анализу таких каналов коммуникации. Авторы исследования основывают свое понимание концепции «обмен знаниями» на модели SECI (социализация, экстернализация, комбинация, интернализация), разработанной теоретиками менеджмента Икудзиро Нонакой и Хиротака Такеути. Эта модель основана на идее неявного знания (*tacit knowledge*) британского философа науки Майкла Поланьи. Опираясь на концепцию Поланьи, авторы выделяют две формы знания — явную и неявную. Явное знание (*know-what*) представляет собой то, что легко сформулировать и кодифицировать, а значит, и передать кому-то другому (например, некие факты). Неявное знание (*know-how*) менее подвластно систематизации и передаче, поскольку представляет собой интуитивное понимание контекста, возникающее исходя из личного опыта. Модель SECI описывает переход от явных и неявных знаний к организационным знаниям. Этот подход можно соотнести с упомянутым выше исследованием [*Li et al.*, 2017] о посредниках между кластерами пользователей, которыми чаще всего оказываются опытные разработчики, понимающие различные стили программирования и контексты выполнения задач.

Одно из наиболее ранних сетевых исследований *GitHub* посвящено изучению клик (*cliques*) разработчиков, то есть таких подмножеств сети, в которых все узлы соединены связями [*Weber*, 2012]. Примером клики будет команда друзей-программистов, которые добавили друг друга в друзья (если речь идет о сети «пользователь — пользователь»), или когда все представители команды участвуют в проектах друг друга (если речь о сети пул-реквестов). На примере изучения кластеров сообщества разработчиков, использующих язык *Python*, авторы показывают обнаружение клики ученых-физиков. Следует отметить, что клика — это понятие, относящееся к структуре сети. Для более полного социологического понимания клики разработчиков как сообщества полезно дополнить данные сетевого анализа информацией о специализации пользователя, его географической локации, месте работы и т. д. Отсюда необходимость анализа биографических данных, которые часто, хотя и не всегда, включаются пользователями в описание профиля.

Исследователями было обнаружено, что взаимодействие пользователей на *GitHub* в основном опосредовано репозиториями, а взаимность социальных связей за пределами репозитория (в виде подписок пользователей друг на друга) имеет низкий уровень [Lima et al., 2014]. При этом активные пользователи (те, кто вносит больше всего изменений в код проектов) не обязательно имеют большое число подписчиков. Одним из интересных элементов этого исследования является анализ влияния географической близости на сотрудничество между пользователями. Авторы обнаруживают, что люди склонны участвовать в проектах с теми, кто находится ближе к ним с точки зрения географии.

Географический аспект сетевого взаимодействия на *GitHub*

В дальнейшем влияние географического фактора на взаимодействие в сообществах *GitHub* рассматривалось в исследовании, посвященном изучению предвзятости и ее связи с географической локацией разработчиков [Rastogi, 2016]. Взаимодействие описывается в контексте сети пул-реквестов, представляющей собой ключевую модель сотрудничества на платформе *GitHub*. Исследовательница рассматривает, как происходит одобрение запросов на включение изменений и каким образом географическая локация пользователя, указанная в его профиле, может повлиять на принятие таких решений. В рамках исследования используются смешанные методы: анализируется сеть пул-реквестов в проектах и проводится опрос среди сабмиттеров (тех, кто делает запрос на внесение изменений) и интеграторов (тех, кто принимает решение о включении или невключении изменений в проект). В сети рассмотрено 70 740 запросов на внесение изменений в проекты. В опросе поучаствовали 2 532 респондента. В результате выявлено, что общность локаций сабмиттера и интегратора повышает вероятность принятия изменений в проект, то есть в сотрудничестве в *open source*-сообществах на *GitHub* играет важную роль местонахождение разработчика (в случае, если оно указывается в профиле). Также было обнаружено, что существует градация локаций с точки зрения их восприятия в контексте включения или невключения изменений в проект. Так, например, гораздо большие шансы на принятие пул-реквестов имеют сабмиттеры из Великобритании, Канады, Японии, Нидерландов и Швейцарии.

Дополнением к анализу влияния географического фактора на построение сетей взаимодействия на *GitHub* может стать подход, предлагаемый в одном из исследований, совмещающих анализ пространственного и временного аспекта взаимодействия [Ma et al., 2017]. Авторы исследования изучают поведение пользователей платформы, в частности, рассматривают темпоральный аспект взаимодействия. В результате они обнаруживают, что активность пользователей в течение дня, недели, месяца или года может меняться в зависимости от географической локации пользователя. К примеру, исследователи обнаружили, что китайские программисты снижают активность разработки во время китайского Нового года.

Темпоральный аспект сетевого взаимодействия на *GitHub*

Темпоральный аспект сетевого взаимодействия на *GitHub* изучается также в контексте длительности сотрудничества. Исследователи рассматривают, как сохраняются или распадаются команды *open source*-разработчиков с течением времени. Большинство таких команд выстраиваются вокруг одного проекта и, соответственно, являются временными, то есть чаще всего разработчики не продолжают сотрудничество за пределами проекта [Wang et al., 2019]. Авторы исследования предлагают понятие межрепозиторных социально-связанных команд (Cross-Repository Socially Connected Teams, CRSC Teams). В ходе исследования было выяснено, что CRSC-команды тесно связаны между собой и в них практически не возникают кластеры. Они могут иметь высокий уровень централизации. На формирование таких команд положительно влияет популярность темы проекта, а популярность языка программирования, напротив, влияет отрицательно.

Темпоральность в сетях *GitHub* изучается также в контексте внутренней динамики сообществ вокруг *open source*-проектов. Авторы исследования, посвященного изучению роста и угасания известности участников *open source*-проектов, создают типологию динамики перемещения пользователя от периферии сообщества к его ядру [El Asri, 2017]. Под периферийным участием считается самый первый вклад (коммит) в проект. В ядре сообщества состоят основные участники, принимающие изменения в код проекта. Авторы статьи изучили виды деятельности, которые могут повлиять на динамику, — это коммиты (внесение изменений в исходный код), комментирование коммитов и т. д. Ключевым действием, которое влияло на динамику перехода из периферии проекта в его ядро, оказалось внесение коммитов в проект. Чем чаще пользователь вносит изменения в исходный код, тем быстрее и с большей вероятностью он станет частью основной команды проекта. Снижение активности пользователя может привести к тому, что он покинет ядро проекта, однако это, как было обнаружено исследователями, связано также с этапом, на котором находится разработка проекта.

Изучение поведения начинающих пользователей в рамках сетевой коммуникации *GitHub* осуществлялось также в контексте сотрудничества между университетами и компаниями. В результате исследованиями было обнаружено, что студенты, участвующие в проектах на *GitHub*, склонны к подражанию, а не к инновациям [Cheng et al., 2020]. Поскольку *GitHub* активно используется в рамках преподавания, то довольно большое число пользователей являются студентами университетов. В рамках социологии науки возможно изучение того, как функционируют научные сообщества на *GitHub* и как на этой платформе осуществляется академическая коммуникация на уровне университетов, образовательных организаций, регионов и стран.

При изучении паттернов смешивания степеней (degree mixing patterns) в сетях коллабораций на *GitHub* обнаруживается следующая закономерность: существует тенденция узлов высокого уровня (то есть имеющих высокую степень взаимодействия) быть связанными с узлами меньшего уровня [Allaho, Lee, 2013]. Это означает, что при разработке программного обеспечения опытные разработчики не замыкаются в кластере экспертов, но активно сотрудничают с новыми разработчиками. Также в другом исследовании тех же авторов было обнаружено, что разработчики, как правило, принимают участие в *open source*-проектах, команды которых имеют

одинаковый размер [Allaho, Lee, 2015]. Различие между коммерческой разработкой программного обеспечения заключается в том, что в *open source*-разработке большую роль играет мотивация, приверженность разработчика проекту и/или сообществу. Также авторами выявляется положительная корреляция между центральностью (в четырех формах — входящая степень, исходящая степень, степень посредничества и степень близости) и вкладом разработчика в проект.

Поскольку платформа *GitHub* позволяет изучать сети на нескольких уровнях, исследователи предлагают различные варианты возможного обобщения сетевых структур платформы. Так, например, в одном из исследований [Hu et al., 2018] предлагается совместить сети, создаваемые на основе четырех видов связей — подписки на людей, подписки на проекты (звезды), форки (ответвления), пул-реквесты (активность по внесению изменений в проекты). В другом исследовании [Oliveira et al., 2022] авторы также предлагают перейти от изучения однородных сетей к гетерогенным сетям. Концепция гетерогенных сетей, представленная авторами, включает те же связи, что и в упоминаемом выше исследовании, но помимо этого подразумевает, что в анализе необходимо учитывать время совместной работы пользователей в репозитории. Это делает актуальным обращение к изучению темпорального аспекта взаимодействия на *GitHub*.

Сетевой анализ платформы *GitHub* в контексте социологии науки и технологий

Платформа *GitHub* представляет особый интерес для социологии науки и технологий, поскольку представляет собой среду, в которой осуществляется научная и технологическая коммуникация. Опираясь на рассмотренные исследования, можно сделать выводы о том, какие аспекты сетевого анализа *GitHub* являются наиболее актуальными в контексте социологического изучения науки и технологий.

Прежде всего, стоит отметить формирование теории социального капитала в контексте изучения интернет-сообществ разработчиков. Благодаря ей открываются возможности для исследования предпосылок научных и технологических инноваций, связанных с программированием. Кроме того, важным элементом изучения таких инноваций может стать социологическое рассмотрение функционирования того технологического инструментария, который используется в сообществах программистов. Прежде всего, речь о языках — как они возникают в сообществах, какие предпосылки способствуют их появлению, как они меняются и какие факторы влияют на их популярность. Поскольку создание и развитие языков программирования зачастую связано с академическим миром (а именно — с исследователями в области компьютерных наук), то этот аспект будет весьма важен для социологии науки.

Изучение сетей *GitHub* может дополнять сетевой анализ научного сотрудничества, осуществляемого в привычном публикационном формате. В связи с этим стоит отметить большое количество исследований по библиометрическому сетевому анализу, осуществляемых отечественными социологами [Мальцева и др., 2024; Мальцева, Ким, 2022]. Поскольку многие из этих исследований посвящены сетевому анализу сообщества российских социологов, то сетевой анализ *GitHub* мог бы дополнить их и помочь в изучении интернет-сообщества социологов, которые задей-

ствованы в разработке и развитии программного обеспечения, используемого для социологических целей. Весьма вероятно, что изучение сетей платформы *GitHub* способно обогатить выстраиваемое исследователями описание научной коммуникации в российской и международной науке.

Следует отметить, что для социологии науки и технологий особенный интерес представляет именно изучение сети форков на *GitHub*, поскольку именно эта сеть представляет собой пространство технологических инноваций (или, как было отмечено ранее, пространство «видеообразования» программного обеспечения). В рамках разработки ПО передовые проекты, создаваемые с нуля, не так распространены, как инновации, возникающие через ответвления от существующих проектов.

Рассматривая исследования, связанные с географическим аспектом взаимодействия на *GitHub*, можно отметить, что, несмотря на преодоление ряда препятствий, затруднявших научную коммуникацию и технологическое сотрудничество (прежде всего это были препятствия, связанные с невозможностью удаленного взаимодействия между разработчиками), все еще существуют факторы, которые мешают такой коммуникации, — это не только уровень образования и профессионализма или репутация (разработчики из одних стран могут рассматриваться в качестве более компетентных, а потому с большей вероятностью приниматься в сети сотрудничества), но также культурно-языковые различия, которые влияют на коммуникацию разработчиков.

Социологическое изучение разработки ПО связано также с непосредственными рабочими процессами. Отсюда необходимость сочетать различные методы исследования. Важным шагом в этом направлении стала статья, посвященная цифровой этнографии в контексте изучения платформенной занятости [Ritter, 2023]. Автор показывает, как этнографические полевые исследования могут быть объединены с цифровыми методами, в частности с сетевым анализом. В исследовании рассматривается переплетение цифровых и физических контекстов, а именно то, как влияет на труд проникновение цифровых платформ в рабочие процессы. Автор исследования приходит к выводу, что сетевой анализ цифровых платформ, на которых осуществляется взаимодействие, позволяет обнаружить то, что не выявляется с помощью стороннего наблюдения, участия социолога в рабочем процессе или изучения бюрократических структур, в которых находятся работники. Хотя офлайн-взаимодействие работников имеет большое значение для изучения их профессиональной деятельности, исследование взаимодействия на уровне цифровых платформ может позволить более полно изучить рабочие процессы. Например, как уже было отмечено ранее, *GitHub* дает возможность для анализа вклада каждого конкретного работника в общую деятельность коллектива. Социологическое изучение разработки ПО невозможно без изучения платформы *GitHub*, поскольку она представляет собой основную площадку для коммуникации и взаимодействия программистов. Отсюда важность рассматриваемых исследований для социологии технологий.

Одной из разновидностей коммуникации на платформе *GitHub* является взаимодействие научных работников, использующих платформу для преподавания или публикации программного обеспечения, используемого в исследованиях. Поэтому применение сетевого анализа в этой области весьма актуально, однако стоит отметить, что исследований по этой теме мало и они, как правило, связаны с взаимодействием университетов и коммерческих компаний в области ИТ [Cheng et al., 2020]. Иногда исследователи в рамках общего сетевого анализа *GitHub* рассматривают со-

общества ученых [Weber, 2012], однако эти исследования осуществляются в отрыве от традиции социологического изучения науки и производства научного знания. Можно сделать вывод о том, что синтез достижений социологии науки и возможностей сетевого анализа способен многое дать изучению научной и технологической коммуникации в интернет-сообществах. Особенно это актуально в контексте распространения цифровых методов в гуманитарных и социальных науках: такие направления, как *computational social science* (вычислительные социальные науки) и *digital humanities* (цифровые гуманитарные науки), непосредственно связаны с размещением программного обеспечения на *GitHub*, а также с использованием этой платформы для преподавания.

Заключение

Следует сказать о перспективах в области сетевого анализа сообществ *GitHub*, и прежде всего тех возможностях, которые открываются перед социологами науки и технологий в этом контексте. На основе анализа исследований в этой области можно прийти к выводу, что наиболее изученной здесь является тема сетевого взаимодействия *open source*-разработчиков. Исходя из этого, перспективным шагом могло бы стать включение этих наработок в сферу социологии науки и технологий, а также социологии интернет-сообществ (как профессиональных сообществ программистов, так и сообществ по интересам). Другой областью социологии, которая могла бы получить пользу от сетевого анализа платформы *GitHub*, является социология труда.

Как уже было сказано, на *GitHub* представлены не только с профессиональные сообщества программистов, но также сообщества программистов-любителей и, что следует отметить отдельно, сообщества ученых. Поскольку использование программирования широко распространено в современном академическом сообществе, социология науки могла бы использовать сетевой анализ для изучения коммуникации ученых в рамках *GitHub*. В настоящий момент академические сообщества на *GitHub* слабо изучены. Однако такие исследования могли бы дать более глубокое понимание того, как происходит взаимодействие и коммуникация в современной науке, особенно если сетевой анализ *GitHub* будет осуществляться в сочетании с библиометрическим анализом научных сетей.

Также следует сказать, что для более плодотворного социологического изучения *GitHub* социологам науки и технологий необходимо концептуальное осмысление тех наработок, которые существуют в виде эмпирических исследований платформы. Для этого необходимо поместить эти наработки в контекст современной социологической теории, к примеру, рассмотреть их через методологию, предлагаемую реляционной социологией — направлением, которое является довольно близким к социальному сетевому анализу. Если такая интеграция теоретической социологии и эмпирических исследований сетевого взаимодействия на *GitHub* будет осуществлена, то это поможет дальнейшему развитию социологии научной и технологической коммуникации в интернет-сообществах.

Литература

- Мальцева Д.В., Ким А.В.* Профессиональное онлайн-сообщество российских социологов: тематика обсуждений и структура социальных взаимодействий // Мониторинг общественного мнения: Экономические и социальные перемены. 2022. № 3 (169). С. 151–174.
- Мальцева Д.В., Ким А.В., Шеглова Т.Е.* Изучение сетей коллабораций российских социологов: методологический подход и его апробация // Вестник Томского гос. ун-та. Философия. Социология. Политология. 2024. № 79. С. 221–232.
- Allaho M.Y., Lee W.-C.* Analyzing the Social Networks of Contributors in Open Source Software Community // Applications of Social Media and Social Network Analysis. Lecture Notes in Social Networks / Eds. P. Kazienko, N. Chawla. Cham: Springer, 2015. P. 57–75.
- Allaho M.Y., Lee W.-C.* Analyzing the Social Ties and Structure of Contributors in Open Source Software Community // Proceedings of the 2013 IEEE / ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13). 2013. P. 56–60.
- Badashian A.S., Stroulia E.* Measuring User Influence in GitHub: The Million Follower Fallacy // Proceedings of the 3rd International Workshop on CrowdSourcing in Software Engineering. 2016. P. 15–21.
- Bana R., Arora A.* Influence Indexing of Developers, Repositories, Technologies and Programming Languages on Social Coding Community GitHub // 2018 Eleventh International Conference on Contemporary Computing (IC3). 2018. P. 1–6.
- Batista N.A., Brandão M.A., Alves G.B., da Silva A.C., Moro M.M.* Collaboration Strength Metrics and Analyses on GitHub // Proceedings of the International Conference on Web Intelligence. 2017. DOI: 10.1145/3106426.3106480.
- Biazzini M., Baudry B.* “May the Fork Be with You”: Novel Metrics to Analyze Collaboration on GitHub // Proceedings of the 5th International Workshop on Emerging Trends in Software Metrics. 2014. P. 37–43.
- Blincoe K., Sheoran J., Goggins S., Petakovic E., Damian D.* Understanding the Popular Users: Following, Affiliation Influence and Leadership on GitHub // Information and Software Technology. 2016. Vol. 70. P. 30–39. DOI: 10.1016/j.infsof.2015.10.002.
- Celińska D.* Coding Together in a Social Network: Collaboration among GitHub Users // Proceedings of the 9th International Conference on Social Media and Society (SMSociety '18). 2018. P. 31–40. DOI: 10.1145/3217804.3217895.
- Cheng X., Zhang Z., Yang Y., Yan Z.* Open Collaboration between Universities and Enterprises: A Case Study on GitHub // Internet Research. 2020. Vol. 30. No. 4. P. 1251–1279. DOI: 10.1108/INTR-01-2019-0013.
- El Asri I., Kerzazi N., Benhiba L., Janati M.* From Periphery to Core: A Temporal Analysis of GitHub Contributors' Collaboration Network // Collaboration in a Data-Rich World. Vol. 506 / Eds. L.M. Camarinha-Matos, H. Afsarmanesh, R. Fornasiero. Cham: Springer International Publishing, 2017. P. 217–229.
- Fu E., Zhuang Y., Zhang J., Zhang J., Chen Y.* Understanding the User Interactions on GitHub: A Social Network Perspective // 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). 2021. P. 1148–1153.
- Hu Y., Wang S., Ren Y., Choo K.-K.R.* User Influence Analysis for GitHub Developer Social Networks // Expert Systems with Applications. 2018. Vol. 108. No. 3. P. 108–118. DOI: 10.1016/j.eswa.2018.05.002.
- Hu Y., Zhang J., Bai X., Yu S., Yang Z.* Influence Analysis of GitHub Repositories // SpringerPlus. 2016. Vol. 5. No. 1. DOI: 10.1186/s40064-016-2897-7.
- Jiang J., Lo D., He J., Xia X., Kochhar P.S., Zhang L.* Why and How Developers Fork What from Whom in GitHub // Empirical Software Engineering. 2017. Vol. 22. No. 1. P. 547–578. DOI: 10.1007/s10664-016-9436-6.

Lee S., Baek H., Oh S. The Role of Openness in Open Collaboration: A Focus on Open-Source Software Development Projects // *ETRI Journal*. 2020. Vol. 42. No. 1. P. 196–204. DOI: 10.4218/etrij.2018-0536.

Li L., Goethals F., Baesens B., Snoeck M. Predicting Software Revision Outcomes on GitHub Using Structural Holes Theory // *Computer Networks*. 2017. Vol. 114. P. 114–124. DOI: 10.1016/j.comnet.2016.08.024.

Lima A., Rossi L., Musolesi M. Coding Together at Scale: GitHub as a Collaborative Social Network // arXiv: 1407.2535v1. 2014.

Ma Y., Li H., Hu J., Xie R., Chen Y. Mining the Network of the Programmers: A Data-Driven Analysis of GitHub // *Proceedings of the 12th Chinese Conference on Computer Supported Cooperative Work and Social Computing*. 2017. P. 165–168.

Mezouar M.E., Zhang F., Zou Y. An Empirical Study on the Teams Structures in Social Coding using GitHub Projects // *Empirical Software Engineering*. 2024. Vol. 24. No. 3. P. 3790–3823. DOI:10.1007/s10664-019-09700-1.

Mohan P., Narayan P., Sharma L. Egocentric Analysis of GitHub User Network // *2021 2nd International Conference for Emerging Technology (INCET)*. 2021. P. 1–7.

Newton O.B., Fiore S.M., Song J. Developing Theory and Methods to Understand and Improve Collaboration in Open Source Software Development on GitHub // *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2018. Vol. 62. No. 1. P. 1118–1122. DOI: 10.1177/1541931218621256.

Oliveira G.P., Moura A.F.C., Batista N.A., Brandão M.A., Hora A., Moro M.M. How Do Developers Collaborate? Investigating GitHub Heterogeneous Networks // *Software Quality Journal*. 2023. Vol. 31. No. 1. P. 211–241. DOI: 10.1007/s11219-022-09598-x.

Rastogi A. Do Biases Related to Geographical Location Influence Work-Related Decisions in GitHub? // *Proceedings of the 38th International Conference on Software Engineering Companion*. 2016. P. 665–667.

Raymond E.S. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media, Sebastopol, CA, 1999. P. 21–22.

Ritter C. *Digital Ethnography: Understanding Platform Labour from Within* // *Methodological Approaches for Workplace Research and Management* / Eds. C. Tagliaro, M. Orel, Y. Hua. Abingdon, Oxon; New York, NY: Routledge, 2023. P. 54–69. DOI: 10.1201/9781003289845-4.

Strzalkowski T., Harrison T., Sa N., Katsios G., Khoja E. GitHub as a Social Network // *Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2018. Advances in Intelligent Systems and Computing*. Vol. 787 / Ed. T. Ahram. Cham: Springer, 2019. P. 379–390.

Tantisuwankul J., Nugroho Y.S., Kula R.G., Hata H., Rungsawang A., Leelaprute P., Matsumoto K. A Topological Analysis of Communication Channels for Knowledge Sharing in Contemporary GitHub Projects // *Journal of Systems and Software*, 158, 2019. DOI: 10.1016/j.jss.2019.110416.

Wang D., Cao J., Qian S., Qi Q. Investigating Cross-Repository Socially Connected Teams on GitHub // *2019 26th Asia-Pacific Software Engineering Conference (APSEC)*. 2019. P. 490–497.

Weber N.M. Combined Methods, Thick Descriptions: Languages of Collaboration on GitHub // *Proceedings of the American Society for Information Science and Technology*. 2012. Vol. 49. No. 1. P. 1–4. DOI: 10.1002/meet.14504901347.

Zöller N., Morgan J.-H., Schröder T. A Topology of Groups: What GitHub Can Tell Us About Online Collaboration // *Technological Forecasting and Social Change*. 2020. Vol. 161. No. 2. P. 120291.

A Review of Research in the Field of Network Analysis of *GitHub* Platform Communities

DMITRIY S. SEMENKO

The National Research University “Higher School of Economics”,
Moscow, Russia;
e-mail: dssemenko@hse.ru

Software development plays an important role in the modern world. That’s why researchers are increasingly turning to the study of programmers’ communities, as well as communication in these communities. Collaboration platforms are an important part of the social space of the programmers’ community. These are systems that are designed to host and manage projects. The *GitHub* platform is the largest of these systems. More than 100 million developers host their projects on *GitHub*, take part in the development of projects created by other programmers, or simply observe what happens to the projects they are interested in. *GitHub* is also actively used in the modern academic community for teaching or publishing software used for scientific purposes. This is why studying communities on this platform is important for the sociology of science and technology. Collaboration of software developers on the *GitHub* platform is one of the largest and most complex forms of social interaction on the Internet. This article analyzes the approaches that are proposed in SNA (social network analysis) research of communities on this platform. Most of these publications are related to the study of software development in open source communities. A review of these publications describes the types of networks that exist on the *GitHub* platform, their structural features and communication channels that allow software developers to interact. Approaches to studying the geographical and temporal aspects of network interaction are also considered. A review of publications in this area will help prepare the transition from the study of homogeneous networks to the study of heterogeneous networks (networks that include various entities and their connections — users, repositories, forks, subscriptions to users and repositories). It will also help to understand the potential of using social network analysis of *GitHub* communities within the framework of the sociology of science and technology, as well as the sociology of work.

Keywords: sociology of programming, programmers, developers, platforms, *GitHub*, networks, network analysis, communication, internet communities, open source software.

Acknowledgment

The research was carried out as part of the Program for Fundamental Research of National Research University “Higher School of Economics” (HSE University).

References

Allaho, M.Y., Lee, W.-C. (2015). Analyzing the Social Networks of Contributors in Open Source Software Community, in P. Kazienko, N. Chawla (Eds.), *Applications of Social Media and Social Network Analysis. Lecture Notes in Social Networks* (pp. 57–75), Cham: Springer.

Allaho, M.Y., Lee, W.-C. (2013). Analyzing the Social Ties and Structure of Contributors in Open Source Software Community, *Proceedings of the 2013 IEEE / ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*, 56–60.

Badashian, A.S., Stroulia, E. (2016). Measuring User Influence in GitHub: The Million Follower Fallacy, *Proceedings of the 3rd International Workshop on CrowdSourcing in Software Engineering*, 15–21.

Bana, R., Arora, A. (2018). Influence Indexing of Developers, Repositories, Technologies and Programming Languages on Social Coding Community GitHub, *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1–6.

Batista, N.A., Brandão, M.A., Alves, G.B., da Silva, A.C., Moro, M.M. (2017). Collaboration Strength Metrics and Analyses on GitHub, *Proceedings of the International Conference on Web Intelligence, August 23–26, 2017*. DOI: 10.1145/3106426.3106480.

Biazzini, M., Baudry, B. (2014). “May the Fork Be with You”: Novel Metrics to Analyze Collaboration on GitHub, *Proceedings of the 5th International Workshop on Emerging Trends in Software Metrics*, 37–43.

Blincoe, K., Sheoran, J., Goggins, S., Petakovic, E., Damian, D. (2016). Understanding the Popular Users: Following, Affiliation Influence and Leadership on GitHub, *Information and Software Technology*, vol. 70, 30–39. DOI: 10.1016/j.infsof.2015.10.002.

Celińska, D. (2018). Coding Together in a Social Network: Collaboration among GitHub Users, *Proceedings of the 9th International Conference on Social Media and Society (SMSociety '18)*, 31–40. DOI: 10.1145/3217804.3217895.

Cheng, X., Zhang, Z., Yang, Y., Yan, Z. (2020). Open Collaboration between Universities and Enterprises: A Case Study on GitHub, *Internet Research*, 30 (4), 1251–1279. DOI: 10.1108/INTR-01-2019-0013.

El Asri, I., Kerzazi, N., Benhiba, L., Janati, M. (2017). From Periphery to Core: A Temporal Analysis of GitHub Contributors' Collaboration Network., in L.M. Camarinha-Matos, H. Afsarmanesh, R. Fornasiero (Eds.), *Collaboration in a Data-Rich World*, vol. 506 (pp. 217–229), Cham: Springer International Publishing.

Fu, E., Zhuang, Y., Zhang, J., Zhang, J., Chen, Y. (2021). Understanding the User Interactions on GitHub: A Social Network Perspective, *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 1148–1153.

Hu, Y., Wang, S., Ren, Y., Choo, K.-K. R. (2018). User Influence Analysis for GitHub Developer Social Networks, *Expert Systems with Applications*, 108 (3), 108–118. DOI: 10.1016/j.eswa.2018.05.002.

Hu, Y., Zhang, J., Bai, X., Yu, S., Yang, Z. (2016). Influence Analysis of GitHub Repositories, *SpringerPlus*, 5 (1), 1268. DOI: 10.1186/s40064-016-2897-7.

Jiang, J., Lo, D., He, J., Xia, X., Kochhar, P.S., Zhang, L. (2017). Why and How Developers Fork What from Whom in GitHub, *Empirical Software Engineering*, 22 (1), 547–578. DOI: 10.1007/s10664-016-9436-6.

Lee, S., Baek, H. Oh, S. (2020). The Role of Openness in Open Collaboration: A Focus on Open-Source Software Development Projects, *ETRI Journal*, 42 (1), 196–204. DOI: 10.4218/etrij.2018-0536.

Li, L., Goethals, F., Baesens, B., Snoeck, M. (2017). Predicting Software Revision Outcomes on GitHub Using Structural Holes Theory, *Computer Networks*, vol. 114, 114–124. DOI: 10.1016/j.comnet.2016.08.024.

Lima, A., Rossi, L., Musolesi, M. (2014). Coding Together at Scale: GitHub as a Collaborative Social Network, *arXiv: 1407.2535v1*.

Ma, Y., Li, H., Hu, J., Xie, R., Chen, Y. (2017). Mining the Network of the Programmers: A Data-Driven Analysis of GitHub. *Proceedings of the 12th Chinese Conference on Computer Supported Cooperative Work and Social Computing*, 165–168.

Maltseva, D.V., Kim, A.V. (2022). Professional'noye onlayn-soobshchestvo rossiyskikh sotsiologov: tematiki obsuzhdeniy i struktura sotsial'nykh vzaimodeystviy [Professional online

community of Russian sociologists: topics of discussion and structure of social interactions], *Monitoring obshchestvennogo mneniya: Ekonomicheskiye i sotsial'nyye peremeny*, no 3 (169), 151–174 (in Russian).

Maltseva, D.V., Kim, A.V., Shcheglova, T.E. (2024). Izucheniye setey kollaboratsiy rossiyskikh sotsiologov: metodologicheskiy podkhod i yego aprobatsiya [Studying collaborative networks of Russian sociologists: a methodological approach and its testing], *Vestnik Tomskogo gos. un-ta. Filosofiya. Sotsiologiya. Politologiya*, no. 79, 221–232 (in Russian).

Mezouar, M.E., Zhang, F., Zou, Y. (2019). An Empirical Study on the Teams Structures in Social Coding using GitHub Projects, *Empirical Software Engineering*, 24 (3), 3790–3823. DOI:10.1007/s10664-019-09700-1.

Mohan, P., Narayan, P., Sharma, L. (2021). Egocentric Analysis of GitHub User Network, *2021 2nd International Conference for Emerging Technology (INCET)*, 1–7.

Newton, O.B., Fiore, S.M., Song, J. (2018). Developing Theory and Methods to Understand and Improve Collaboration in Open Source Software Development on GitHub, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 1118–1122. DOI: 10.1177/1541931218621256.

Oliveira, G.P., Moura, A.F.C., Batista, N.A., Brandão, M.A., Hora, A., Moro, M.M. (2023). How Do Developers Collaborate? Investigating GitHub Heterogeneous Networks, *Software Quality Journal*, 31 (1), 211–241. DOI: 10.1007/s11219-022-09598-x.

Rastogi, A. (2016). Do Biases Related to Geographical Location Influence Work-Related Decisions in GitHub? *Proceedings of the 38th International Conference on Software Engineering Companion*, 665–667.

Raymond, E.S. (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (pp. 21–22), O'Reilly Media, Sebastopol, CA.

Ritter, C. (2023). Digital Ethnography: Understanding Platform Labour from Within, in C. Tagliaro, M. Orel, Y. Hua (Eds.), *Methodological Approaches for Workplace Research and Management* (pp. 54–69), Abingdon, Oxon; New York, NY: Routledge. DOI: 10.1201/9781003289845-4.

Strzalkowski, T., Harrison, T., Sa, N., Katsios, G., Khoja, E. (2019). GitHub as a Social Network, in T. Ahram (Ed.), *Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2018. Advances in Intelligent Systems and Computing*, vol. 787 (pp. 379–390), Cham: Springer.

Tantisuwankul, J., Nugroho, Y.S., Kula, R.G., Hata, H., Rungsawang, A., Leelaprute, P., Matsumoto, K. (2019). A Topological Analysis of Communication Channels for Knowledge Sharing in Contemporary GitHub Projects, *Journal of Systems and Software*, vol. 158, 110416. DOI: 10.1016/j.jss.2019.110416.

Wang, D., Cao, J., Qian, S., Qi, Q. (2019). Investigating Cross-Repository Socially Connected Teams on GitHub, *2019 26th Asia-Pacific Software Engineering Conference (APSEC)*, 490–497.

Weber, N.M. (2012). Combined Methods, Thick Descriptions: Languages of Collaboration on GitHub, *Proceedings of the American Society for Information Science and Technology*, 49 (1), 1–4. DOI: 10.1002/meet.14504901347.

Zöller, N., Morgan, J.-H., Schröder, T. (2020). A Topology of Groups: What GitHub Can Tell Us about Online Collaboration, *Technological Forecasting and Social Change*, 161 (2), 120291.