УДК 004.65:[336.743-021.131]-047.44
https://doi.org/10.33186/1027-3689-2022-9-84-97

Блокчейн-анализ рынка биткоинов. (Часть 1)

Игорь Макаров[1], Антуанетта Шоар[2]

[1]Лондонская школа экономики Houghton Street Лондон WC2A 2AE Соединённое Королевство, i.makarov@lse.ac.uk
[2]Школа менеджмента MIT Sloan School of management 100 Main Street, E62-638 Кембридж, Массачусетс 02142 и NBER, aschoar@mit.edu

Аннотация. В настоящей статье представлен подробный анализ сети Биткоин (Bitcoin) и её основных участников, проведённый авторитетными специалистами – Игорем Макаровым из Лондонской школы экономики и Антуанеттой Шоар из Массачусетского технологического института – по поручению Национального бюро экономических исследований (NBER) – частной организации в США. Сеть Биткоин детерминируется как новая база данных, включающая большое количество общедоступных и проприетарных источников для связывания адресов биткоинов с реальными объектами и обширный набор алгоритмов для извлечения информации о поведении основных участников рынка. Анализ экосистемы Биткоин состоял из трёх основных этапов. Во-первых, проанализированы объём транзакций и сетевая структура основных участников блокчейна. Во-вторых, задокументированы концентрация и региональный состав майнеров, которые осуществляют проверку (верификацию) и обеспечивают целостность реестра блокчейна (гроссбуха, леджера). В-третьих, рассмотрена концентрация собственности крупнейших держателей биткоинов. Установлено, что владельцами трети всех выпущенных биткоинов являются 10 тыс. индивидуальных инвесторов. Делается вывод, что высокая концентрация делает рынок первой в мире криптовалюты уязвимым перед гипотетической атакой хакеров.

Переводчик статьи отмечает, что переложение текста с английского языка[1] на русский было весьма затруднительным в связи с новизной финансовой тематики и широким использованием его авторами распространённого на Западе, однако нового для нас термина *entity* (сущность). Несмотря на данный факт, представляется необходимым ознакомить читателей с технологией бит-

[1] Оригинальный текст: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf.

коинов, что будет иметь практическую пользу для библиотечно-информационного сообщества.

Ключевые слова: криптовалюта, биткоин, блокчейн, транзакции, майнеры, концентрация собственности

Для цитирования: Макаров И., Шоар А. Блокчейн-анализ рынка биткоинов. (Часть 1) / И. Макаров, А. Шоар // Научные и технические библиотеки. 2022. № 9. С. 84–97. https://doi.org/10.33186/1027-3689-2022-9-84-97

Введение

Созданная более десяти лет назад криптовалюта буквально совершила революцию в денежно-кредитной политике. Значительно поднявшись в цене, сегодня она находится в центре внимания широкой общественности и вызывает противоречивые мнения. Отличительной чертой криптовалюты является децентрализованная система платежей или средств сбережения вне традиционного государственного контроля. Технология блокчейн в основе криптовалюты заменяет зависимость от нескольких организаций, ведущих централизованные учётные записи, таких как банки или сети кредитных карт, на сеть с большим набором децентрализованных и анонимных участников – агентов. Отсутствие централизованной подотчётности и анонимность пользователей часто рассматриваются как основные преимущества, однако в то же время мешают своевременной диагностике состояния здоровья системы, создают множество проблем для регулирующих органов и вводят новые источники систематических рисков.

Биткоин, оригинальная криптовалюта, по-прежнему остаётся самой крупной и популярной монетой, причём её рыночная капитализация выше, чем у всех остальных монет, вместе взятых. Она часто рассматривается как шаблон или эталон для других новых монет. Сегодня многие призывают к ещё более широкому внедрению биткоинов – в качестве инструмента государственных инвестиций либо законного платёжного средства. Это давление ставит в сложное положение регулирующие органы, которые хотят найти правильный баланс между защитой общественных интересов и поддержкой инноваций. Несмотря на то, что биткоины существуют более десяти лет, по-прежнему имеет-

ся много открытых вопросов об их использовании, концентрации собственности, а также структуре основных элементов, которые составляют основу одноимённой экосистемы. Представляется необходимым проведение анализа сети Биткоин и её участников для принятия решения о том, как и каким образом интегрировать эту валюту в традиционную финансовую систему.

В данной статье нами была предпринята попытка дать ответы на эти вопросы с помощью моделирования новой базы данных, которая позволила нам задокументировать эволюцию рынка биткоинов и его различных участников. Для создания этой базы мы использовали большое количество общедоступных и проприетарных источников, которые связывают биткоин-адреса с реальными объектами, и разработали набор алгоритмов, имеющих коммерческий характер блокчейна Биткоин для извлечения информации о поведении основных участников рынка. Созданная нами база данных Биткоин на сегодняшний день считается наиболее полной из тех, которые используются в академических исследованиях.

Мы провели три базовых этапа анализа, в которых основное внимание было уделено наиболее крупным участникам экосистемы Биткоин. Во-первых, проанализировали объём транзакций и структуру сети основных участников блокчейна. Во-вторых, задокументировали концентрацию и региональный состав майнеров, обеспечивающих целостность реестра блокчейн (гроссбуха, леджера). И, наконец, рассмотрели концентрацию собственности крупнейших держателей биткоинов.

Объём транзакций и структура сети Биткоин. Для начала мы установили, что 90% объёма транзакций в блокчейне Биткоин не привязано к экономически значимой деятельности, но является побочным продуктом конструкции биткоин-протокола, а также анонимностью участников. Поскольку блокчейн Биткоин – это публичный реестр (гроссбух), все потоки оплаты между адресами прекрасно наблюдаются. Следовательно, многие биткоин-пользователи применяют стратегии, предназначенные для предотвращения их отслеживания, перемещая средства по длинным цепочкам из нескольких адресов и разделяя платежи между ними; в результате чего в большом количестве генерируются ложные объёмы транзакций. Мы разработали алгоритмы для их филь-

трации и фиксации экономически значимых платежей, проводимых между реальными организациями.

Мы доказали, что подавляющее большинство биткоин-транзакций между реальными объектами предназначено для торговых и спекулятивных целей. Начиная с 2015 г. 75% реального объёма транзакций биткоинов было связано с биржами или подобными биржам объектами, такими как онлайн-кошельки, внебиржевые столы, крупные институциональные трейдеры. Другие известные организации несут ответственность только за незначительную часть от общего объёма. Например, незаконные транзакции, мошенничество и азартные игры вместе составляют менее 3%. Доля объёма, связанная с майнерами, ещё меньше.

Биржи не только генерируют наибольший объём, но и являются наиболее активно связанными узлами в сети Биткоин. В частности, у них самый высокий показатель сконцентрированности собственных значений. Кроме того, большая часть объёма транзакций состоит из перекрёстного (межбиржевого) обмена. Высокий кросс-обмен потоков является следствием современной структуры рынка. В отличие от традиционных регулируемых бирж, рынки криптовалют состоят из множества неинтегрированных и независимых бирж без каких-либо гарантий того, что инвесторы получат самые крупные суммы при совершении сделок. В результате согласованность цены биткоинов на биржах оказывается зависима от арбитражёров и спекулянтов, которые через них осуществляют торговлю. В поддержку этой идеи мы наглядно продемонстрировали, что торговля схожими валютными парами имеет более высокий кросс-обмен.

Наши оценки незаконных транзакций намного меньше, чем указывается в других источниках (например, [**1**]). Одна из причин такого различия заключается в том, что мы располагаем гораздо более подробным и исчерпывающе полным документом идентификации участников блокчейна. В указанной же работе авторы опираются на условно вычисляемую сеть, в которой любой биткоин-адрес рекурсивно классифицируется как принадлежащий незаконному объекту, если большинство его транзакций осуществляется с адресами, ранее признанными незаконными. Однако этот метод приводит к значительному завышению незаконных объёмов, поскольку не идентифицирует различия между реальными пользователями и ложными.

Мы доказали, что сконцентрированность собственных значений может служить новым и полезным показателем для ранжирования объёма и важности обменов, поскольку он основан на кросс-бирже потоков биткоинов в блокчейне и поэтому, вероятно, будет более устойчивым к манипуляции, чем другие меры.

Сильная взаимосвязанность бирж имеет важные последствия для обеспечения прозрачности и отслеживаемости транзакций, а в особенности для обеспечения соблюдения норм «знай своего клиента» (KnowYour-Customer, KYC) во всей сети. В настоящее время акцент в регуляторных усилиях по обеспечению большей прозрачности достигается посредством соблюдения этих норм и налоговой отчётности о приросте капитала на уровне отдельных учреждений, таких как биржи или платёжные системы. Однако, если пользователи биткоинов смогут свободно торговать регулируемыми и нерегулируемыми биржами даже со странами с разными уровнями правоприменения, эффективное регулирование системой KYC может оказаться невозможным на уровне отдельных институтов.

Для изучения потоков на рынке биткоинов мы использовали в качестве примера Hydra Market, которая является одной из крупнейших торговых площадок в даркнете. Наш анализ показал, что наибольшее количество организаций, напрямую взаимодействующих с пользователями Hydra Market, в том числе Binance и Huobi – две крупнейшие биржи в мире, не являются биржами, соблюдающими KYC. Как только финансовые потоки поступают на эти биржи, они смешиваются с другими и становятся виртуально не отслеживаемыми, а потому могут быть отправлены впоследствии куда угодно, даже на биржи, которые применяют нормы KYC. Напротив, прямое взаимодействие бирж, соблюдающих KYC, таких как Coinbase или Gemini, с пользователями Hydra Market весьма скромно. Но их косвенное взаимодействие с потоками рынка Hydra Market значительно больше, так как эти потоки направляются через сеть недолговечных кластеров, созданных исключительно с целью запутывания происхождения этих средств.

Эти результаты показывают, что организации, не поддерживающие KYC, служат шлюзом для «отмывания» денег и других «серых» действий. Децентрализованный характер протокола Биткоин упрощает работу этих учреждений – им нужно только иметь свои серверы в стране, в которой власти готовы мириться с их существованием. Если компаниям, выполняющим нормы KYC, разрешено принимать потоки от организаций, которые не соблюдают строгие нормы KYC (а это именно так на текущий момент), то цифровой след имеет очень ограниченное влияние на предотвращение поступления в широкое обращение сомнительных потоков.

Даже если организациям, выполняющим нормы KYC, будет разрешено иметь дело исключительно с организациями, также выполняющими нормы KYC, предотвратить приток испорченных средств будет почти невозможно в случае, если вы не готовы к строгим ограничениям на совершение сделок и установлению контроля над совершениями всех транзакций (похожие схемы работы предоставляют такие компании, как Bitfury Crystal Blockchain или Chainalysis). Потенциальная реализация такого режима предполагает, что объектом мониторинга блокчейна станут де-факто доверенные стороны, необходимые для функционирования сети Биткоин.

Состав биткоин-майнеров. На втором этапе анализа мы исследовали концентрацию и региональный состав биткоин-майнеров, которые несут ответственность за обработку и проверку транзакций сети Биткоин и поддерживают её целостность, за что получают оплату в виде вновь созданных биткоинов.

Проверка протокола сети Биткоин позволила установить, что её функционирование требует честности работы децентрализованных майнеров и безупречности ведения учёта бухгалтерской книги биткоинов. В случае нарушения механизма управления системой финансовая стабильность и безопасность сбережений могут оказаться под угрозой.

Поэтому важно понимать, насколько сконцентрированы майнинговые мощности. В ранее опубликованных работах основное внимание уделялось концентрации майнинговых пулов. По дизайну системы вероятность получения блока и получения вознаграждения за блок в цепочке биткоинов пропорциональна мощности хеширования, затрачиваемого на добычу. Это даёт майнерам возможность объединения своих вычис-

лительных мощностей и совместного страхования друг друга. Как следствие, в майнинге блокчейна Биткоин преобладают майнинговые пулы.

Но в то время как пулы функционируют как агрегаторы хеш-мощности и, следовательно, могут оказывать существенное влияние на протокол Биткоин, они не обязательно контролируют своих майнеров. Как сообщает [**2**], власть оператора пула по отношению к майнеру зависит от лёгкости, с которой майнеры могут перемещать мощность между пулами, что, в свою очередь, зависит от основного распределения майнеров по размеру.

В отличие от общедоступности информации о пулах для майнинга, информации об отдельных майнерах пока нет. Их идентификация осуществляется путём отслеживания распределения вознаграждений за майнинг от 20 крупнейших майнинговых пулов до майнеров, которые на них работают. Поскольку каждый пул использует свой алгоритм для распределения вознаграждений, мы создали отдельные алгоритмы для каждого пула. Насколько нам известно, это первое исследование, в котором установлена точная связь майнеров с их пулами для майнинга.

Наиболее крупные 10% майнеров контролируют 90% майнинговых мощностей и всего 0,1% (около 50 майнеров) – примерно 50% майнинговых мощностей. Кроме того, эта концентрация мощности майнинга является контрцикличной и зависит от цены биткоинов. Она уменьшается после резкого роста цены биткоинов и увеличивается в периоды ценопадения. Таким образом, риски атаки на блокчейн (атаки 51%) увеличиваются при резком падении цены биткоина.

Кроме того, мы доказали, что существует значительная географическая кластеризация майнеров. Как показал наш анализ, подавляющее большинство майнинговых пулов зарегистрировано в Китае (от 60% до 80% по данным на 2015 г. – апрель 2020 г.). Однако это совершенно не означает, что майнеры должны находиться там же. Основные данные об их местонахождении получены путём анализа IP-адресов из нескольких избранных пулов. В тот момент, когда майнер подключается к серверу пула, оператор может видеть IP-адрес майнера (если майнер не использует VPN-адрес). Оператор пула может использовать этот IP-адрес для определения географического положения. Имея возможность отслеживания адресов майнеров и транзакции биткоинов, мы можем видеть, на каких биржах они обналичивают свои вознаграждения. Как правило,

майнеры, находящиеся в определённом регионе, отправляют свои сбережения на биржу, также расположенную в этом регионе.

Чтобы проверить правильность нашего подхода к определению местоположения майнеров с помощью отслеживания того места, где они обналичивают свои биткоин-вознаграждения, мы использовали статистику за апрель 2021 г. по китайской провинции Синьцзян. После разрушительной аварии на угольной шахте правительство остановило добычу угля и отключило электроснабжение всего района почти на двое суток. Отключение электричества повлекло за собой массовое обналичивание биткоинов в регионе. Это помогло нам установить, что многие китайские майнеры располагались именно здесь, вероятно, из-за дешёвых поставок угля, за счёт которых работают электростанции.

Концентрация собственности. Наконец, мы изучили владельцев биткоинов. С момента появления сети Биткоин наблюдается большой интерес к личности крупнейших владельцев биткоинов и размеру их капитала. Существуют веб-сайты, посвящённые отслеживанию их адресов, на которых публикуется так называемый «богатый список», один из самых известных и популярных списков в криптосообществе. Но вопрос концентрации собственности – это не только любопытство. С точки зрения государственной политики, важно, кто получит наибольшую выгоду от повышения цен, которое произойдёт, если регулирующие органы разрешат более широкое внедрение биткоинов, – несколько избранных инвесторов или широкая публика?

Определение концентрации собственности сложнее, чем просто отслеживание авуаров самых богатых адресов, поскольку многие из них принадлежат «холодным кошелькам» бирж и онлайн-кошелькам, в которых хранятся биткоины инвесторов. На основе анализа граф мы разработали набор алгоритмов для классификации адресов, принадлежащих индивидуальным инвесторам или посредникам.

Полученные нами данные свидетельствуют о том, что остаток на счетах у посредников с 2014 г. неуклонно растёт. К концу 2020 г. он составил 5,5 млн биткоинов, примерно треть из них находится в обращении; 8,5 млн биткоинов контролируется индивидуальными инвесторами; высока концентрация биткоинов в холдингах: 1 тыс. крупнейших инвесторов контролируют около 3 млн, а 10 тыс. – около 5 млн.

Следующий этап анализа имел своей целью документирование эволюции объёма вложений различных участников блокчейна. В частности, мы разработали алгоритмы для отделения ложного объёма биткоинов от реального, а затем отобразили сетевую структуру участников. Далее мы осуществили анализ майнеров, их состав и географическую концентрацию, а также зафиксировали концентрацию собственности участников биткоинов.

Данные сети Биткоин. Все транзакции с биткоинами регистрируются в распределённой публичной книге (гроссбухе). Организованные в блоки, они добавляются в реестр в среднем каждые 10 минут. Один блок содержит несколько тысяч биткоин-транзакций, включающих в себя список отправителей и получателей, представленных адресами – псевдонимами, количество отправленных и полученных биткоинов, а также отметку о времени сделки.

Мы осуществили загрузку данных блокчейна, используя программное обеспечение с открытым исходным кодом Bitcoin Core [**3**] и программу BlockSci [**4**] для разбиения необработанных данных на отдельные транзакции. По статистике на 28 июня 2021 г. было создано 689 тыс. блоков из 652 млн транзакций биткоинов и 896 млн адресов, организованных в базе данных блокчейна объёмом более 379 Гб.

Адрес в блокчейне можно рассматривать как банковский счёт. Любой человек может отправлять биткоины на любой адрес. Но чтобы отправить биткоины с заданного адреса, нужно знать пароль, связанный с этим адресом. В отличие от банковских счетов биткоин-адреса могут быть сгенерированы свободно, поэтому обычно один и тот же владелец управляет несколькими адресами, а в некоторых случаях даже десятками миллионов различных адресов. Сообщество Биткоин разработало несколько эвристических алгоритмов для назначения адресов одним и тем же организациям. В качестве отправной точки мы используем наиболее консервативный метод кластеризации адресов, при этом все адреса, которые отправляют биткоины в любой отдельной транзакции, считаются принадлежащими к одному и тому же учреждению либо лицу. Эти алгоритмы оправданы протоколом Биткоин, который требует, чтобы подписывающая транзакцию сторона имела контроль над всеми выходными адресами.

Отметим, что некоторые сервисы, например CoinJoin, позволяют пользователям смешивать биткоины с монетами других пользователей, по этой причине программа BlockSci избегает транзакций CoinJoin в своём алгоритме кластеризации.

На практике пользователю обычно достаточно указать адрес назначения перевода и его сумму. Специальная программа, называемая кошельком, сама решает, с каких адресов отправлять биткоины, эквивалентные той сумме, которую пользователь хочет перевести. Затем этот процесс позволяет алгоритму кластеризации успешно сгруппировать все адреса пользователей вместе. Однако следует подчеркнуть, что пользователь может намеренно скрыть связи между своими адресами; в этом случае кластеризация определяет только нижнюю границу количества отдельных сущностей (людей, организаций).

Чтобы связать кластеры адресов с реальными объектами, мы использовали блоги и веб-сайты с криптовалютой, такие как Reddit, Blockchain.info, bitinfocharts.com, bitcointalk.org, walletexplorer.com, Matbea.com, для всех общедоступных адресов известных компаний Биткоин (бирж, платёжных систем, сайтов азартных игр и др.). Мы дополнили эту информацию современной базой данных криптосущностей от Bitfury Crystal Blockchain – одного из ведущих поставщиков инструментов для борьбы с отмыванием денег и аналитических решений в криптопространстве. Материалом для нашего исследования послужили 1 043 сущности, из них 393 биржи, 86 игорных сайтов, 39 онлайн-кошельков, 33 платёжных процессора, 63 пула для майнинга, 35 мошенников, 227 злоумышленников, 151 нелегальный рынок или сервис.

Объём транзакций блокчейна. Ложный объём. Дизайн блокчейна Биткоин и анонимность многих его пользователей создают массу ложного объёма транзакций, который не привязан к экономически значимым сделкам. В этом разделе мы постарались описать, как мы идентифицируем и отделяем этот объём от реального, то есть платежей за товары и услуги и другие финансовые переводы между двумя сторонами.

Поучительно и полезно начать с рассмотрения конкретного примера транзакции[2], изображённой на рис. 1 «Транзакции

---

[2] Вторую транзакцию в блоке 600000 можно увидеть, например, по адресу https://explorer.btc.com/btc/block/600000.

биткоинов и ложный объём» (оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf). Это типичная транзакция блокчейна биткоинов с большим ложным объёмом. С адреса отправителя «17A16Q...» в левой части гроссбуха отправлены средства по трём адресам. Адрес последнего из трёх получателей, которому отправляется наибольшая сумма, совпадает с адресом отправителя.

В приведённой в пример транзакции участник с адресом «17A16Q...» отправляет валюту со своего баланса по следующим трём адресам: «3QKAn2...», «1F8fDp...» и «17A16Q...». Сумма полученных средств равна отправленной сумме за вычетом небольшой комиссии в размере 0,001 биткоина, которая составляет вознаграждение за блок.

Стоит отметить, что последний из трёх адресов совпадает с адресом отправителя, то есть адрес «17A16Q...» отправляет большую часть своего баланса сам себе. Это означает, что общий объём данной транзакции в блокчейне достаточно большой, однако экономически значимый (реальный) объём, генерируемый в результате транзакции, то есть обмен между разными объектами, невелик.

Вышеупомянутая ситуация, когда участник отправляет свою сумму самому себе или на другой адрес, контролируемый одним и тем же лицом, очень распространён. Отчасти это следствие конструкции протокола Биткоин. Непогашенный остаток по адресу не сохраняется, но рассчитывается из всей истории транзакций, связанных с этим адресом, и будет виден при просмотре реестра биткоинов. Для вычислительной эффективности биткоин-протокол позволяет отправлять только те суммы, которые были ранее получены данным адресом. Например, предположим, что на адрес ранее были получены 5, 7 и 10 биткоинов, поэтому непогашенный остаток составляет 22 биткоина. Чтобы отправить 8 биткоинов с этого адреса, можно либо отправить 10 биткоинов, либо любую из следующих линейных комбинаций: 5 + 7, 5 + 10, 7 + 10, 5 + 7 + 10. Так как в любом случае сумма превышает 8 биткоинов, отправителю необходимо собрать разницу, используя один из своих адресов. Этот процесс создаёт большое количество ложного объёма, который скрывает истинный объём транзакций в цепочке блоков.

Ещё одна распространённая причина возникновения ложного объёма – это анонимность участников блокчейна. Поскольку блокчейн

биткоинов является публичной книгой, все потоки платежей между адресами прекрасно наблюдаются. Поэтому многие пользователи выбирают стратегии, предназначенные для предотвращения отслеживания потоков биткоинов.

Рассмотрим, например, ситуацию, когда хакер требует, чтобы платёж от компании был отправлен на адрес биткоинов, который он контролирует. Поскольку адрес выкупа является общедоступной информацией, если хакер позже отправит биткоины с этого адреса третьему лицу, сторона может легко заявить, что средства поступают в результате незаконной деятельности. Чтобы этого не произошло, хакеры часто пытаются скрыть трассировку, создавая несколько адресов и разделяя первоначальный взнос между ними. Этот процесс обычно повторяется много раз, в результате чего возникают так называемые «цепочки отслаивания», когда средства перемещаются на большое расстояние от одного адреса к другому, что приводит к появлению большого количества записей фиктивных адресов в бухгалтерской книге.

Цепочки отслаивания также широко используются многими биржами, такими как Coinbase и Kraken, и многими пулами майнинга. Эти учреждения каждый раз, когда им нужно создать изменения (как в транзакции на рис. 1), генерируют новый адрес вместо повторного использования старого адреса. Затем этот новый адрес применяется для отправки средств другому лицу, и изменения собираются по другому новому адресу. Процесс обычно повторяется много раз, пока не будет израсходован весь первоначальный баланс. Адреса в цепочках отслаивания обычно используются только для мгновенного получения и немедленной отправки биткоинов с типичным сроком службы 10 часов.

Существует два способа учёта транзакций цепочки отслаивания. Первый способ предполагает возможность изменения алгоритма кластеризации, чтобы добавить адреса в цепочках отслаивания к соответствующим кластерам. Другой подход, которому мы следуем в этой статье, – это возврат переходов в цепочках отслаивания до исходных кластеров и отбрасывание всех любых промежуточных адресов из дальнейшего анализа. Для этого мы разработали эффективный рекурсивный алгоритм, подробно описанный ниже.

Выделение цепочек отслаивания снижает вычислительную нагрузку и приводит к значительному сокращению адресов и кластеров.

Исходная база данных насчитывает 896 млн адресов, однако если удалить адреса в цепочках отслаивания, получится 640 млн. Эти адреса принадлежат 189 млн кластеров, из которых 116 млн являются одноадресными.

На рис. 2 «Разделение объёма на компоненты: внутренний, сквозной и реальный объём» (оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) показано разложение общего объёма блокчейна Биткоин на то, что мы называем внутренним/сквозным/реальным объёмом. Внутренний объём – это объём, который создаётся, когда кластер отправляет биткоины самому себе. Сквозной объём – это переходный объём, связанный с отслаивающимися цепями. Наконец, реальный объём – это оставшийся объём, который представляет собой переводы между кластерами. Он составляет только 10% от общего объёма биткоинов на блокчейне; 90% объёма не привязаны к экономически значимым сделкам.

Верхняя (оранжевая) часть на рисунке показывает сквозной объём, который создаётся, когда пользователи перемещают свои средства по длинным цепочкам из нескольких адресов (так называемым отшелушивающим цепочкам) и делят между ними платежи, чтобы препятствовать отслеживанию потоков. Следующая часть (жёлтая) показывает внутренний объём, который генерируется, когда пользователь (кластер) отправляет сам себе биткоины. Наконец, оставшаяся часть (зеленая) – настоящий (истинный) объём, который представляет собой переводы между кластерами, контролируемыми разными пользователями.

Публикация результатов масштабного анализа сети Биткоин будет продолжена в следующих номерах журнала.

Перевод А. И. Земскова, ГПНТБ России
(Продолжение в следующих номерах журнала.)

Список источников

1. **Foley S., Karlsen J., Putniņš T.** Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? // The Review of Financial Studies. 2019. № 32 (5). P. 1798–1853. https://doi:10.1093/rfs/hhz015

2. **Cong Y, Ulasli M, Schepers H. et al.** Nucleocapsid Protein Recruitment to Replication-Transcription Complexes Plays a Crucial Role in Coronaviral Life Cycle //J Virol. P. 169–176. 2020. № 94 (4). doi: 10.1128/JVI.01925-19

3. **Bitcoin** Core. URL: https://bitcoin.org/en/bitcoin-core/ (дата обращения: 05.08.2022).

4. **BlockSci.** URL: https://github.com/citp/BlockSci (дата обращения: 05.08.2022).

5. **Ron D., Shamir A.** Quantitative Analysis of the Full Bitcoin Transaction Graph. URL: https://eprint.iacr.org/2012/584.pdf (дата обращения: 05.08.2022).

6. **Meiklejohn C., Holmbeck M., Siddiq M. et al.** An Incompatibility between a Mitochondrial tRNA and Its Nuclear-Encoded tRNA Synthetase Compromises Development and Fitness in Drosophila // PLOS Genetics. № 9 (1). doi: 10.1371/journal.pgen.1003238

Информация об авторах

Игорь Макаров – Лондонская школа экономики Houghton Street Лондон WC2A 2AE Соединённое Королевство
i.makarov@lse.ac.uk

Антуанетта Шоар – Школа менеджмента MIT Sloan School of management 100 Main Street, E62-638 Кембридж, Массачусетс 02142 и NBER
aschoar@mit.edu

UDC 004.65:[336.743-021.131]-047.44
https://doi.org/10.33186/1027-3689-2022-9-98-111

Blockchain analysis of the Bitcoin market. (Part 1)

Igor Makarov[1], Antoinette Schoar[2]

[1]London School of Economics Houghton Street London WC2A 2AE UK, i.makarov@lse.ac
[2]MIT Sloan School of Management 100 Main Street, E62-638 Cambridge, MA 02142 and NBER, aschoar@mit.edu

Abstract. The detailed analysis of the Bitcoin network and its main participants. The expert authors (Igor Makarov, London School of Economics, Antoinette Schoar, MIT Sloan School of Management) completed the study authorized by the National Bureau of Economic Research (NBER), the US-based private agency. The Bitcoin network is defined as a new database comprising many of public and proprietary sources to link bitcoin address to real object, and an extensive set of algorithms to extract information on market key players behavior. Three major pieces of analysis of the Bitcoin eco-system were conducted. First, the authors analyze the transaction volume and network structure of the main participants on the blockchain. Second, they document the concentration and regional composition of the miners which are the backbone of the verification protocol and ensure the integrity of the blockchain ledger. Finally, they analyze the ownership concentration of the largest holders of Bitcoin. The researchers found that 1/3 of all bitcoins issued were owned by 10,000 individual investors. They conclude that the high concentration makes the first cryptocurrency market vulnerable to hypothetical hacker attack. The translator notes that paraphrasing English text in Russian was rather challenging due to the newness of the financial agenda and introduction of the term *entity* extensively used in the Western countries though new to Russia. Nevertheless, it is necessary to introduce readers to the bitcoin technology which will be also practical and useful for the library and information community.

Keywords: cryptocurrency, bitcoin, blockchain, transaction, miner, multiple ownership

Cite: Makarov I., Shoar A. Blockchain analysis of the Bitcoin market. (Part 1) / I. Makarov, A. Shoar // Scientific and technical libraries. 2022. No. 9. P. 98–111. https://doi.org/10.33186/1027-3689-2022-9-98-111

Introduction

Cryptocurrencies have seen a remarkable growth in value and public attention since their inception more than a decade ago. Opinions about the impact of cryptocurrencies range all the way from being a revolution in financial access to a threat to financial stability and monetary policy. A distinguishing feature of cryptocurrencies is the promise of a decentralized system of payments or store of value outside the traditional nexus of government scrutiny. The blockchain technology at the heart of cryptocurrencies replaces the reliance on a few centralized record keepers, such as banks or credit card networks, with a large set of decentralized and anonymous agents. The absence of centralized accountability and the anonymity of its users are often viewed as major benefits by crypto supporters, but it hinders the timely diagnosis of the health of the system, generates many challenges for regulators, and introduces new sources of systematic risk.

Bitcoin, the original cryptocurrency, is still the largest and most popular coin, with a market cap that is larger than all the other coins combined. It is often seen as a template or point of comparison for other new coins. Many industry participants are now calling for even wider Bitcoin adoption, either as a public investment vehicle or legal tender. These pressures put regulators who want to find the right balance between protecting the public interest and allowing innovation in a difficult position. There are still many open questions about the utilization of bitcoin, its ownership concentration as well as the structure of core entities that form the backbone of the Bitcoin ecosystem, despite being in existence for more than ten years. A better understanding of the Bitcoin network and its participants is required for any decision about how and whether to integrate Bitcoin into the traditional financial system.

In this paper, we aim to shed light on these open questions by developing a novel database that allows us to document the evolution of the Bitcoin market and its different participants over time. To build this database we use a large number of public and proprietary sources that link Bitcoin addresses to real entities and develop a suite of algorithms that

use the semi-public nature of the Bitcoin blockchain to extract information about the behavior of the main market participants. We believe that this is the most complete Bitcoin database used in academic research to date.

We conduct three major pieces of analysis that focus on the main participants of the blockchain eco-system. First, we analyze the transaction volume and network structure of the main participants on the Bitcoin blockchain. Second, we document the concentration and regional composition of miners which ensure the integrity of the blockchain ledger. Finally, we analyze the ownership concentration of the largest holders of Bitcoin.

Transaction Volume and Network Structure. We first document that 90% of transaction volume on the Bitcoin blockchain is not tied to economically meaningful activities but is the byproduct of the Bitcoin protocol design as well as the preference of many participants for anonymity. Because the Bitcoin blockchain is a public ledger all payment flows between addresses are perfectly observable. Therefore, many bitcoin users adopt strategies designed to impede the tracing of bitcoin flows by moving their funds over long chains of multiple addresses and splitting payments among them resulting in a large amount of spurious volume. We develop algorithms to filter out this spurious volume and trace economically meaningful payments between real entities on the Bitcoin network.

We show that the vast majority of Bitcoin transactions between real entities are for trading and speculative purposes. Starting from 2015, 75% of real bitcoin volume has been linked to exchanges or exchange-like entities such as on-line wallets, OTC desks, and large institutional traders. In contrast, other known entities are only responsible for a minor part of total volume. For example, illegal transactions, scams and gambling together make up less than 3% of volume[1]. The fraction of volume explained by miners is even smaller.

---

[1] Our estimates of illegal transactions are much smaller than the previous literature found, see for example Foley et al. (2019). One reason for this difference is that we have a much more detailed and comprehensive identification of participants on the blockchain. The prior work had to rely on an imputed network of illegal entities where any Bitcoin address recursively is classified as belonging to an illegal entity if the majority of its transactions is with addresses that themselves were previously classified as illegal. However, this method leads to significant overstatement of illegal volume, since it does not discriminate between real users and spurious volume.

Exchanges not only generate the most volume, but they are also the most connected nodes in the Bitcoin network. In particular, they have the highest measure of eigenvalue centrality[2]. Furthermore, a large fraction of exchange volume consists of cross-exchange flows. The high cross-exchange flows are the consequence of the current market structure. Different from traditional, regulated exchanges, cryptocurrency markets consist of many non-integrated and independent exchanges without any provisions to ensure that investors receive the best price when executing trades. As a result, the consistency of the Bitcoin price across exchanges depends on arbitrageurs and speculators who trade across them. In support of this idea, we show that exchanges that trade similar currency pairs have higher cross-exchange flows.

The strong interconnectedness of exchanges has important implications for the transparency and traceability of transactions, and especially the enforcement of KnowYour-Customer (KYC) norms, across the network. The current regulatory efforts focus on creating greater transparency through enforcement of KYC norms and capital gains tax reporting at the level of individual institutions, such as exchanges or payment processors. However, if users of Bitcoin can freely trade across regulated and unregulated exchanges or even countries with different enforcement levels, effective KYC regulation might not be possible at the level of individual institutions.

We use the example of Hydra Market, which is one of the largest dark net marketplaces, to study flows in this market. Our analysis shows that the highest volume entities interacting directly with Hydra Market users are non-KYC exchanges, including Binance and Huobi which are two of the largest exchanges worldwide. Once the flows arrive at these exchanges, they get mixed with other flows and become virtually untraceable, and so can be sent anywhere afterwards, even to exchanges that enforce KYC norms. In contrast, the direct interaction of KYC exchanges,

---

[2] See Section 3.3 for the definition and details. We show that the eigenvalue centrality can serve as a new and useful measure for ranking the volume and importance of exchanges because it is based on the cross-exchange Bitcoin flows on the blockchain, and therefore, is likely to be more resilient to manipulation than other measures.

such as Coinbase or Gemini, with Hydra Market users is modest. But their indirect interaction with flows originating from Hydra market is significantly larger, since these flows are channeled through a network of short-lived clusters, solely created for the purpose of obfuscating the origin of these funds.

These results highlight that non-KYC entities serve as a gateway for money laundering and other gray activities. The decentralized nature of the Bitcoin protocol makes it easy for these entities to operate – they only need to have their servers in a country where the authorities are willing to tolerate their existence. If KYC entities are allowed to accept flows from entities that are not following strict KYC norms (the current state), then the digital footprint has a very limited effect on preventing tainted flows from entering into wide circulation.

Even if KYC entities were restricted to deal exclusively with other KYC entities, preventing inflows of tainted funds would still be nearly impossible, unless one was willing to put severe restrictions on who can transact with whom and make every transaction subject to the approval of a blockchain "monitoring entity", e. g. similar to what companies like Bitfury Crystal Blockchain[3] or Chainalysis are providing. Note that if this regime was to realize, the blockchain monitoring entities would become de facto trusted parties essential for the functioning of the Bitcoin network. But this is exactly what the Bitcoin protocol aims to overcome.

Composition of Bitcoin Miners. In a second major piece of analysis, we study the concentration and regional composition of Bitcoin miners, which are responsible for processing and verifying Bitcoin transactions and maintaining the integrity of the Bitcoin blockchain. For this service, miners are rewarded with newly created Bitcoins and transaction fees.

A proof of work protocol like Bitcoin requires a majority of decentralized miners to be honest for its record keeping function to work. If a single miner or a set of colluding miners were to command a majority of the mining power in the network, the ledger could become controlled by the colluding group and result in the infamous 51% attack, in which the group can alter the previously verified records. The possibility of such attacks creates systemic risks for financial stability and potentially even

---

[3] https://crystalblockchain.com/

for national security if a large fraction of citizens uses Bitcoin as a store of value.

It is therefore important to understand how concentrated the mining capacity is. The previous literature has mainly focused on mining pool concentration. By design, the probability of mining a block and obtaining a block reward in the Bitcoin blockchain is proportional to the hashing power spent on mining. This provides strong incentives for miners to pool their computing power and co-insure each other. As a consequence, mining in the Bitcoin blockchain is dominated by mining pools.

But while pools function like aggregators of hashing capacity and can therefore have substantial influence over the Bitcoin protocol, they do not necessarily control their miners. As Cong et al. (2020a) emphasize, the power that a pool operator has vis a vis the miner depends on the ease with which miners can shift capacity across pools, which in turn depends on the underlying size distribution of the miners.

Unlike information about mining pools, which is commonly available, information about individual miners is not readily available. We identify individual miners by tracking the distribution of mining rewards from the largest 20 mining pools to the miners that work for them. Since each pool uses its own algorithm to distribute rewards, we build separate algorithms for each pool. To the best of our knowledge, this is the first study that accurately links miners to their mining pools.

We show that the Bitcoin mining capacity is highly concentrated and has been for the last five years. The top 10% of miners control 90% and just 0.1% (about 50 miners) control close to 50% of mining capacity. Furthermore, this concentration of mining capacity is counter cyclical and varies with the Bitcoin price. It decreases following sharp increases in the Bitcoin price and increases in periods when the price drops or. Thus, the risk of a 51% attack increases in times when the Bitcoin price drops precipitously or following the halving events.

In addition, we show that there is significant geographic clustering of miners. While it has been previously discussed that a large majority of mining pools are registered in China, this does not automatically mean that miners have to be located in China. So far, the main data about miners' location has come from the analysis of miners' IP 4 addresses from a

few select pools. When a miner connects to a pool server, the pool operator can see the IP address of the miner. Unless a miner uses a VPN address, the pool operator can use this IP address to determine the geographical location.

Here, we utilize a new approach, which takes advantage of our ability to trace miners on the blockchain. Since we can trace miners' addresses and Bitcoin transactions, we can see at which exchanges they use to cash out their rewards. The idea is that miners in a particular region would most likely send their rewards to an exchange that is also in this region. Using our approach we show that starting in 2015 and until April 2020 a majority of mining capacity, between 60% to 80% is located in China, which is consistent with anecdotal evidence.

In order to verify the validity of our approach of identifying miner locations by looking at where miners cash out their Bitcoin rewards, we use a recent incidence in April 2021 in the Xinjiang province of China. After a devastating coal mining accident, the government shut down coal mining and electricity supply for the entire area. Many Chinese Bitcoin miners are located in this province due to the cheap supply of coal powered electricity. Of course, not all Chinese miners are located in this area and thus we do not use it as a test of the mining capacity in China. But the shutdown of electricity for more than two days allows us to identify a set of miners for which we can be sure that they are physically located in China since they had to stop their operations. Using this strategy, we confirm that these Chinese miners, indeed utilize the cashing out policies that we had conjectured.

Ownership concentration. Finally, we study the ownership and concentration of Bitcoin holdings. Since the inception of Bitcoin, there has been intense interest in the question of who are the largest owners of Bitcoin, and how much do they actually own. There are websites dedicated to tracking the addresses with the largest Bitcoin holdings, the so called "rich list," one of the most well-known and widely followed lists in the crypto community. But the question of ownership concentration is not only a matter of curiosity and intrigue. From a public policy perspective, it is important to understand who is positioned to benefit most from any price appreciation that would happen if regulators allow a broader adoption of Bitcoin. Are these a select few investors or the general public?

Determining the concentration of ownership is more complicated than just tracking the holdings of the richest addresses, since many of the largest addresses belong to cold wallets of exchanges and online wallets, which hold Bitcoin on behalf of many investors. We develop a suite of algorithms based on graph analysis to classify addresses into those belonging to individual investors or those belonging to intermediaries[4].

We show that the balances held at intermediaries have been steadily increasing since 2014. By the end of 2020 it is equal to 5.5 million bitcoins, roughly one-third of Bitcoin in circulation. In contrast, individual investors collectively control 8.5 million bitcoins by the end of 2020. The individual holdings are still highly concentrated: the top 1000 investors control about 3 million BTC and the top 10,000 investors own around 5 million bitcoins.

The rest of the paper is structured to first discuss the data sources and the construction of the data set. The next section documents the evolution of volume to different participants on the blockchain, in particular, we develop algorithms to separate spurious volume from real volume and then map the network structure of participants. In the following section we analyze miners, their composition and geographic concentration. And finally we document the ownership concentration of Bitcoin participants.

1. Related Literature

Our paper contributes to a fast-growing literature on cryptocurrencies and blockchains. Raskin and Yermack (2016) and Hardle et al. (2020) provide a broad perspective on the economics of cryptocurrencies and the blockchain technology they are built upon.

Budish (2018), Abadi and Brunnermeier (2018), and Biais et al. (2019) study consensus mechanisms and limitations of the proof-of-work protocol, the core innovation of this new technology.

Athey et al. (2016), Cong et al. (2020b), Pagnotta and Buraschi (2018), Sockin and Xiong (2020), and Han and Makarov (2021) develop different theoretical frameworks to study bitcoin adoption and bitcoin pricing and highlight that beliefs about adoption are central for Bitcoin

[4] See Section 5 for a detailed description of the identification.

pricing. Schilling and Uhlig (2019) propose a model, in which a cryptocurrency such as Bitcoin coexists and competes with a traditional governmentissued fiat money.

A number of papers study the economics of Bitcoin mining. Prat and Walter (2021) examines the relationship between the Bitcoin price and the investment in hashing capacity. Easley et al. (2019) and Huberman et al. (2021) develop equilibrium models of Bitcoin mining fees. Cong et al. (2020a) propose a theory of mining pools and suggest that mining pools escalate miners' arms race and significantly increase the energy consumption of proof-of-work-based blockchains. Ferreira et al. (2019) model the joint behavior of miners, mining pools, and firms producing specialized mining equipment. We contribute to this literature by developing a suite of algorithms to identify individual miners on the blockchain. This data is the first to trace individual miners and allows us to study their concentration and regional composition.

Similar to our paper, Foley et al. (2019) use the Bitcoin blockchain data to examine the prevalence of illegal transactions on the Bitcoin blockchain. Wallet-level blockchain data are also used by Griffin and Shams (2020) to study whether tether issuance affects bitcoin prices. In comparison to the earlier literature, we develop a novel database that not only has a much more comprehensive classification of participants on the blockchain, but also eliminates spurious volume. This granular data allows us to attribute economically meaningful transactions more precisely and to provide a detailed analysis of the evolution of the Bitcoin market.

2. Data

All bitcoin transactions are recorded on a distributed public ledger, the so-called blockchain. Transactions are organized in blocks that are added to the ledger every 10 minutes on average. Each block contains a few thousand transactions. A typical Bitcoin transaction includes a list of senders and recipients represented by pseudonymous addresses, the number of bitcoins sent and received, and a time-stamp of the transaction.

We download the blockchain data using the open-source software of Bitcoin Core and use the BlockSci program to parse the raw data into in-

dividual transactions[5]. As of June 28, 2021, there have been 689,000 blocks of 652 million Bitcoin transactions and 896 million addresses organized in a blockchain database of more than 379 GB in size.

An address on the blockchain can be thought of as a bank account. Anyone can send bitcoins to any address. But to send bitcoins from a given address one needs to know a password associated with this address. Unlike bank accounts, Bitcoin addresses can be generated freely, so typically the same entity controls several addresses, and in some cases, even tens of millions of different addresses.

The Bitcoin community developed several heuristics to assign addresses to the same entity. As a starting point, we use the most conservative method to cluster addresses whereby all addresses that send bitcoins in any single transaction are deemed to belong to the same entity[6].

This heuristic is justified by the Bitcoin protocol that requires the party that signs a transaction to have control of all output addresses

In practice, a user typically only needs to specify the destination addresses and the amounts to be transferred. A special piece of software, called a wallet, then decides which addresses to send bitcoins from to cover a given amount that the user wants to transfer. This process then allows the clustering algorithm to successfully group all user's addresses together. It should be stressed however, that with a little bit of effort, a user can deliberately conceal the connections between his different addresses by making sure that no two addresses are ever used in the same transaction. As a result, this clustering heuristics only produces a lower bound for the true number of distinct entities.

To link address clusters to real entities we scrape cryptocurrency blogs and websites, such as Reddit, Blockchain.info, bitinfocharts.com, bitcointalk.org, walletexplorer.com, and Matbea.com for all publicly avai-

---

[5] Bitcoin Core and BlockSci are available at https://bitcoin.org/en/bitcoin-core/ and https://github.com/citp/BlockSci, respectively.

[6] See Ron and Shamir (2012) or Meiklejohn et al. (2013). Bitcoin mixing services, such as CoinJoin, let users mix their coins with other users, and are designed to confuse this heuristic. The BlockSci accounts for that and avoids CoinJoin transactions in its clustering algorithm. See BlockSci documentation for more details.

lable addresses of prominent Bitcoin entities such as exchanges, payment processors, gambling sites, and others. We supplement this information with the state-of-the-art database of crypto entities from Bitfury Crystal Blockchain. Bitfury Crystal Blockchain is one of the leading providers of anti-moneylaundering tools and analytic solutions in the crypto space.

To the best of our knowledge, we have the most complete information about crypto entities that have been used in academic research up to this point. Our data cover 1,043 different entities. These include 393 exchanges, 86 gambling sites, 39 on-line wallets, 33 payment processors, 63 mining pools, 35 scammers, 227 ransomware attackers, 151 dark net market places and illegal services.

3. Bitcoin Blockchain Volume

3.1. Spurious Volume

The design of the Bitcoin blockchain and the preference of many of its users for anonymity creates a lot of spurious volume that is not tied to economically meaningful transactions. In this section, we describe how we identify and separate this volume from the real volume, i. e. payments for goods and services and other financial transfers between two parties. It is instructive to start by looking at a particular example, see the transaction depicted in

Figure 1[7]. In this transaction, the address "17A16Q..." sends its balance to the following three addresses "3QKAn2...", "1F8fDp...", and "17A16Q...". The amount received is equal to the amount sent except for a small fee of 0.001 bitcoins, which is a part of the block reward. Notice that the last of the three addresses is the same as the sending address, that is, the address "17A16Q..." sends the majority of its balance to itself. This means the overall volume this transaction generates on the blockchain is large. However, the economically meaningful volume generated in the transaction (the real volume), which is the volume between different entities, is small.

The above situation where an address sends its balance to itself or to another address controlled by the same entity is very common. In part,

---

[7] This is the second transaction in block 600,000 and can be seen e. g., at https://explorer.btc.com/btc/block/600000.

it is a consequence of the design of the Bitcoin protocol. The outstanding balance of an address is not stored in the address but is imputed from the whole history of transactions involving this address by traversing back the Bitcoin ledger. For computational efficiency, the Bitcoin protocol allows one to send only the amounts that have been previously received by an address. For example, suppose an address previously received 5, 7, and 10 bitcoins, so the outstanding balance is 22 bitcoins.

To send 8 bitcoins from this address one can either send 10 bitcoins, or any of the following linear combinations: 5 + 7, 5 + 10, 7 + 10, 5 + 7 + 10. Since in any case, the amount is larger than 8 bitcoins the sender needs to collect the difference using one of his addresses. This process creates a large amount of spurious volume that obscures the true volume of transactions on the blockchain.

Another common reason for spurious volume is the preference of blockchain participants for anonymity. Because the bitcoin blockchain is a public ledger all payment flows between addresses are perfectly observable. Many Bitcoin users, therefore, adopt strategies designed to impede the tracing of bitcoin flows.

Consider, for example, a situation where a hacker demands payment from a company to be sent to a Bitcoin address he controls. Since the ransom address is public information, if the hacker later sends bitcoins from this address to a third party, the party could easily flag funds as coming from illegal activity. To prevent this from happening, hackers often try to obfuscate tracing by creating multiple addresses and splitting the initial payment among them. This process is usually repeated many times resulting in the so-called “peeling chains”, where funds travel a long distance from one address to another leading to a large amount of fictitious volume on the ledger.

Peeling chains are also commonly used by many exchanges, such as Coinbase and Kraken, and many mining pools. These entities, every time they need to collect a change as in the transaction in Figure 1: Bitcoin transactions and spurious volume (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf), generate a new address instead of re-using the old address. This new address is then used to send funds to another entity, and the change is col-

lected in another new address. This process is usually repeated many times until all initial balance is spent. The addresses used in peeling chains are usually used only to receive and immediately send bitcoins with a typical lifetime span of 10 hours.

There are two ways how one can account for peeling chain transactions. First, one could modify the clustering algorithm to add addresses in peeling chains to the corresponding clusters. The other approach, which we follow in this paper, is to backtrack volume in peeling chains to the original clusters and discard any intermediate addresses from further analysis. To backtrack this volume we develop an efficient recursive algorithm detailed in the Appendix.

Factoring out peeling chains reduces the computational burden and results in significant reduction of addresses and clusters. While the original database has 896 million addresses, after we remove addresses in peeling chains we end up with 640 million addresses. Theses addresses belong to 189 million clusters, of which 116 million clusters are single-address clusters.

Figure 2: Decomposition of volume: Internal, pass-through, and real volume (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) shows the decomposition of total Bitcoin blockchain volume into what we call internal, pass-through, and real volume. Internal volume is the within-cluster volume, that is, the volume that is generated when a cluster sends bitcoins to itself. The pass-through volume is the transitory volume associated with peeling chains. Finally, the real volume is the remaining volume, which represents transfers between clusters. This volume accounts only for 10% of the total Bitcoin volume on the blockchain, with 90% of the Bitcoin volume on the blockchain not tied to economically meaningful transactions.

References

1. **Foley S., Karlsen J., Putniņš T.** Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? // The Review of Financial Studies. 2019. № 32 (5). P. 1798–1853. https://doi:10.1093/rfs/hhz015

2. **Cong Y, Ulasli M, Schepers H. et al.** Nucleocapsid Protein Recruitment to Replication-Transcription Complexes Plays a Crucial Role in Coronaviral Life Cycle // J Virol. P. 169–176. 2020. № 94 (4). doi: 10.1128/JVI.01925-19

3. **Bitcoin** Core. URL: https://bitcoin.org/en/bitcoin-core/ (accessed: 05.08.2022).

4. **BlockSci.** URL: https://github.com/citp/BlockSci (accessed: 05.08.2022).

5. **Ron D., Shamir A.** Quantitative Analysis of the Full Bitcoin Transaction Graph. URL: https://eprint.iacr.org/2012/584.pdf (accessed: 05.08.2022).

6. **Meiklejohn C., Holmbeck M., Siddiq M. et al.** An Incompatibility between a Mitochondrial tRNA and Its Nuclear-Encoded tRNA Synthetase Compromises Development and Fitness in Drosophila // PLOS Genetics. № 9 (1). doi: 10.1371/journal.pgen.1003238

Information about the authors

Igor Makarov – London School of Economics Houghton Street London WC2A 2AE UK
i.makarov@lse.ac

Antoinette Schoar – MIT Sloan School of Management 100 Main Street, E62-638 Cambridge, MA 02142 and NBER
aschoar@mit.edu