ЭЛЕКТРОННЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

УДК 004.65:[336.743-021.131]-047.44
https://doi.org/10.33186/1027-3689-2022-10-85-100

Блокчейн-анализ рынка биткоинов. (Часть 2)

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 проведён на трёх основных этапах. 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 За-

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

паде, однако нового для нас термина *entity* (сущность). Несмотря на данный факт, представляется необходимым ознакомить читателей с технологией биткоинов, 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 2) / И. Макаров, А. Шоар // Научные и технические библиотеки. 2022. № 10. С. 85–100. https://doi.org/10.33186/1027-3689-2022-10-85-100

3.2. Реальный объём

Теперь рассмотрим экономически значимую, не ложную часть объёма транзакций биткоинов. Чтобы понять, для каких целей используется биткоин, мы отслеживаем его потоки между различными типами объектов в блокчейне. В список организаций входят биржи, онлайн-кошельки, платёжные системы, сайты азартных игр, сервисы микширования, нелегальные сервисы и пулы для майнинга. Мы идентифицируем эти организации из большого количества общедоступных и коммерческих источников, как описано в разделе 2. Данные.

Такие криптовалютные биржи, как Coinbase, Binance или Kraken, и онлайн-кошельки Blockchain.info и BixIn являются представителями основных типов организаций, в которых биткоин можно хранить, а также торговать им. Теоретически биржи должны предоставлять платформы для торговли биткоинами за фиатные валюты и другие монеты, а онлайн-кошельки – специализироваться на сервисах хранения. Однако на практике разница между ними зачастую невелика. В большинстве случаев оба типа объектов предлагают и ту, и иную функции. Поэтому в обзоре об использовании биткоинов мы сгруппировали эти объекты вместе. Платёжные системы, такие как BitPay или CoinPayments, облегчают оплату онлайн-магазинам, продавцам азартных игр и другим организациями, принимающим криптовалюту как средство оплаты товаров и услуг. В список организаций, оказывающих незаконные услуги, входят торговые площадки в «тёмной» сети, такие как Hydra Market, многочисленные кошельки с программами-

вымогателями и компании, занимающиеся мошенничеством. Сервисы смешивания или тумблеры, такие как Bitcoin Fog и кошелёк Wasabi, – это сайты, которые позволяют своим клиентам объединять средства, чтобы скрыть адрес их отправки.

Другой список учреждений и лиц из числа основных компонентов системы Биткоин – это майнинг-пулы и майнеры. Мы идентифицируем индивидуальных майнеров, отслеживая их вознаграждение в крупнейших майнинговых пулах (процедура отслеживания майнеров описана в разделе 4 и в приложении). Всего идентифицировано 248 тыс. майнеров.

Как обсуждалось ранее, псевдонимный характер биткоина затрудняет возможность связать адрес с реальным объектом, стоящим за ним. Таким образом, идентификация сущностей является неполной по смыслу, поскольку полагается на то, что субъект либо добровольно раскрывает свой адрес, либо информация об адресе юридического лица обнаруживается в ходе взаимодействия с ним.

Чтобы решить проблему неполной идентификации юридических лиц и организаций и убедиться, что мы не упустили из виду крупных игроков блокчейна, нами было проанализировано 10 тыс. крупнейших неизвестных кластеров с наибольшим объёмом неидентифицированных биткоинов. Из этого множества кластеров мы выбрали те, которые либо получают регулярные потоки от майнеров, либо получают более 50% своих поступлений от известных бирж (или отправляют более 40% на известные биржи). Эти пороговые уровни определены по шаблонам транзакций известных нам организаций. Для типичной биржи 53% оттока биткоинов (отправлений) идёт на другие биржи, а 52% притока биткоинов поступает от других бирж. По всем остальным организациям эти цифры значительно ниже. Например, типичный игорный сайт отправляет на биржи 21%, а получает от них 29% от общего потока. Мы обнаружили, что 4 507 кластеров удовлетворяют указанным выше условиям. Вместе они поставляют 63% биткоинов, поступающих в крупнейшие 10 тыс. кластеров. В дальнейшем мы будем ссылаться на такие кластеры, как биржи типа LEOTD (Likely Exchanges, OTC brokers, or Trading Desks – вероятные биржи, внебиржевые брокеры или торговые столы).

Основываясь на этой классификации участников, мы показываем среднемесячный объём транзакций, который создан этими учреждениями в цепочке блоков с начала 2015 г. по май 2021 г. (рис. 3, оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf). Объём рассчитывается как сумма биткоинов, отправленных организациям различных типов в определённый месяц. На рис. A объём биткоинов показан в миллионах (BTC), а на рис. B – в процентах от общего ежемесячного объёма.

Мы видим, что большая часть объёма генерируется транзакциями с участием бирж и кластеров типа LEOTD. Объём, поступающий на известные биржи, составляет около 40% от общего объёма, ещё 20% приходится на объём, поступающий от 11 организаций типа LEOTD. Чтобы подчеркнуть доминирующую роль бирж и LEOTD, мы разделили объём, попадающий в категорию Other (все неизвестные кластеры, не являющиеся LEOTD), на две части: поступающий с бирж, LEOTD и остальной. Это распределение показывает, что объём от бирж и LEOTD до Other занимает ещё 20% объёма. Таким образом, объём, связанный с биржей и торговыми системами, составляет около 80% от общего. Другим известным организациям по данным за 2020 г. принадлежит только незначительная часть от общего объёма. Например, на незаконные операции, мошенничество и азартные игры приходится менее 3% от общего объёма. Доля, связанная с майнерами, ещё меньше.

Этот анализ объёма отличает доминирование транзакций, связанных с торговлей и со спекуляциями в блокчейне. На первый взгляд кажется, что он расходится с более ранними версиями, результаты которых указывали на преобладание незаконных транзакций в блокчейне. В частности, в работе Foley et al. (2019) более 46% транзакций связаны с незаконными операциями. Разница между расчётами происходит по двум причинам: во-первых, Foley et al. (2019) намеренно исключает из своих расчётов все связанные с биржей объёмы, чтобы сосредоточиться только на оплате товаров и услуг. Мы показали выше, что торговля является основным видом деятельности в блокчейне, поэтому этот выбор сильно меняет знаменатель дроби. Во-вторых, оценка объёмов в Foley et al. (2019) основана на надуманной сети незаконных кластеров, в которой любой кластер рекурсивно считается незаконным, если

большинство его транзакций совершается с ранее идентифицированными незаконными кластерами. Несмотря на то, что этот метод вменения интуитивно привлекателен, он не делает различий между реальными пользователями и недолговечными транзитными кластерами, которые существуют исключительно для пресечения отслеживаний. В разделе 3.5 мы покажем, что этот тип паразитных потоков обычно составляет очень большую часть незаконных сделок. В результате объём, рассчитанный этим методом, вероятно, завышает экономическую ценность незаконных сделок[2].

Наш результат, конечно, не означает, что незаконная деятельность в блокчейне Биткоин не является проблемой с точки зрения социального обеспечения. Мы солидарны с общей обеспокоенностью тем, что псевдонимный характер Биткоина способствует злоупотреблениям, таким как незаконная деятельность, уклонение от уплаты налогов или даже взятки. Несмотря на то, что объёмы незаконных торгов BTC оставались относительно стабильными в последние несколько лет, сумма за незаконную деятельность в долларах увеличилась, так как выросла долларовая стоимость BTC. Мы вычислили чистый поток биткоинов по незаконным организациям в период с 2020 г. с разбивкой по типам. Мы считаем, что на адреса, идентифицированные как мошеннические, поступает около 550 млн долларов. Около 16 млн долларов поступает в виде установленных выкупов и более 1,6 млрд долларов в виде «тёмных» платежей и даркнет-сервисов. Около 1,7 млрд долларов поступает на адреса, связанные с азартными играми, ещё 1,4 млрд долларов тратится на оплату за микширование.

Мы считаем важным оценить масштабы транзакционной активности, чтобы понять, каковы основные движущие силы, влияющие на стоимость биткоинов. Наши результаты не подтверждают идею о том, что высокая оценка криптовалют основана на незаконных сделках. Мы предполагаем, что большинство биткоин-транзакций связаны со спекуляциями.

---

[2] Точное сравнение наших результатов с предыдущей работой затруднительно, потому что мы используем существенно больший набор идентифицированных сущностей.

3.3. Центральность сети

В предыдущем разделе мы показали, что биржи криптовалют несут ответственность за большую часть объёма транзакций в сети Биткоин и, вероятно, будут играть в ней доминирующую роль. Чтобы уточнить наше понимание роли бирж, мы проанализировали структуру сети Биткоин. Мы рассматривали наиболее релевантные кластеры, то есть кластеры, идентичность которых известна и которые входят в 10 тыс. кластеров с наибольшим объёмом. С учётом этих ограничений рассмотрим 11 043 учреждения, выполняющих более 55% от общего объёма транзакций. Из-за быстро меняющейся эволюции экосистемы Биткоин мы фокусируемся на периоде с 2018 г. до конца 2020 г. У нас осталось 6 248 сущностей. Для представления этой сети мы используем ориентированный взвешенный сетевой граф, где узел *i* соответствует кластеру *i,* а ребро от узла *i* до *j* соответствует общему потоку биткоинов за период 2018–2020 гг. из кластера *i* в кластер *j*. Полученная сеть состоит из 6 248 узлов и 622 тыс. рёбер. Каждая сущность получает и отправляет в среднем биткоины другим ста объектам (см. график). График 4 показывает подмножество этого сетевого графа Биткоин. Для простоты иллюстрации мы оставляем только узлы, получившие не менее 500 тыс. биткоинов за выбранный период (рис. 4, оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf).

Сеть самых крупных объектов состоит из 23 объектов и 492 рёбер[3].

Узел и размер ребра пропорциональны объёму, полученному объектом, и объёму транзакций между двумя разными объектами. В случае, когда два кластера отправляют потоки друг другу, направление края между этими кластерами соответствует наибольшему потоку, а ребро отмечено красным сегментом.

В статье для построения этой и других сетей мы используем пакет программного обеспечения Graphia, доступный по адресу https://graphia.app/ (Freeman et al., 2020).

---

[3] В направленной сети рёбра от узла *i* до *j* и от узла *j* до *i* считаются отдельными.

Из 23 объектов три (BitGo, Xapo и BixIn) – это онлайн-кошельки, 18 – идентифицированные биржи, два – неизвестные организации (скорее всего, это неидентифицированные биржи или крупные внебиржевые кассы). Они активно взаимодействуют с известными биржами и получают солидную сумму майнерских вознаграждений: 1 252 биткоина и 4 795 биткоинов соответственно.

На рис. 4 проиллюстрирована высокая степень взаимосвязанности между основными биржами. Мы видим, что они образуют почти полный граф, в котором каждый узел соединяется со всеми остальными. И это при том, что биржи работают в разных регионах. Например, Bithumb и Upbit – корейские биржи; bitFlyer – японская; Bitstamp, Coinbase, Gemini и Kraken ориентируются на пользователей из США и Европы, а Huobi, BixIn, OKEx и OceanEx – из Китая. Высокая степень взаимосвязанности важна для регулирования требований KYC, о которых мы скажем в разделе 3.5.

Binance, Huobi и Coinbase являются крупнейшими и самыми активными участники сети Биткоин (рис. 4). Для формальной количественной оценки важности различных сущностей мы вычисляем центральность собственных значений каждой из них в полной сети. Центральность собственных значений для объекта i – это i-й компонент вектор x, который является решением уравнения на собственный вектор:

$$A_x = \lambda_x \qquad (1)$$

где матричные элементы A_{ij} задаются полными потоками биткоинов от объекта i к j в течение 2018–2020 гг., а λ – наибольшее собственное значение, связанное с вектором матрицы A. Центральность собственного вектора учитывает не только полный объём, полученный сущностью, но также структуру сети Биткоин и даёт больший вес кластерам, которые получают большой объём от других кластеров, получающих большой объём[4].

[4] Для более подробной информации см. Newman (2010), 7.1.2. Центральность собственного вектора, основанная на измерениях других сетей, например на общем количестве транзакций, даёт аналогичные результаты.

Список 25 крупнейших организаций с наибольшей центральностью в сети Биткоин приведён на рис. 5 (оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) . Подтверждая наши предыдущие наблюдения (рис. 4), Binance, Coinbase и Huobi демонстрируют высшую меру центральности. Другие биржи (рис. 4) также входят в число 25 наиболее центральных субъектов. Это не должно вызывать удивления, поскольку все эти биржи часть очень плотной сети.

Центральность собственных значений (уравнение 1) подтверждает доминирующую роль бирж в сети Биткоин. В заключение отметим, что потенциально центральность собственных значений может стать новым и полезным показателем значимости биржи.

Другими популярными критериями, которыми ранее пользовались многие агрегаторы данных (например, CoinMarketCap), являются объём торгов вне сети, трафик веб-сайта или количество подписчиков в Twitter. Желательным качеством любого рейтингового показателя должна быть устойчивость к манипуляциям. К сожалению, ни один из существующих критериев не защищён от манипуляций полностью[5].

Из-за того, что центральность собственных значений основана на потоках биткоинов между биржами на блокчейне, для улучшения своего положения в рейтинге биржа должна иметь возможность отправлять большое количество биткоинов на другие биржи. Это может оказаться значительно дороже, чем заниматься демоторговлей (посылать куда-либо средства и тут же получать их обратно) или покупать трафик веб-сайта. Поэтому разумно полагать, что центральность собственных значений может быть более устойчива к манипуляциям, чем другие критерии.

3.4. Межбиржевые потоки

Наш анализ в разделах 3.2 и 3.3 показывает, что большую часть объёма биткоинов составляют потоки между биржами. Чем обусловлены эти потоки? Чтобы ответить на этот вопрос, необходимо признать, что рынки криптовалюты состоят из многих неинтегрированных бирж,

---

[5] См., например, отчёт компании по надзору за рынком криптовалюты BTI Verified: https://btiverified.com/crypto-market-data-report-2020/.

которые имеют независимых владельцев, существуют параллельно и находятся в разных странах. Большинство этих бирж самостоятельно функционирует как традиционные фондовые рынки, где трейдеры подают заказы на покупку и продажу, а биржа завершает сделки на основе централизованной книги заявок. Однако на рынке криптовалют, в отличие от традиционных регулируемых фондовых рынков, нет гарантии того, что инвесторы получат лучшую цену при совершении сделок[6].

Отсутствие таких механизмов повышает роль арбитров, торгующих на разных биржах и обеспечивающих согласованные цены. Предположим, обменный курс между биткоинами и какой-либо другой валютой, скажем С, различается на двух биржах. Идеальной арбитражной сделкой будет обмен биткоинов на С на бирже, где курс обмена высок, и обмен С на биткоины на бирже с невысоким курсом обмена. Затем биткоины и С переносятся между биржами, реализуется безрисковая прибыль.

Вышеописанная сделка сталкивается с небольшими препятствиями, если С является криптовалютой, так как псевдонимный характер криптовалют делает их невосприимчивыми к любому контролю за движением капитала. В случае, если С является фиатной валютой, возможность репатриировать средства из одной страны в другую может быть затруднена из-за трансграничного контроля за капиталом. Рынок может стать потенциально сегментированным.

В работе Игоря Макарова и Антуанетты Шоар (2020) показано, что в период 2017–2018 гг. отклонения цен на криптовалюту на разных биржах были существенными и повторяющимися, что указывает на значительную сегментацию рынка. А арбитражные спреды для обменов в разных странах были намного больше, чем в пределах одной страны. Они определённо связаны с трансграничным контролем за движением капитала[7].

---

[6] Регулирование со стороны Комиссии по ценным бумагам и биржам США (SEC) в рамках положения о Национальной лучшей ставке и предложении (NBBO) в Соединённых Штатах требует от брокеров исполнения клиентских сделок по наилучшим доступным на нескольких биржах ценам.

[7] Сам по себе контроль за трансграничным капиталом может быть мотивом для трансграничных потоков. Поскольку биткоины не подлежат контролю, их можно использовать как средство его обхода. Однако одно это не объясняет потоки между криптовалютными биржами и между биржами одной страны.

Приведённое выше обсуждение предполагает, что:

Биржи внутри страны с сильным контролем за капиталом могут быть более взаимосвязаны друг с другом, чем с другими биржами.

Биржи, торгующие схожими валютными парами, могут демонстрировать более высокие межбиржевые потоки.

Чтобы проверить эти предположения, мы вычисляем два критерия сходства пары бирж. Один основан на схожести криптовалютных пар, торгуемых на каждой бирже. А другой – на сходстве взаимодействия двух бирж с другими биржами в блокчейне Биткоин. Чтобы вычислить схожесть валютных пар, мы используем данные частной фирмы Kaiko, собирающей торговую информацию о криптовалютах с 2014 г. Данные Kaiko охватывают только подмножество бирж, которые мы можем идентифицировать в блокчейне Биткоин, но это крупнейшие биржи. Полученный набор состоит из 57 бирж.

Для каждой биржи мы рассматриваем все торгуемые валютные пары, в которых одной из валют является биткоин. Другой валютой может быть фиатная валюта, стабильные монеты или другие криптовалюты. В общей сложности у нас 4 360 валютных пар на 57 биржах, причём среднее количество валютных пар на бирже составляет 13. Для каждой биржи *i* и криптовалютной пары *j* рассчитываем общий объём торгов за период 2018–2020 гг., деноминированный в биткоинах *(vij)*. Затем мы нормализуем объём на каждой бирже по евклидовой норме:

$$\hat{v}_{ij} = \frac{v_{ij}}{\sqrt{\sum_j v_{ij}^2}},$$

и используем евклидово расстояние между векторами

$$\mathbf{v_i} = \{\hat{v}_{ij}\}_{j=1}^{N} \text{ and } \mathbf{v_k} = \{\hat{v}_{kj}\}_{j=1}^{N}$$

как меру сходства бирж *i* и *k*.

Чтобы вычислить сходство бирж на основе потоков биткоинов, сначала мы вычислим матрицу перекрёстных потоков *A*. Каждый элемент a_{ij} матрицы *A* является средним биткоин-потоком с биржи *i* на биржу *j* и наоборот.

Как и раньше, нормализуем объём на каждой бирже по евклидовой норме:

$$\hat{a}_{ij} = \frac{a_{ij}}{\sqrt{\sum_j a_{ij}^2}},$$

и используем евклидово расстояние между векторами

$$\mathbf{a_i} = \{\hat{a}_{ij}\}_{j=1}^{N} \text{ and } \mathbf{a_k} = \{\hat{a}_{kj}\}_{j=1}^{N},$$

где *j* = от 1, чтобы получить меру сходства бирж *i* и *k* на блокчейне Биткоин.

Чтобы увидеть, выделяют ли две построенные меры сходства одну и ту же группу бирж, мы применяем алгоритмы кластеризации K-medoids к каждой из мер сходства. Алгоритмы K-medoids пытаются сгруппировать похожие биржи вместе, чтобы минимизировать внутрикластерную сумму расстояний между биржами. Это популярный метод кластеризации во многих случаях[8].

На рис. 6 (оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) показан результат применения алгоритмов кластеризации K-medoids на основе меры сходства валютных пар. Мы видим, что есть четыре заметные группы обменов. Это американо-европейские (группа 5), корейские (3), биржи японской группы (4) и Tether (2). Резкая кластеризация корейских и японских бирж объясняется тем, что в качестве базовых на этих биржах используются национальные фиатные валюты, торговля осуществляется небольшим количеством валютных пар. Аналогичная ситуация наблюдается на американо-европейских биржах, где в качестве базовой валюты обслуживаются и доллар, и евро, причём доллар обычно более популярен. Большинство бирж 2-й группы – это только криптовалютные

[8] В нашем анализе мы используем процедуру K-medoids из модуля Python sklearn extra (см. Hastie et al. (2001), 14.3.10 для более подробной информации). Алгоритмы кластеризации K-medoids принимают в качестве параметра количество кластеров. Поскольку для его определения нет строгих правил, мы по умолчанию используем значение 8.

биржи, которые не предлагают торговлю против фиатной валюты. Обычно эти биржи в качестве базовой валюты используют Tether и размещают большое количество различных криптовалют для торговли. Также есть несколько разрозненных бирж с менее популярными базовыми валютами. Например, Coinfloor использует британский фунт стерлингов, BitBay – польский злотый, а ACX – австралийский доллар.

Затем мы применили алгоритмы кластеризации K-medoids к измерению сходства цепочки биткоинов (рис. 7, оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf). Сравнивая рис. 6 и 7, можно увидеть, что разница в расстоянии между биржами внутри кластера и биржами из разных кластеров не так ярко выражена, как в случае кластеризации на основе криптовалютных пар. Это не должно удивлять, так как интеграция биржи зависит не только от торгуемых криптовалют, но и от средств контроля за капиталом, действующих в данной стране. Две биржи в разных странах могут быть разделены значительным расстоянием между криптовалютными парами, если они используют разные фиатные валюты, но могут быть очень похожими на расстоянии блокчейна Биткоин, если они работают в странах без контроля за движением капитала.

Тем не менее метод кластеризации основных групп бирж, основанный на двух показателях сходства, в целом даёт аналогичные результаты (рис. 6, 7). Особенно отчётливо в обоих случаях сгруппированы вместе корейские и японские биржи. Американские, европейские и Tether разделены менее чётко. Например, Poloniex, которая является только криптовалютной биржей, теперь сгруппирована вместе с Coinbase, Bitstamp, «Близнецы», Kraken и Bitfinex. Это согласуется с результатами работы Игоря Макарова и Антуанетты Шоар (2020), в которой показано, что американо-европейские биржи и биржи Tether интегрированы лучше, чем биржи в Корее и Японии.

3.5. Обеспечение соблюдения норм KYC для биткоин-транзакций

Мы завершаем наше исследование объёма потоков биткоинов анализом, связанным с теневой экономикой. Нежёсткое (лёгкое) регулирование и анонимность сделали криптовалюты популярными у всех, кто хочет избежать юридической или нормативной проверки, укло-

ниться от уплаты налогов. Сторонники криптовалюты часто указывают на то, что она по-прежнему превосходит наличные из-за своего цифрового следа. Цифровой след действительно накладывает некоторые ограничения на анонимность транзакций и в некоторых случаях помогает найти правонарушителей, однако есть и существенные ограничения.

Чтобы понять проблемы, связанные с соблюдением норм «Знай своего клиента» (KYC), полезно и поучительно рассмотреть сеть, сосредоточенную на Рынке Гидры (Hydra Market), который является одной из крупнейших торговых площадок в «тёмной» сети[9].

Hydra Market начал работу в 2015 г. и с тех пор быстро растёт. Мы ориентируемся на период 2020 г. – июнь 2021 г. За этот период на Hydra Market поступило 147 620 биткоинов из 514 855 кластеров. Биткоины отправлены в 315 359 кластеров. Эти 514 855 отправляющих в Hydra Market кластеров, в свою очередь, получили свои потоки из 3 291 180 кластеров, а вышеупомянутые 315 359 кластеров-получателей отправили потоки в 500 544 кластера, из них 116 131 новые кластеры.

На рис. 8 (оригинал рисунка здесь: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) изображена получившаяся сеть, где для простоты иллюстрации мы сохраняем только узлы, которые отправляют не менее тысячи биткоинов в этой сети. При вычислении объёма мы намеренно исключаем объём между любыми кластерами, которые принадлежат к списку известных или крупных объектов, который мы изучали в разделах 3.3 и 3.4.

Размер узла отражает общее количество биткоинов, отправленных с Hydra Market соответствующему лицу или организации. Размер связующей линии пропорционален объёму потоков между двумя разными объектами. В случае, когда два кластера отправляют потоки друг другу, длина линии между этими кластерами соответствует наибольшему потоку, а линия изображена в красном сегменте. Оранжевый цвет показывает идентифицированные кластеры, зелёный отмечает неизвестные кластеры большого объёма, бирюзовый – короткоживущие кластеры

---

[9] https://www.bloomberg.com/news/articles/2021-02-01/darknet-market-had-a-record-2020-ledby-russian-bazaar-hydra.

с продолжительностью жизни менее одного месяца. Оставшиеся кластеры отмечены фиолетовым цветом.

На рис. 8 продемонстрировано, что объектами с наибольшим объёмом, напрямую взаимодействующими с Hydra Market, являются биржи, не поддерживающие KYC, такие как LocalBitcoins, Bitzlato, Binance, Huobi и Totalcoin[10].

Поступая на биржи, потоки смешиваются и становятся практически неотслеживаемыми, поэтому впоследствии могут быть отправлены куда угодно.

Рисунок также демонстрирует, что прямое взаимодействие Hydra Market с теми биржами, которые пытаются обеспечить соблюдение норм KYC, такими как Coinbase и Gemini, скромны, но их взаимодействие с соседними кластерами значительно шире. Например, Coinbase напрямую отправила и получила 196 и 126 биткоинов с Hydra Market соответственно. Но она отправила 530 тыс. и получила 218 тыс. биткоинов через соседние кластеры.

Мы видим (рис. 8), что большинство потоков на Coinbase и обратно происходят через короткоживущие кластеры, которые в большинстве случаев создаются с единственной целью сокрытия происхождения средств. Типичная транзакция включает смешивание испорченных средств (те, которые можно отследить до Hydra Market) с «чистыми» (не связанными с незаконными сделками). Каждое смешивание снижает долю испорченных потоков. Процесс повторяется несколько раз, пока полученные потоки не станут достаточно чистыми, чтобы отправить их в KYC-биржи.

Каковы последствия? Во-первых, учреждения, не участвующие в KYC, служат для отмывания денег и другой серой деятельности. Децентрализованный характер протокола Биткоин упрощает работу этих организаций – им нужно только иметь свои серверы в стране, где власти готовы мириться с их существованием. Если организациям KYC разрешено принимать потоки от организаций, которые не соблюдают строгие нормы KYC (текущее состояние дел), то цифровой след очень ограниченно предотвращает попадание загрязнённых потоков в широкое распространение. Возможность торговать «монетами конфиденциаль-

[10] https://bitshills.com/best-non-kyc-crypto-exchanges/.

ности», такими как Monero, и растущая популярность платформ DeFi способствуют реализации этих стратегий отмывания денег.

Во-вторых, даже если бы учреждениям, соблюдающим KYC, разрешить иметь дело исключительно с другими KYC-организациями, предотвратить приток испорченных средств по-прежнему будет почти невозможно, разве что кто-нибудь наложит строгие ограничения на то, кто с кем может совершать сделки, и сделает все транзакции подлежащими одобрению аналитическими компаниями типа блокчейн, такими как как Bitfury Crystal Blockchain и Chainalysis. Если бы такой режим был реализован, эти фирмы стали бы фактически доверенными сторонами, необходимыми для функционирования сети Биткоин. Но протокол Биткоин предназначен именно для того, чтобы обойти ограничения.

Операции с наличными деньгами и их хранение требуют значительных затрат и создают операционные риски, а операции с криптовалютами и их хранение практически бесплатны (за исключением колебаний стоимости биткоинов). Чем более широкое распространение получит биткоин, тем проще будет использовать его для транзакций без участия регулируемых юридических лиц, и тем более привлекательным он станет для должностных преступлений и теневой экономики.

Перевод А. И. Земскова, ГПНТБ России
(Продолжение в следующем номере 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/ (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

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

DIGITAL INFORMATION RESOURCES

UDC 004.65:[336.743-021.131]-047.44
https://doi.org/10.33186/1027-3689-2022-10-100-114

Blockchain analysis of the Bitcoin market. (Part 2)

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 2) / I. Makarov, A. Shoar // Scientific and technical libraries. 2022. No. 10. P. 100–114. https://doi.org/10.33186/1027-3689-2022-10-100-114

3.2. Real volume

We now focus on the economically meaningful, non-spurious, part of Bitcoin volume. To understand for what purposes Bitcoin is utilized, we trace Bitcoin flows between different types of entities on the blockchain. Our list of known entities includes exchanges, on-line wallets, payment processors, gambling sites, mixing services, illegal services, and mining pools. We identify these entities from a large number of public and proprietary sources as described in the data section.

Cryptocurrency exchanges such as Coinbase, Binance, or Kraken, and on-line wallets such as Blockchain.info and BixIn are one of the major types of entities where Bitcoin can be stored and traded. Exchanges in theory provide platforms to trade Bitcoin against fiat currencies and other coins, while on-line wallets specialize in custodian services. However, in practice,

the difference between exchanges and on-line wallets is often slim. Both types of entities in many cases offer both functions. Therefore, we group these entities together when providing a general overview of Bitcoin utilization. Payment processors, such as BitPay or CoinPayments, facilitate payments by on-line shops, gambling, and other entities that accept cryptocurrencies as means of payment for good and services. Illegal services include dark net marketplaces such as Hydra Market, numerous ransomware wallets, and entities engaged in scams. Mixing services or tumblers such as Bitcoin Fog and Wasabi wallet are sites that allow their customers to pool together their funds in order to obfuscate where the coins are being sent from.

Another set of entities that are a core component of the Bitcoin system are mining pools and miners. We identify miners by tracing rewards distribution of the largest mining pools to individual miners. We describe how we trace miners in Section 4 and in the Appendix. Overall, we identify 248,000 miners in the data.

As previously discussed, the pseudonymous nature of Bitcoin makes it difficult to link an address to the real-world entity behind them. Thus the identification of entities is incomplete almost by design, since it relies on an entity either voluntarily disclosing its addresses or learning about an entity's addresses in the course of interaction with it.

To address the problem of incomplete identification of entities and to make sure that we are not missing major players on the blockchain, we analyze the top 10,000 unknown clusters with the largest Bitcoin volume, for which we were not able to find an identity. Out of this universe of clusters, we select those that either receive regular flows from miners or receive more than 50% of its inflow from known exchanges and send more than 40% of its outflow to known exchanges. These thresholds are determined from the transaction patterns of known entities. For a typical exchange, 53% of its Bitcoin outflow goes to other exchanges, and 52% of its Bitcoin inflow comes from other exchanges. These numbers are significantly lower for all other entities. For example, a typical gambling site sends 21% and receives 29% of its total flows from exchanges. We find that 4507 clusters satisfy the above conditions. Taken together they account for 63% of the bitcoins

flowing to the largest 10,000 clusters. In what follows, we refer to these clusters as LEOTD, Likely Exchanges, OTC brokers, or Trading Desks.

Based on this classification of participants, in Figure 3 (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) we plot the average monthly transaction volume that is generated by these different types of entities on the blockchain from the beginning of 2015 until May 2021. The volume is calculated as the amount of bitcoins that are sent to different types of entities in a given month. Figure A shows the volume in BTC and Figure B as the percentage of the total monthly volume.

We see that the majority of the volume is generated by transactions involving exchanges and LEOTD clusters. Volume flowing to known exchanges constitutes about 40% of total volume and another 20% of the volume is generated by volume flowing to LEOTD. To highlight the dominant role of exchanges and LEOTDs, we split volume that goes to the Other category, which consists of all unknown clusters that are not LEOTDs, into two parts: volume coming from exchanges and LEOTD and the rest. This decomposition shows that volume from exchanges and LEOTD to Other explains another 20% of the volume. Thus, exchange and trading desk related volume constitutes about 80% of the total volume. Other known entities are only responsible for a minor part of total volume as of the end of 2020. For example, illegal transactions, scams, and gambling together make up less than 3% of the volume. The fraction of volume explained by miners is even smaller.

This analysis of volume underscores the dominance of trading and speculation related transactions on the blockchain, and at first glance seems to be at odds with earlier results that emphasized the prevalence of illegal transactions on the blockchain. Most notably, Foley et al. (2019) estimates that more than 46% of transactions are due to illegal transactions. The difference between their calculations and ours comes from two main sources. First, Foley et al. (2019) intentionally drop all exchange-related volumes from their calculations, since they want to focus only on payments for goods and services. Since we show above that trading constitutes the main activity on the blockchain, this choice severely changes the denominator. Second, the estimate of volume in Foley et al. (2019) is based on an imputed

network of illegal clusters where any cluster recursively is deemed illegal if the majority of its transactions is with previously identified illegal clusters. While intuitively appealing, this imputation method does not discriminate between real users and short-lived pass-through clusters that exist solely to obfuscate tracing. We show in Section 3.5 that this type of spurious volume is typically a very large part of illegal transactions. As a result, volume imputed by this method is likely to overstate the economic value of illegal trades[1].

Our results of course do not mean that illegal activities on the Bitcoin blockchain are not a problem from the perspective of social welfare. We agree with the general concern that the pseudonymous nature of Bitcoin facilitates malfeasance such as illegal activities, tax evasion, or even bribes. Even though the BTC volume of illegal trades has stayed relatively stable in the last few years, the dollar amount of illegal activities increased, since the dollar value of BTC went up. We compute the net flow of bitcoins to illegal entities over 2020, broken down by their specific types. We calculate that there are about $550 million flowing to addresses that have been identified as scams, about $16 million in identified ransom payments, and more than $1.6 billion for dark net payments and dark net services. In addition, there are about $1.7 billion flowing to addresses affiliated with gambling and another $1.4 billion in mixing services.

In sum, we think it is important to get the magnitudes of transaction activities right in order to understand what are the ultimate drivers of Bitcoin value. Our results do not support the idea that the high valuation of cryptocurrencies is based on the demand from illegal transactions. Instead, they suggest that the majority of Bitcoin transactions is linked to speculation.

3.3. Network centrality

In the previous section, we show that cryptocurrency exchanges are responsible for the majority of volume on the Bitcoin network, and are therefore likely to play a dominant role in the network. To sharpen our understanding of the role exchanges play, we now analyze the structure of the Bitcoin network.

---

[1] The exact comparison of our results to the prior paper is difficult because we use a substantially larger set of identified entities.

In our network analysis, we restrict our attention to the most relevant clusters, i.e. clusters for which we know their identity and that are in the top 10,000 highest volume clusters. With these filters, we have 11,043 entities, which account for more than 55% of the total volume. Because of the rapidly changing evolution of the Bitcoin ecosystem, we focus on the most recent time period: from 2018 to the end of 2020, which leaves us with 6248 entities. To represent this network, we use a directed weighted network graph, where a node *i* corresponds to cluster *i* and an edge (or link) from node *i* to *j* corresponds to the total Bitcoin flows over the period 2018–2020 from cluster *i* to cluster *j*. The resulting network consists of 6248 nodes and 622K edges. Each entity receives and sends bitcoins to the other 100 entities in the graph, on average. Figure 4 (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) plots a subset of this Bitcoin network graph, where for ease of illustration we retain only nodes that received at least 500,000 bitcoins over the period from 2018 to the end of 2020[2].

The network of the largest entities consists of 23 entities and 492 edges[3].

The node and edge size are proportional to the volume received by the entity and the volume between two different entities. In the case when two clusters send flows to each other, the direction of the edge between these clusters agrees with the largest flow, and the edge is marked with a red segment.

Out of 23 entities, three entities (BitGo, Xapo, and BixIn) are on-line wallets, 18 are identified exchanges, and two are unknown entities. The two unknown entities are likely to be unidentified exchanges or large OTC desks. They actively interact with known exchanges and receive a large amount of miners' rewards; 1 252 and 4 795 bitcoins, respectively.

Figure 4 reveals a high degree of interconnectedness between the major exchanges. We can see that they form an almost complete graph, where each node connects to all others. This is despite the fact that these exchanges operate in different regions. For example, Bithumb and Upbit are Korean

---

[2] To plot this and other networks in this paper we use Graphia package software available at https://graphia.app/, Freeman et al. (2020).

[3] In a directed network, an edge from node i to j and an edge from node j to i count as separate edges.

exchanges, bitFlyer is Japanese, Bitstamp, Coinbase, Gemini, and Kraken are geared towards US and European users, and Huobi, BixIn, OKEx, and OceanEx towards Chinese. The high degree of interconnectedness has important implications for KYC regulation which we address in Section 3.5.

Inspection of Figure 4 further shows that Binance, Huobi, and Coinbase are the largest and the most active participants in the Bitcoin network. To formally quantify the importance of different entities, we compute the eigenvalue centrality of each entity in the full network. The eigenvalue centrality for an entity i is the i^{th} component of vector x, which is the solution to the eigenvector equation:

$$A_x = \lambda_x, \quad (1)$$

where matrix elements A_{ij} are given by the total Bitcoin flows from entity i to j over 2018-2020, and λ is the largest eigenvalue associated with the eigenvector of matrix A. The eigenvector centrality takes into account not only the total volume received by an entity but also the structure of the Bitcoin network and gives larger weights to clusters that receive large volume from clusters that receive large volume themselves[4].

Figure 5 (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) shows the top 25 entities with the largest Bitcoin network centrality. Confirming our earlier observation based on Figure 4, Binance, Coinbase, and Huobi have the highest measure of centrality. Other exchanges from Figure 4 are also among the most central 25 entities. This should not come as surprise since all these exchanges are part of a very dense network.

The eigenvalue centrality (1) confirms the dominant role of exchanges in the Bitcoin network. We conclude this section by noticing that the eigenvalue centrality can potentially serve as a new and useful measure of the importance of exchanges.

Other popular measures that have previously been used by many data aggregators such as CoinMarketCap include (1) off-chain exchange trading volume, website traffic, or the number of Twitter followers. A desirable property for any ranking measure is to be resilient to manipulation.

[4] See Newman (2010), 7.1.2 for more details. Eigenvector centrality based on other network measures such the total number of transactions produces similar results.

Unfortunately, none of the existing measures seem to be fully manipulation-proof[5].

Since the eigenvalue centrality measure is based on the cross-exchange bitcoin flows on the blockchain, to improve its position in the ranking an exchange would have to send back and forth large amounts of bitcoins to other exchanges. This can prove significantly more costly than simply engaging in wash trading or buying website traffic. Therefore it is reasonable to believe that the eigenvalue centrality can be more resilient to manipulation than other measures.

3.4. Cross-exchange flows

Our analysis in Sections 3.2 and 3.3 shows that a large part of the Bitcoin volume is driven by cross-exchange flows. What explains these flows? To answer this question, it is important to recognize that cryptocurrency markets consist of many non-integrated exchanges that are independently owned and exist in parallel within and across countries. On an individual basis, the majority of these exchanges function like traditional equity markets where traders submit buy and sell orders, and the exchange clears trades based on a centralized order book. However, in contrast to traditional, regulated equity markets, the cryptocurrency market lacks any provisions to ensure that investors receive the best price when executing trades[6].

The absence of such mechanisms increases the importance of arbitrageurs who trade across different exchanges and ensure consistent prices across them. Suppose an exchange rate between Bitcoin and some other currency, say *C*, is different across two exchanges. An ideal arbitrage trade would be to exchange Bitcoin for *C* on the exchange where the exchange rate is high and exchange *C* for Bitcoin on the exchange with a low exchange rate; then transfer Bitcoin and *C* between exchanges and realize the risk free profit.

---

[5] See, for example, a report from cryptocurrency market surveillance firm BTI Verified: https://btiverified.com/crypto-market-data-report-2020/.

[6] For example, the US Securities and Exchange Commission (SEC)'s National Best Bid and Offer (NBBO) regulation in the United States requires brokers to execute customer trades at the best available prices across multiple exchanges.

The above trade faces few obstacles if *C* is a cryptocurrency since by design, the pseudonymous nature of cryptocurrencies makes them immune to any capital controls. However, when *C* is a fiat currency the ability to repatriate funds from one country to another may be obstructed by cross-border capital controls, and the market can become potentially segmented.

In Makarov and Schoar (2020) we indeed show that in the period 2017–2018 there were large and recurring deviations in cryptocurrency prices across exchanges, pointing to significant market segmentation. We also showed that the arbitrage spreads were much larger for exchanges across different countries than within the same country, and were positively linked to cross-border capital controls[7].

The above discussion suggests that (1) exchanges within a country with strong capital controls can be more interconnected with each other than with other exchanges, and (2) exchanges that trade similar currency pairs can see higher cross-exchange flows.

To test these predictions, we compute two measures of similarity of a pair of exchanges. One is based on the similarity of the cryptocurrency pairs traded on each exchange. And the other is based on the similarity of the interaction of the two exchanges with other exchanges on the Bitcoin blockchain. To compute the currency-pair similarity we use Kaiko data, a private firm that has been collecting trading information about cryptocurrencies since 2014. The Kaiko data cover only a subset of exchanges that we can identify on the Bitcoin blockchain, but these are the largest exchanges. The joint set consists of 57 exchanges.

For each exchange, we consider all traded currency pairs where one of the currencies is Bitcoin. The other currency could be a fiat currency, stable coins, or other cryptocurrencies. In total, we have 4,360 currency pairs across 57 exchanges, with the median number of currency pairs on an exchange being 13. For each exchange i and cryptocurrency pair j, we compute the total trading volume in the period 2018-2020 denominated

[7] Cross-border capital controls can be a motif for cross-exchange flows themselves. Since Bitcoin is not subject to capital controls one can use it as a means to bypass them. This alone, however, cannot explain flows between crypto-only exchanges and flows across exchanges within the same country.

in Bitcoin, v_{ij}. Next, we normalize the volume on each exchange by the Euclidean norm:

$$\hat{v}_{ij} = \frac{v_{ij}}{\sqrt{\sum_j v_{ij}^2}},$$

and use the Euclidean distance between vectors

$$\mathbf{v_i} = \{\hat{v}_{ij}\}_{j=1}^N \text{ and } \mathbf{v_k} = \{\hat{v}_{kj}\}_{j=1}^N$$

measure of similarity of exchanges *i* and *k*.

To compute the exchange similarity based on the Bitcoin flows we first calculate the matrix of cross-exchange flows, *A*. Each element a_{ij} of matrix *A* is the average of Bitcoin flows from exchange ***i*** to exchange ***j*** and vice versa. As before, we normalize the volume on each exchange by the Euclidean norm

$$\hat{a}_{ij} = \frac{a_{ij}}{\sqrt{\sum_j a_{ij}^2}},$$

and use the Euclidean distance between vectors

$$\mathbf{a_i} = \{\hat{a}_{ij}\}_{j=1}^N \text{ and } \mathbf{a_k} = \{\hat{a}_{kj}\}_{j=1}^N$$

to obtain a measure of similarity of exchanges *i* and *k* on the Bitcoin blockchain.

To see if the two constructed similarity measures isolate the same group of exchanges we apply the K-medoids clustering algorithms to each of the similarity measures. The K-Medoids algorithms tries to group similar exchanges together to minimize the withincluster sum of distances between exchanges. It is a popular clustering method available in many packages[8].

Figure 6 (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) shows the result of the

[8] We use KMedoids routine from Python module sklearn extra in our analysis, see Hastie et al. (2001), 14.3.10 for more details.

application of K-medoids clustering algorithms based on the currency-pair similarity measure. We can see that there are four no- table groups of exchanges. These are US-European (group 5), Korean (group 3), Japanese group (4), and Tether (group 2) exchanges. The sharp clustering of Korean and Japanese exchanges reflects the fact that these exchanges use the national fiat currency as the base currency and trade a small number of currency pairs. A similar situation holds for US-European exchanges where both dollar and euro serve as a base currency, with the dollar usually being more popular. The majority of group 2 exchanges are crypto-only exchanges, which do not offer an opportunity to trade against a fiat currency. These exchanges usually use Tether as a base currency and list a large number of different cryptos for trading. There are also a few isolated exchanges that have less popular base currencies. For example, Coinfloor uses British pound, BitBay Polish zloty, and ACX Australian dollar.

Next, we apply the K-medoids clustering algorithms to the Bitcoin blockchain similarity measure. Figure 7 (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) shows the results[9]. Comparing Figures 6 and 7, we can see that the difference in distance between exchanges within a cluster and exchanges from different clusters is not as pronounced as in the case of clustering based on the cryptocurrency pairs. This should not come as a surprise since exchange integration depends not only on the cryptocurrencies traded but also on the capital controls that are in place in a given country. Two exchanges in different countries can be well separated in the cryptocurrency pair distance if they use different fiat currencies but can be very similar in the Bitcoin blockchain distance if they operate in countries without capital controls.

Nevertheless, Figures 6 and 7 show that clustering of major groups of exchanges based on the two similarity metrics produces broadly similar results. In particular, the Korean and Japanese exchanges in both cases are grouped together. The US-European and Tether exchanges are less clearly separated. For example, Poloniex, which is crypto-only exchange, is now

[9] K-medoids clustering algorithms takes the number of clusters as a parameter. Since there is no rigorous theory to determine it, we use a default value of 8.

grouped together with Coinbase, Bitstamp, Gemini, Kraken, and Bitfinex. This is again consistent with our results in Makarov and Schoar (2020), where we show that the US-European and Tether exchanges are better integrated than exchanges in Korea and Japan.

3.5. Enforcement of KYC norms for Bitcoin Transactions

We conclude our study of Bitcoin volume with the analysis of flows associated with the shadow economy. Light regulation and the anonymity of cryptocurrencies have made them a popular choice for anyone who wants to evade legal or regulatory scrutiny or engage in tax evasion. Proponents of cryptocurrencies often like to point out that cryptocurrencies are still superior to cash because of their digital footprint. While the digital footprint indeed imposes some constraints to the anonymity of transactions, and in some cases helped catch offenders, it is important to realize that there are strong limitations.

To understand the challenges of enforcing Know-Your-Customer (KYC) norms, it is instructive to consider a network centered on Hydra Market, which is one of the largest dark net marketplaces[10].

Hydra Market has been in operation since 2015 and has been growing rapidly since then. We focus on the most recent period, 2020 – June 2021. Over this period, Hydra market received 147,620 bitcoins from 514,855 clusters and sent them to 315,359 clusters. The 514855 sending clusters in turn received their flows from 3,291,180 clusters, and the 315,359 sending clusters sent to 500,544 clusters, of which 116,131 are new clusters. be very similar in the Bitcoin blockchain distance if they operate in countries without capital controls.

Figure 8 (view original figure here: https://www.nber.org/system/files/working_papers/w29396/w29396.pdf) depicts the resulting network, where for ease of illustration we retain only nodes that send at least 1000 bitcoins *within* this network. When computing volume in this network we intentionally exclude volume between any clusters that

[10] https://www.bloomberg.com/news/articles/2021-02-01/darknet-market-had-a-record-2020-led-by-russian-bazaar-hydra.

belong to the list of known or high volume entities, which we studied in Sections 3.3 and 3.4.

The node size reflects the total amount of bitcoins sent from Hydra Market to a corresponding entity. The edge size is proportional to the volume between two different entities. In the case when two clusters send flows to each other, the direction of the edge between these clusters agrees with the largest flow, and the edge is depicted with a red segment. The orange color shows identified clusters, the green color marks unknown high volume clusters, the turquoise color shows short-lived clusters with a lifespan below one month, the purple color marks the remaining clusters.

Figure 8 reveals that the highest volume entities interacting directly with Hydra Market are non-KYC exchanges such as LocalBitcoins, Bitzlato, Binance, Huobi, and Totalcoin[11]. Once the flows arrive at these exchanges they get mixed with other flows and become virtually untraceable, and so can be sent anywhere afterwards.

The figure also shows that direct interactions of Hydra Market with those exchanges that try to enforce KYC norms, such as Coinbase and Gemini, are modest; but their interaction with the neighboring clusters is significantly larger. For example, Coinbase directly sent and received 196 and 126 bitcoins from Hydra Market, respectively. But it sent 530,000 and received 218,000 bitcoins via the neighboring clusters.

Looking at Figure 8 we can see that the majority of flows to and from Coinbase occur through short-lived clusters, which in most cases are created for the sole purpose of obfuscating the origin of funds. A typical transaction involves mixing tainted funds (those that can be traced to Hydra Market) with "clean" (not traceable to illegal transactions). Each mixing reduces the share of tainted flows. The process is repeated several times until the resulting flows become clean enough to send them to KYC exchanges.

What are the implications of the above analysis? First, 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.

---

[11] https://bitshills.com/best-non-kyc-crypto-exchanges/.

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. The ability to trade "privacy coins" such as Monero and the increasing popularity of DeFi platforms further facilitate these money laundering strategies.

Second, 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 type of blockchain analytics companies such as Bitfury Crystal Blockchain and Chainalysis. Note that if this regime was to realize these firms would become the de facto trusted parties essential for the functioning of the Bitcoin network. But this is exactly what the Bitcoin protocol is designed to circumvent. If trusted parties exist there are simpler and more efficient solutions than the Bitcoin protocol, e. g., a permissioned blockchain.

Finally, notice that while transacting in cash and storing cash involve substantial costs and operational risks, transacting in cryptocurrencies and storing them are essentially costless (apart from fluctuation in value). The wider the adoption of Bitcoin is, the easier it will be to use it for transactions without ever having to touch regulated entities, and the more attractive it will become for malfeasance and shadow economy.

(To be continued.)

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. **Ron D., Sham.** 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