Estudio de la red Bitcoin

Marcelo Ortega

Abstract— **Bitcoin es un sistema de pago digital que desde su aparición en 2009 no ha dejado de popularizarse mientras que su valor se incrementa exponencialmente. Al ser un sistema descentralizado, la información de cuánto dinero es poseído por cada usuario no es confiada a un nodo central, sino que es registrada en una base de datos distribuida a la que cualquier usuario puede tener acceso. Esta base de datos guarda la información de todas las transacciones existidas desde el inicio de la red, lo que puede ser modelado como un grafo en el que los nodos correspondan a direcciones Bitcoin y las aristas que los unen a los movimientos de capital ocurridos. La actividad de los usuarios en Bitcoin puede ser analizada a través del estudio de este grafo, utilizando caminatas aleatorias para estudiar las tendencias de los flujos de capital.**

I. INTRODUCCIÓN

Bitcoin es un sistema de pago digital que permite la transferencia de esta moneda digital de una persona a otra de forma descentralizada, sin depender del beneplácito de una entidad central como puede ser un banco o el proveedor de una tarjeta de crédito. Propuesto en 2008 por [18], y puesto en funcionamiento por primera vez en Enero del 2009, en sus comienzos el sistema era usado solo por un pequeño grupo de aficionados a la criptografía y con el transcurrir de los años la comunidad y su uso se fue expandiendo a la par de su precio con respecto al dólar.

Bitcoin se basa en un sistema de base de datos distribuida llamado Blockchain para mantener un registro inmutable de todas las transacciones realizadas desde el inicio y lograr un consenso entre sus participantes sobre su veracidad. Cada transacción de bitcoins debe ser aceptada por todos los participantes de la red, quienes validarán que el remitente realmente cuente con los fondos transferidos y que estos no hayan sido gastados previamente. Una vez aceptada, la transacción es ingresada al registro y los bitcoins transferidos quedarán en posesión del destinatario. La validación de las transacciones es efectuada por nodos llamados mineros, que ofrecen a la red su capacidad de cómputo a cambio del cobro de una tarifa en cada movimiento. No se necesita ningún permiso especial para realizar esta llamada minería, y si bien los mineros son de alguna forma los que controlan la red al poder decidir cuáles transacciones validar y cuales no, es de su interés actuar de forma legal para mantener la confianza en la red y por lo tanto su valor. Mediante este sistema Bitcoin soluciona el problema del doble gasto, un defecto potencial del dinero digital por el cual una misma moneda podría gastarse más de una vez, por tratarse de un archivo que puede falsificarse. Notablemente, Bitcoin soluciona este problema presidiendo de un actor central confiable que determine en todos los casos si una moneda fue gastada previamente o no.

Pese a lo novedoso de la solución tecnológica propuesta en Bitcoin, en los últimos años el interés hacia él ha girado más bien en torno a el constante incremento de su precio. La cotización de bitcoin en dólares estadounidenses alcanzó la paridad en Febrero del 2011, y seis años mas tarde, en Octubre del 2017, sobrepasaba por primera vez los $5.000 [8]. La evolución de su precio es seguida regularmente por la prensa internacional, en noticias que alternan a través de los meses entre caídas en su precio y denuncias a la veracidad de su valor [5], [4], [3], y reportes de máximos históricos en su precio alcanzados [12], [7], [6]. En las opiniones con respecto al futuro de Bitcoin, se puede encontrar una polarización entre quienes pregonan continuidad en el aumento de su precio, a menudo personas asociadas a emprendimientos en sistemas de pagos digitales [9]; y sus detractores quienes consideran que su evolución corresponde a una burbuja financiera que acabará explotando y desplomando rápidamente su precio, a menudo asociados a corporaciones bancarias y grupos inversores [2].

La ausencia de una autoridad central en el sistema es la principal característica que diferencia a Bitcoin y otras monedas digitales basadas en blockchain de los sistemas de pago convencionales. Esta descentralización del sistema implica que el registro de transacciones es de libre acceso, cualquier persona puede descargarlo y proceder a validar la legalidad de todas las transacciones realizadas desde la primera hasta la más reciente. La disponibilidad de este registro permite el análisis de la información completa del sistema; en contraste con los sistemas de pago convencionales en los que la información de las transacciones digitales es reservada para proteger la privacidad de sus usuarios; en Bitcoin es posible analizar en profundidad todas las transacciones, sus remitentes, destinatarios y las cantidades transferidas.

Aprovechando la abundancia de datos del sistema financiero Bitcoin, este trabajo se propone investigar el comportamiento de las transacciones realizadas y que relaciones se pueden establecer entre el comportamiento general del sistema y los cambios en la cotización del bitcoin. Concretamente, se propone modelar el sistema como un grafo en el que los vértices corresponden a usuarios de bitcoin y las aristas que los unen a los flujos de dinero resultado de las transacciones observadas en el registro histórico, para aplicar en el análisis postulados y resultados de la Teoría de Redes.

I-A. Trabajos previos

La primera publicación académica centrada en analizar el sistema bitcoin a través de los datos sobre sus transacciones fue publicada por Ron y Shamir [20] en el 2012. En ella se obtienen algunas métricas básicas de la red, como la distribución del tamaño de las transacciones y del capital

acumulado en las diferentes direcciones. Se encuentra que la mayor parte de la riqueza es concentrada sin ser movilizada en unas pocas direcciones, y que aunque que la mayoría de las transacciones mueven pequeñas fracciones de bitcoin, otros cientos de ellas operan con más de 50.000 BTC. El trabajo prosigue analizando patrones de conducta en las transacciones, principalmente en usuarios asociados a actividades ilícitas que intentan ocultar el origen de sus fondos. Este estudio sobre los límites del anonimato en la red es retomado por Meiklejohn et al. [17] quienes realizan un etiquetado de las distintos usuarios, clasificándolos en sitios de intercambio, servicios, cuentas personales y vendedores, para investigar como fueron gastados fondos ilícitos asociados a el sitio Silk Road, un mercado online cerrado por el FBI en 2013.

Kondor et al [16] se enfocan en estudiar la repartición de riqueza en bitcoin, encontrando que esta converge después del 2011 a una distribución de estiramiento exponencial estable, y que las conexiones entre los nodos siguen una conexión preferencial sublineal. Investigando la correlación entre la distribución de riqueza y la topología de la red, identifican una relación entre el grado y la riqueza de un nodo, implicando que la habilidad de atraer nuevas conexiones esta relacionada a la de adquirir riqueza. Todos estos trabajos analizan la totalidad de las transacciones realizadas hasta el momento. En 2015, cuando el tamaño de la red ya se había incrementado varios órdenes mas, [1] utiliza para su estudio el grafo resultado de aislar las transacciones realizadas en intervalos de tiempo dados. En él se realiza un planteo inicial de la utilización de métricas de la red junto con información de fuentes externas como análisis de sentimiento en comentarios extraídos de Twitter para predecir las fluctuaciones en la cotización de bitcoin.

II. Modelado de la red

Bitcoin puede ser visto como una cadena de transacciones que mueven valores de bitcoin de un lugar a otro. Estos lugares donde se almacenan los bitcoins reciben el nombre de direcciones y son identificados por claves públicas alfanuméricas. No hay restricciones a cuantas direcciones pueda poseer un mismo usuario, y no hay forma de relacionar inequívocamente la pertenencia de dos o más direcciones al mismo usuario, ni su verdadera identidad. Bitcoin es llamado un sistema pseudo-anónimo, ya que mientras por un lado todas las transacciones son públicas, la identidad de los usuarios no es registrada en ningún momento.

Una transacción es un movimiento de bitcoins de un grupo de cuentas a otro. Se extrae valor de una o más direcciones de entrada, y se deposita este valor repartido en cantidades arbitrarias entre una o más direcciones de salida. Dada esta potencial multiplicidad de las direcciones tanto de entrada como de salida, no es posible en la mayoría de los casos determinar a qué dirección de salida fue el valor aportado por una dirección de entrada, sino que solo se puede observar que una transacción recibió determinadas cantidades de bitcoin de un conjunto de direcciones de entrada, y la suma de esos valores fue repartida de determinada forma entre elas direcciones de salida.

El primer modelado de la red realizado identifica dos tipos de vértices: direcciones y transacciones, y los une con aristas ponderadas etiquetadas como de entrada o salida según el papel que haya tomado cada dirección en la transacción, donde su peso es directamente la cantidad de bitcoins que transfirieron.

Los trabajos previos mencionados anteriormente utilizaban la totalidad del registro histórico de transacciones para la construcción del grafo que analizaban, en Abril 2013 Meiklejohn et al. [17] reporta haber importado aproximadamente 16 millones de transacciones y 12 millones de direcciones diferentes. Desde entonces el uso de Bitcoin se ha extendido y popularizado mundialmente, y el tamaño de la red se ha disparado de forma acorde. En Mayo del 2017, un proyecto público presentado en GitHub para convertir el registro histórico a una base de datos orientada a grafos Neo4j [10] reporta la conversión genera un grafo de 1.500 millones de nodos y 2.500 millones de relaciones entre ellos. Operar con este gigantesco volumen de datos requiere una infraestructura de gran escala que excede las posibilidades de este trabajo, teniendo en consideración además que los algoritmos de análisis en grafos tienen generalmente un orden cuadrático en su complejidad.

Un usuario puede realizar una transacción Bitcoin en cualquier momento, pero esta no tendrá efecto inmediato sino que deberá esperar a ser procesada por los nodos mineros y ser añadida en un bloque al registro distribuido. Cada bloque contiene una cantidad variable de transacciones, es generado cada un intervalo de tiempo también variable, y es la unidad mínima de tiempo de la red, ya que no existen certezas en cuanto al momento en una transacción fue enviada para su aprobación (distintos nodos observarán distintos tiempos) pero sí del instante en que fue aprobado un nuevo bloque.

Para el análisis presentado en este trabajo se consideran las transacciones ingresadas dentro del rango de bloques que va del 481935 al 483735, correspondientes al período de tiempo del 25 de Agosto al 5 de Septiembre del 2017. En la gráfica II se muestra el histograma de frecuencias de la cantidad de direcciones y transacciones de los bloques dentro de este rango. Este rango se dividió en intervalos de 200 bloques con el fin de reducir el tamaño de los grafos a analizar y lograr mayor estabilidad en el comportamiento de cada intervalo analizado. Esta agrupación de los bloques logra menor varianza en la cantidad de transacciones y direcciones presentes en cada uno. En II se puede ver cómo fue variando el precio del bitcoin en dólares americanos, según datos extraídos de la casa de cambio CoinBase, publicados en [11].

Analizar directamente el grafo de transacciones y direcciones presenta la dificultad de no poder extraer conclusiones del comportamiento de cada dirección ya que por lo general, un usuario maneja múltiples direcciones y usa una misma para solo unas pocas transacciones antes de transferir el dinero a otra, también de su pertenencia. Un mejor análisis de perfiles de uso sería logrado si se pudiera, idealmente, conocer con certeza que direcciones pertenecen a un mismo

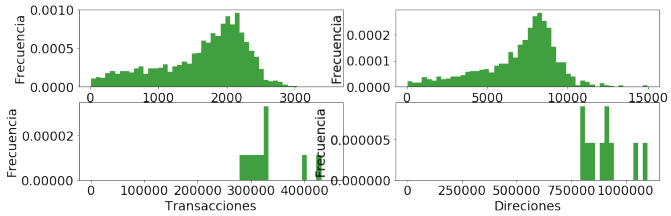

Fig. 1. Frecuencia de cantidad de transacciones y direcciones

Los histogramas superiores corresponden a las frecuencias de distribución de la cantidad de transacciones y direcciones presentes en cada bloque. Los de abajo muestran como se reduce la varianza (también la cantidad) si se agrupan en intervalos de 200 bloques.

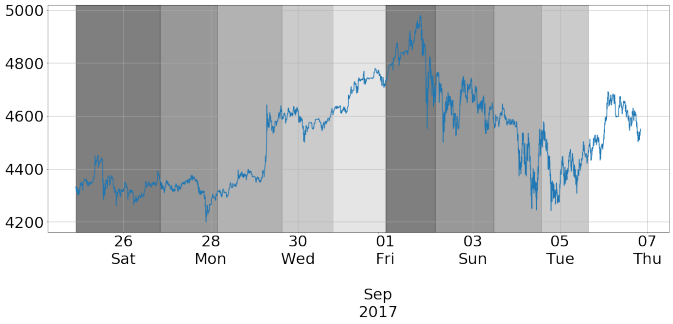

Fig. 2. Cotización BTC/USD durante el rango estudiado

Las áreas grises corresponden a los diferentes intervalos de 200 bloques en los que se dividió el rango.

usuario. Como fue antes dicho, por la naturaleza pseudo-anónima de Bitcoin no es posible determinar la autoría de las direcciones, aunque realizando ciertas asunciones se puede asociar direcciones para las que se cree con cierto grado de certeza que pertenecen al mismo usuario. La heurística presentada en [14] asume que si dos o más direcciones son entradas de una misma transacción, entonces son controladas por el mismo usuario. Sus efectos son transitivos y se expanden más allá de una sola transacción; e.g, si se observa una transacción con direcciones A y B como entradas, y otra con entradas B y C, entonces se concluye que las tres corresponden al mismo usuario. Esta asunción se basa en que las transacciones compartidas son poco frecuentas en Bitcoin.

La heurística de asociación fue aplicada al grafo por medio de una rutina que exploraba todas las transacciones y definía todas las direcciones de entrada como pertenecientes a una misma clase de equivalencia. Los nodos de las múltiples direcciones de una misma clase fueron sustituidos por un solo nodo de un nuevo tipo: ”Wallet”, que representa un conjunto de direcciones controladas por un mismo usuario.

Adicionalmente, se clasificaron las Wallets identificadas en cuatro categorías según el rol del usuario que las controla: ”Mineros”, Äpuestas”para direcciones pertenecientes a sitios de apuestas, Ëxchanges”para las direcciones utilizadas por

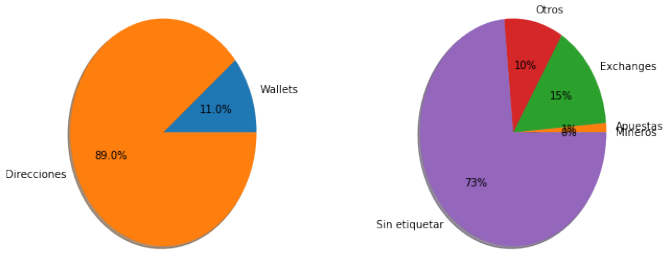

Fig. 3. Resultados de aplicar la heurística y etiquetado

La primer gráfica muestra la proporcionalidad entre nodos Direcciones y Wallets existentes en el nuevo grafo obtenido tras aplicar la heurística, la segunda muestra la cantidad de Wallets etiquetadas.

las casas de cambio (en adelante referidas como exchanges) y Ötros”, donde se agrupan direcciones de diversa índole como tiendas online, servicios, etc. Esta información fue extraía del sitio Wallet-Explorer, que asocia direcciones siguiendo la misma heurística utilizada en este trabajo y adicional ente las clasifica basándose en información extraída de Internet (foros, sitios web que exponen directamente su dirección, etc).

Al aplicar la heurística en cada grafo intervalo se logró reducir en promedio 37 % la cantidad de nodos Direcciones (o nodos Direcciones/Wallets luego de su aplicación). Pese a esto, la proporción de Wallets frente a Direcciones es de solo 10 %, como se puede ver en la gráfica II, y el etiquetado tampoco logró aportar información sobre el 75 % de las Wallets. Si bien se pudo reducir considerablemente el tamaño del grafo, aún se tiene una mayoría de entidades para las que no se tiene información sobre su perfil de uso.

Los grafos reducidos y etiquetados de cada intervalo fueron unidos nuevamente y cargados en una base de datos orientada a grafos Neo4J, sobre la cual se realizará el análisis.

III. Análisis de los datos

La primer característica a examinar de la red de transacciones es su tendencia a agruparse en componentes conectados. Si las transacciones se dan entre grupos reducidos de usuarios que solo interactúan con unos pocos, entonces el grafo de la red tendrá muchas componentes conexas aisladas entre si, y el valor de lo que pueda aportar el análisis de la red sobre el funcionamiento del sistema será limitado.

Para medir esta conectividad de la red, se ejecutó el algoritmo Union Find, que encuentra en un grafo cuales son los conjuntos de nodos para los que cada nodo es alcanzable por cualquier otro dentro de su mismo conjunto. Se trató el grafo como no dirigido, por lo que una Dirección o Wallet se considera conectada con una Transacción igualmente sea una salida o una entrada. Para evaluar la evolución de esta conectividad con respecto al tiempo, se la calculó cada 100 bloques dentro del rango estudiado, para cada uno se limitaba la red a las transacciones y direcciones aparecidas

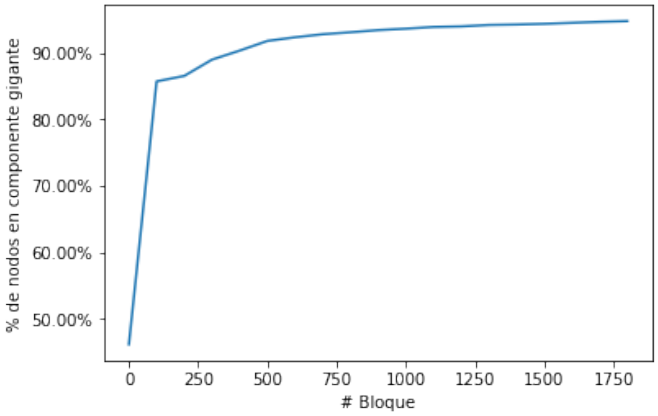

Fig. 4. Evolución del tamaño de la componente gigante

Se grafica el crecimiento del tamaño relativo del componente gigante a medida que se consideran las transacciones de más bloques en la conectividad.

en los bloques anteriores. Así, se encuentran primero las componentes conexas del primer bloque, luego de todos los bloques hasta el 100, de todos los bloques hasta el 200 y así sucesivamente hasta cubrir todo el rango. En la gráfica III se muestra como el tamaño del componente gigante relativo al tamaño de la red es creciente cuanto más bloques se tengan en consideración.

III-A. Análisis de exchanges

Dentro los cuatro tipos de usuarios que distinguimos en la clasificación realizada, los Wallets etiquetadas como exchanges se distinguen tanto por su cantidad como por su importancia. Estas corresponden a direcciones bitcoin utilizadas por sitios en los que usuarios compran y venden bitcoins por otras monedas, ya sean digitales (Ethercoin, Monero) o tradicionales (Dólares americanos, Euros). El exchange oficia como un intermediario entre los usuarios que realicen el intercambio, recibiendo los bitcoins de uno, la moneda ofrecida en contrapartida del otro, y cuando se asegura que ambas transacciones sean válidas, transfiere a cada uno el valor que intercambió. De esta forma, confiando en el exchange, los usuarios tiene la garantía de que el intercambio será exitoso ya que el exchange retendía los fondos hasta que ambos hayan enviado su parte correspondiente, y los devolverá si la otra parte falla.

En la gráfica III-A se muestra la relación que existe entre el volumen total de transacciones dentro de cada bloque y el de aquellas que corresponden a nodos etiquetados como exchanges. Si bien la proporcionalidad no es constante, ambas series mantienen cierta correlación, muchas veces alcanzando máximos o mínimos en entornos cercanos. Al comparar la cantidad de bitcoins enviados desde exchanges con los recibidos (gráfica III-A), se encuentra que ambas coinciden o se encuentran en entornos muy cercanos la mayor parte del tiempo, confirmando que estos solo actúan

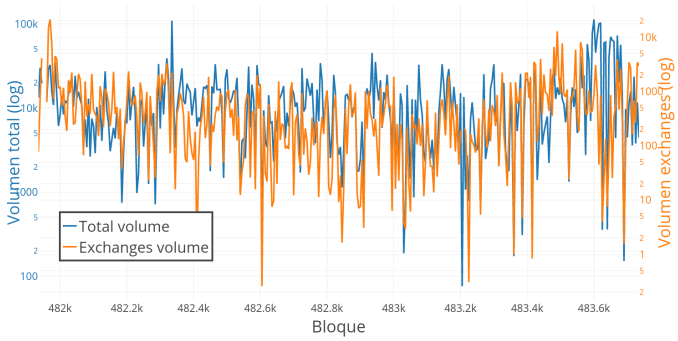

Fig. 5. Transacciones de exchanges con respecto al total

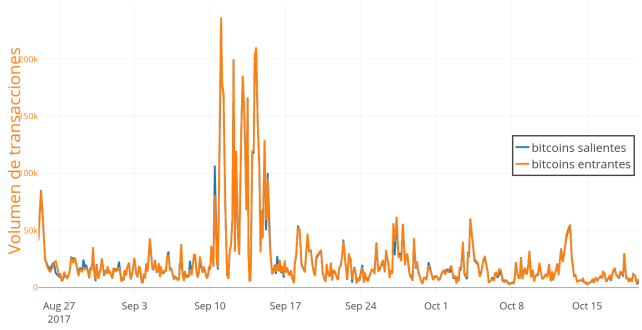

Fig. 6. Transacciones entrantes y salientes de exchanges

como intermediarios reteniendo solo el monto correspondiente a las tarifas que cobran por su servicio.

Con el fin de tener otra métrica de la importancia de los nodos clasificados como exchanges además del volumen de sus transacciones, evaluamos su centralidad. La centralidad es un concepto abstracto con el cual se busca poder capturar la importancia de un nodo en su grafo. Mientras que medir el volumen de transacciones es un atributo intrínseco de los nodos, la centralidad de estos dependerá de su localización en la red, i.e de dónde procede este volumen de transacciones y con qué frecuencia un usuario se relaciona con estos exchanges. Si la compra y venta de bitcoins a cambio de otras monedas es una actividad frecuente de los usuarios, entonces la centralidad de los exchanges será alta; si por el contrario solo un reducido grupo la realiza, entonces su centralidad será baja.

La centralidad es un concepto abstracto en la medida de que depende de la fijación de un criterio que cuantifique la importancia de un nodo con respecto a la red. Existen tres criterios que son ampliamente utilizadas en análisis de redes: el de cercanía (c̈loseness centrality”) que asocia la centralidad de un vértice a su cercanía al resto de nodos en la red, la de intermediación (”betweeness centrality”) que cuantifica la frecuencia con la que un nodo actúa como intermediario en el camino mas corto entre los otros nodos

de la red , y la de vector propio (ëigenvector centrality") que entiende la centralidad como ser accedido por nodos importantes.

Tanto la centralidad de intermediación como la de cercanía serían las indicadas para obtener una medida de la importancia de los exchanges en la red. Si la actividad de intercambio es frecuente, entonces sucederá que tanto la distancia de los exchanges a otros nodos será reducida (habrán intercambiado o se habrán relacionado con alguien que lo haya hecho), así como también el camino mas corto entre dos nodos de la red será frecuentemente a través de un exchange. Calcular cualquiera de estas dos medidas implica calcular el camino más corto entre todos los nodos de la red, lo cual tiene complejidad $\Theta(|V^3|)$ aplicando el algoritmo de Floyd-Warshall, o más eficientemente en un grafo disperso como el de la red Bitcoin $\Theta(|V^2 log|V| + |V||E|)$ utilizando el algoritmo propuesto por Brandes [15]. Dado el tamaño del grafo de transacciones de Bitcoin, aún reducido al rango de bloques estudiado, la aplicación de un algoritmo de este órden es inviable.

Page Rank es un método propuesto por Sergey Brin y Larry Page [19] para computar un ranking de importancia entre los nodos de la red. Basado en paseos aleatorios, la medida centralidad que Page Rank asigna a cada nodo corresponde a la probabilidad de que sea visitado en una exploración aleatoria de la red. Page Rank es un método iterativo, por lo que su tiempo de ejecución puede ser acotado tanto como se quiera.

La implementación de Page Rank provista por la libería de algoritmos Neo4J [13] fue aplicada en el grafo, los diez primeros nodos ordenados por su ranking son listados en la tabla I. Dos nodos sobresalen por su importancia, los correspondientes a los exchanges Bittrex.com y HaoBTC.com. Los otros nodos presentan un ranking considerablemente mas bajo, pero no corresponde a direcciones identificadas como exchanges. Esas wallets tampoco se distinguen por agrupar una gran cantidad de direcciones, pero al analizar sus transacciones se encontró que todas habían realizado un gran número de movimientos que compartían el mismo patrón: sacaban bitcoins de la dirección, una parte eran enviados a otra cuenta pero otra volvían a ser enviados a la misma dirección. Estas transacciones tenían entrada y salida en la misma dirección, creando lazos a sí misma que incrementan su puntaje en el Page Rank y explican el alto ranking logrado por estas direcciones.

Vista la centralidad de los exchanges en la red y la relación entre sus transacciones y las transacciones totales de la red, puede considerarse que su actividad representa un indicador de la actividad del totalidad de la red. Consideramos así que toda tendencia en el comportamiento de la red tendrá su consecuencia en el comportamiento de los exchanges y lo mismo vale para su recíproco. Esto, si bien no deja de ser una asunción que debería ser analizada en más profundidad, permite disponer de un punto de partida para la investigación del funcionamiento de la red, ya que en lugar de analizar la totalidad de la red, lo haremos desde la relación de ésta con los nodos exchanges, acotando así el volumen de datos.

TABLE I
NODOS CON MÁS PAGERANK

Rank	Nodo	Cant. direcciones	Vol. Transacciones
15712	Bittrex.com	40070	191710
13199	HaoBTC.com	1630	353408
5022	No reconocido	5	103517
4857	No reconocido	1	69828
4764	No reconocido	3	48674
4343	No reconocido	2	109524
4037	No reconocido	12	43426
3952	No reconocido	4	72840
3706	No reconocido	2010	44589
2792	No reconocido	19	32552

III-B. Interacciones de la red con los exchanges

Nos interesa entonces lograr una visión de la actividad de la red que permita dilucidar cuales son los usos que hoy en día sus usuarios le están dando a Bitcoin. Ya que el valor de un bitcoin esta dado solamente por las reglas de oferta y demanda en su compra y venta, sin existir ninguna institución que respalde su valor, entonces el constante crecimiento de su cotización debería de estar relacionado a crecimiento de su demanda (usuarios que quieren comprar bitcoins) que no es correspondido por un correspondiente crecimiento en la oferta (usuarios que venden sus bitcoins). Una moneda sea digital o tradicional solo tiene valor en tanto sea aceptada como medio de intercambio en la compra y venta de activos y servicios, por lo que esta desproporcionalidad entre oferta y demanda puede explicarse por dos razones: o los usuarios que disponen de bitcoin no ven necesidad de pasar su valor a otras monedas, ya que lo utilizan como moneda para su actividad económica regular, o simplemente lo acumulan porque confían en que su valor seguirá aumentando, y encontrarán en el futuro un mejor momento para venderlo.

En esta sección se propone una forma de cuantificación para el segundo escenario en que los usuarios estań principalmente adquiriendo bitcoins para esperar venderlos a mayor precio en el futuro. Si esto sucede, se debería encontrar una gran cantidad de direcciones para las cuales su única acción en la red fue adquirir bitcoins directamente de un exchange y o bien los vendieron posteriormente o bien los siguen manteniendo esperando un mejor momento para venderlos.

Se busca una métrica que exprese una magnitud de la probabilidad que tiene un bitcoin que fue adquirido por un usuario a través de un exchange de alejarse una distancia determinada de él antes de volver a ser intercambiado nuevamente. Si la compra y venta de bitcoins es una actividad esporádica que solo es realizada por un usuario para adquirir su primer capital, entonces este valor debería ser alejado, indicando que lo más probable es que el bitcoin fluya entre usuarios bastantes veces antes de volver a ser vendido en un exchange. Si por el contrario, en un escenario extremo los usuarios solo adquieren bitcoin para venderlo posteriormente, entonces no habrá transacciones entre usuarios y el bitcoin volverá inmediatamente a un exchange con probabilidad uno.

Se recurre entonces a una caminata aletaria (random walk") que modele los recorridos posibles de un bitcoin

saliente de un exchange. La caminata comienza en un nodo exchange escogido al azar, y en cada paso el siguiente nodo a visitar se sortea entre todos los vecinos posibles, asignándole a cada uno una probabilidad de ser visitado proporcional a la cantidad de bitcoins que se le envió. Intuitivamente, este paseo aleatorio modela el recorrido de un bitcoin a través de la red, con la salvedad de que no es exactamente un bitcoin en particular al que se está siguiendo, ya que las transacciones generalmente mueven fracciones de unidad.

El uso de una caminata aleatoria permite ponderar el camino a seguir por el flujo de dinero que haya ido a esa dirección, dando mas importancia a los grandes flujos de dinero, los de mas incidencia en la economía de la red. No se obtiene una medida de la actividad de los usuarios en la red, sino de la actividad de los bitcoins en circulación. Una observación más correcta si se toma en cuenta que los movimientos de más capital son los más influyentes. Por otra parte, la caminata aleatoria es un proceso iterativo, lo que permite más control sobre el tiempo de ejecución, particularmente de interés cuando se trabaja con un gran volumen de datos como en este caso.

Data: Grafo de transacciones-direcciones Bitcoin
Result: Largo del camino p que se recorrió hasta encontrar otro exchange
Seleccionar un exchange al azar;
$\mathbf{largo\_camino\_recorrido} := 1$;
while *No se haya llegado a otro exchange* **do**
  Seleccionar una transacción saliente t, con probabilidad proporcional a la cantidad de bitcoins transferidos en cada transacción.;
  Seleccionar una dirección de salida de la transacción t, con probabilidad proporcional a la cantidad de bitcoins transferidos a cada dirección de salida. ;
  $\mathbf{largo\_camino\_recorrido} + +$;
end

Algorithm 1: Algoritmo de caminata aleatoria

El Algoritmo 1 muestra el pseudocódigo correspondiente a una instancia de la caminata aleatoria. Se realizaron 2000 iteraciones y se graficaron histogramas que muestran la frecuencia de ocurrencia de un camino para los distintos largos registrados. En muchos casos, la caminata aleatoria llega a un fin forzado al encontrar un nodo para el que no existían transacciones de salida. En estos casos la caminata es interrumpida, la distancia recorrida registrada y su histograma se muestra en el histograma izquierdo de la Figura III-B. La distancia recorrida por las caminatas que haya logrado arribar nuevamente a un exchange se muestran en el histograma derecho de la Figura III-B.

En el histograma se ve que, para el caso de las caminatas que se interrumpen antes de llegar a un exchange, la distancias mas cortas son las más frecuentes, por lo que hay una gran cantidad de bitcoins que son adquiridos en un exchange pero no fueron gastados en el período de tiempo considerado. Para el caso de las caminatas que sí logran llegar de vuelta a un exchange, lo más frecuente es llegar en un camino

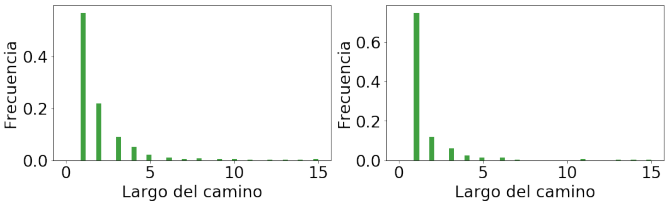

Fig. 7. Distribución de frecuencia de las caminatas aleatorias

El histograma de la izquierda muestra el largo del camino máximo alcanzada por las caminatas que no pudieron acabar llegando a un exchange. En la derecha la distribución del largo de los caminos para las caminatas que llegaron nuevamente a un exchange.

de largo dos, por lo que una gran cantidad de bitcoins son adquiridos en exchanges e inmediatamente vendidos nuevamente. Los datos son el resultado de 2000 caminatas aleatorias, de las que solo el 20 % lograba encontrar un camino de vuelta al exchange.

Los resultados podrían sugerir que una gran cantidad de bitcoins son adquiridos para luego solo permanecer inactivos, y que de los que son adquiridos y movilizados, la mayor cantidad lo hace simplemente para volver a ser intercambiado nuevamente. Esto podría corresponderse con una actividad especulativa en el que algunos usuarios adquieren bitcoins confiando en que su valor incremente a futuro, mientras que la estrategia de otros es tomar una posición mas activa y comprar o vender según los cambios en cotización en una escala de tiempo menor. De todos modos, el alcance de este trabajo es muy acotado, por el rango de días considerado y por la cantidad de caminatas aleatorias realizadas, por lo que se carece de suficiente fundamento para sacar semejantes conclusiones.

IV. Conclusiones

El libre acceso a los datos de todas las transacciones en los sistemas de dinero digital emergente ofrece la oportunidad de estudiar estas economías desde la perspectiva del análisis de redes. Hasta el momento, la información financiera era de acotado alcance y difícil acceso, su análisis era basado en índices y cifras basadas en estimaciones estadísticas. Si bien la digitalización de las transacciones es cada vez mayor y abarca un gran porcentaje incluso de la actividad doméstica de pequeña escala, esta información es protegida por razones de privacidad.

Existe sin embargo un ámbito en donde la información sobre la actividad financiera es abundante, y es en su mismo centro: los mercados bursátiles donde se compran y venden acciones y bonos. En ellos la digitalización es total y la disponibilidad de datos alta, existiendo registros de la evolución segundo a segundo de los precios y las órdenes de compra y venta existentes. Esta abundancia de datos ha sido el motor de una amplia investigación en finanzas, resultado de una competencia entre actores del mercado por lograr explotar de una forma más óptima la información disponible

para maximizar su ganancia. En Bitcoin y cualquier sistema de dinero digital distribuido donde la información es total, todo movimiento de capital existido está registrado, las posibilidades de análisis son ilimitadas. De extenderse el uso de este tipo de sistemas, sea Bitcoin o cualquier otro sucesor en algún futuro, es de esperar que esta abundancia de información sea el insumo para una nueva explosión en la investigación y generación de conocimiento en las finanzas.

En este trabajo se propuso un acercamiento a las posibilidades que el análisis de la red Bitcoin ofrece. Particularmente, se intentó buscar una forma de medir la actividad de la red que brinde información sobre el uso que sus usuarios le están dando actualmente, y poder distinguir si la popularización de bitcoin y la suba de su precio se corresponde a un incremento de la actividad entre sus usuarios, haciendo que el pago en bitcoins sea cada vez mas utilizado como un medio de intercambio, o si por el contrario la actividad entre usuarios es nula y estos solo adquieren bitcoins con la intención de venderlos posteriormente a un mayor precio.

La investigación realizada encontró que su principal motivación, la abundancia de información en Bitcoin, también era su principal impedimento. El volumen de datos disponible es tan grande que se tuvo que limitar a analizar un pequeño período de tiempo, limitando así las conclusiones que se pueden extraer. El rango de diez días estudiado no es representativo y no permite extrapolar de él una conclusión sobre la actividad que la red está teniendo en por ejemplo, el último año.

Aún trabajando dentro de este rango, el volumen de datos es enorme, y la generación del modelo, realizándola en un computador personal, conlleva días y dificulta mucho la agilidad en los ciclos de prueba y corrección de errores detectados. El alcance para este trabajo que se tenía en la planificación inicial tuvo que ser considerablemente reducido ante estas dificultades, y de hecho solo un 20 % del tiempo invertido en este trabajo pudo ser destinado al análisis una vez finalizado el modelado y obtenido el grafo de transacciones depurado.

Aún con estas limitaciones, se demostró que un análisis de la red es posible, y que determinadas características de la red se corresponden a aspectos reales, como por ejemplo que los nodos con más centralidad en la red correspondan a exchanges. El modelado utilizado utiliza una heurística ya propuesta para asociar direcciones que con alta probabilidad pertenecen al mismo usuario y reducir así la cantidad de nodos de la red sin perder significado. El uso de información externa permite reducir aún más esta anonimidad e identificar a los nodos exchanges. Este proceso de generación del modelo ha probado ser correcto y útil, y si bien en este trabajo se realizó para un pequeño período, contando con una infraestrucutra adecuada se podrían acelerar los tiempos usando una mayor capacidad de cómputo.

La anonimidad en la red, pese a ser reducida con las heurísticas utilizadas, imposibilita en análisis de la actividad en términos de usuarios. Un mismo usuario puede potencialmente poseer miles de direcciones diferentes sin que sea posible asociarlas entre sí, y como consecuencia para una transacción no se puede determinar si corresponde efectivamente a una transferencia de dinero entre diferentes usuarios o si simplemente es un ajuste de balances entre direcciones del mismo propietario. Analizar la red en términos de la actividad de usuarios implica una alta incertidumbre, que en este trabajo se propuso evitar analizando en su lugar los flujos de dinero. Para este enfoque adoptado, la pertenencia de diferentes cuentas a un mismo usuario no es un problema a considerar, a la vez que atribuye más importancia a movimientos de grandes capitales que a los de pequeño capital.

Se propuso analizar los flujos de bitcoins utilizando una caminata aleatoria que recorra el grafo de transacciones, eligiendo en cada nodo un sucesor según una distribución de probabilidad proporcional al dinero enviado a cada dirección. Realizando esta caminata repetidas veces, se puede extraer una distribución de probabilidad de sus recorridos más frecuentes, que significará que una mayor cantidad de capital se mueve en esa dirección. Al ser un método iterativo, permite acotar su tiempo de ejecución y paralelizar un gran número de instancias simultáneamente.

Esta caminata aleatoria fue utilizada para evaluar la distancia a la que los bitcoins vendidos en un exchanges tienden a alejarse antes de volver a ser vendidos nuevamente en otro exchange. Se encontró que lo mas frecuente para los bitcoins adquiridos en exchanges es no ser movilizados nuevamente, y que los que lo hacen y vuelven eventualmente a un exchange lo hacen en una distancia corta, frecuentemente en la siguiente transacción. Esto podría reflejar que los usuarios están utilizando bitcoin principalmente con fines especulativos y que no hay una actividad económica real en la que las transacciones entre usuarios sean frecuentes. No obstante, por todas las razones antes mencionadas, el alcance de este trabajo es limitado y se debería de estudiar el comportamiento en más profundidad, analizando un período de tiempo mayor antes de poder extraer conclusiones sostenibles.

References

[1]

[2] Cnbc - "bitcoin is in a bubble, and here's how it's going to crash". https://www.cnbc.com/2017/09/13/bitcoin-is-in-a-bubble-and-heres-how-its-going-to-crash-ron-insana.html. Accedida: 2017-15-09.

[3] Cnc - "bitcoin falls below $3000 as chinese regulator orders cryptocurrency exchanges to shut shop". http://www.moneycontrol.com/news/world/bitcoin-falls-below-3000-as-chinese-regulator-orders-cryptocurrency-exchanges-to-shut-shop-2389121.html. Accedida: 2017-15-09.

[4] Cnc - "jpmorgan ceo jamie dimon says bitcoin is a 'fraud' that will eventually blow up". https://www.cnbc.com/2017/09/12/jpmorgan-ceo-jamie-dimon-raises-flag-on-trading-revenue-sees-20-percent-fall-for-the-third-quarter.html. Accedida: 2017-15-09.

[5] Cnn - "bitcoin bubble may have burst". http://money.cnn.com/2013/04/12/investing/bitcoin-bubble/index.html. Accedida: 2017-15-09.

[6] Coindesk - "$4,880: Bitcoin price climbs to another all-time high". https://www.coindesk.com/4880-bitcoin-price-climbs-to-another-all-time-high/. Accedida: 2017-15-09.

[7] Cointelegraph - "bitcoin reaches all time high daily trading volume at $5 bln". https://cointelegraph.com/news/bitcoin-reaches-all-time-high-daily-trading-volume-at-5-bln. Accedida: 2017-15-09.

[8] Cotización promedio de bitcoin en usd. https://www.coindesk.com/price/. Accedida: 2017-15-09.

[9] Cryptofrance - ”pourquoi une prédiction à 1 million de dollars par bitcoin n’est pas si farfelue”. https://www.crypto-france.com/pourquoi-une-prediction-de-1-million-de-dollars-par-bitcoin-nest-pas-si-farfelue/. Accedida: 2017-15-09.

[10] Github: Bitcoin to neo4j. https://github.com/in3rsha/bitcoin-to-neo4j. Accedida: 2017-22-11.

[11] Kaggle: Bitcoin historical data. https://www.kaggle.com/mczielinski/bitcoin-historical-data. Accedida: 2017-10-10.

[12] Rt - ”bitcoin value passes $1,000 for first time ever”. https://www.rt.com/news/bitcoin-value-thousand-dollars-392/. Accedida: 2017-15-09.

[13] Neo4j graph algorithms, 2017.

[14] Elli Androulaki, Ghassan O. Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. *Evaluating User Privacy in Bitcoin*, pages 34–51. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[15] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25:163–177, 2001.

[16] Dániel Kondor, Márton Pósfai, István Csabai, and Gábor Vattay. Do the rich get richer? an empirical analysis of the bitcoin transaction network. *CoRR*, abs/1308.3892, 2013.

[17] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. *Commun. ACM*, 59(4):86–93, 2016.

[18] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system,” http://bitcoin.org/bitcoin.pdf.

[19] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[20] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. *IACR Cryptology ePrint Archive*, 2012:584, 2012.