利用“多阶朋友”的比特币交易记录去匿名化

崔家铭(515030910359)

1. 背景介绍

比特币(Bitcoin)被部分观点认为是一种去中心化，非普遍全球可支付的电子加密货币，而多数国家则认为比特币属于虚拟商品，并非货币。比特币由中本聪(Satoshi Nakamoto)于2009年1月3日，基于无国界的对等网络，用共识主动性开源软件发明创立。任何人皆可参与比特币活动，可以通过称为挖矿的电脑运算来发行。比特币协议数量上限为2100万个，以避免通货膨胀问题。使用比特币是通过私钥作为数字签名，允许个人直接支付给他人，不需经过如银行、清算中心、证券商等第三方机构，从而避免了高手续费、繁琐流程以及受监管性的问题，任何用户只要拥有可连接互联网的数字设备皆可使用。2017年8月1日由于对区块容量大小的争议，比特币进行分叉，新分支名为Bitcoin Cash。两个分叉当中Bitcoin Cash 主张在区块链上进行扩容并保持无中介支付功能，Bitcoin 则主张应该创建有第三方的第二层网络。

区块链(blockchain)是用分布式数据库识别、传播和记载信息的智能化对等网络，也称为价值互联网。中本聪在2008年，于《比特币白皮书》中提出“区块链”概念，并在2009年创立了比特币社会网络，开发出第一个区块，即“创世区块”。区块链共享价值体系首先被众多的加密货币效仿，并在工作量证明上和算法上进行了改进，如采用权益证明和SCrypt算法。随后，区块链生态系统在全球不断进化，出现了首次代币发售ICO、智能合约区块链以太坊、“轻所有权、重使用权”的资产代币化共享经济以及区块链国家。目前，人们正在利用这一共享价值体系，在各行各业开发去中心化电脑程序(Decentralized applications, Dapp)，在全球各地构建去中心化自主组织和去中心化自主社区(Decentralized autonomous society, DAS)。

2. “多阶朋友”及其概念

在本文中，我们这里所指的“多阶朋友”，实质上是引述了在社交网络中常见的朋友概念，并在社交网络中将其延展而成。在社交网络中，我们通常将两个节点或者两个用户之间建立的单向或双向的长期信息交流关系称之为朋友关系，而所谓的二阶朋友，则是指当两个用户之间共同拥有着某一个朋友，从而使得两者之间虽然没有直接的朋友关系，但仍然可以实现信息或观点的间接交换。显然的，二阶朋友的关系较之正常的一阶朋友而言，其强度和效力都弱了很多。但考虑到通常而言在一个社交群体中，某个节点或用户的朋友数量实际上是比较有限的条件下时，这一概念可以有效的将节点更加充分的与网络中的各个元素建立关系，从而可以进行更为深入的分析。而类似的，如果我们将这种二阶的关系进一步的扩展，那么我们就可以得到三阶乃至更高阶的朋友概念。在本文中，出于研究的方便起见，我们将主要将注意力，集中在二阶朋友上。从而在不失一般性的条件下，对多阶朋友在区块链去匿名化的的作用加以分析。

3. 比特币交易现状

到目前为止，绝大部分的比特币交易都是处于一种商业化的运行模式下，特别是随着各个比特币交易平台的出现，和比特币与显示货币的价值挂钩与兑换，比特币交易市场，已经成为了一个有着大量高频交易的市场。在2017年，比特币的比值节节攀升，而在2018年年初，其币值也出现了50%以上的贬值，其跌幅甚至远高于普通的股票或债券、期货市场。如图1所示，其为2018.5.27时比特币的比值以及其在2015年6月之后至今的币值变化：

图 1 比特币币值变化

4. 比特币交易记录的去匿名化

根据比特币的交易记录与模式，我们可以建立阐述比特币交易的地址变化的模型，并根据相关的经济学知识设定一定的规则并对交易记录进行去匿名化操作。

在这里，我们的交易信息来自于 https://blockchain.info/网站，如图所示，其可以提供 2013 年至今的全部交易记录。

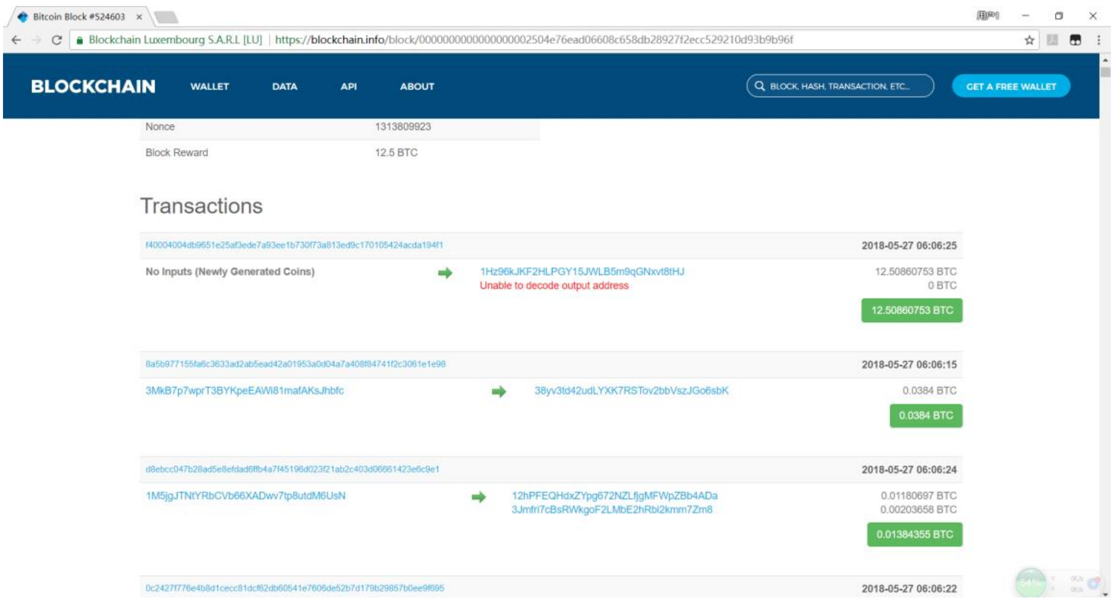

图 2 比特币交易记录

而相关的币值信息，我们这里的来源则是 https://dc-charts.com/raw_btc.php?ex=0&cu=0&tz=6&ar=1 网站，如图 3 所示。

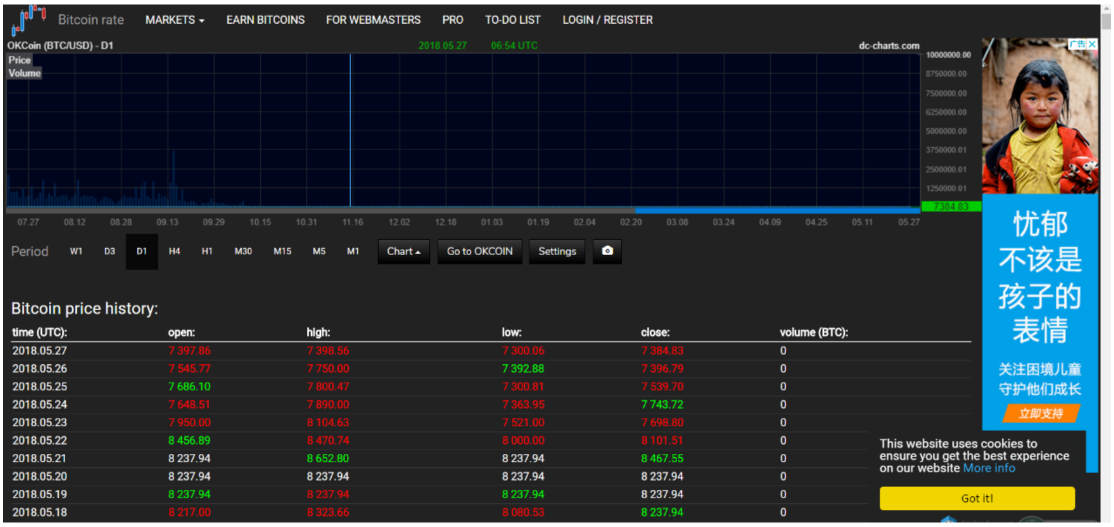

图 3 比特币币值记录

依此，我们找到了 2018.4.23 比特币当天的全部交易记录，共计 226201 笔交易，涉及 607424 个地址。

接下来，我们通过分析其相关的性质，建立了如下三条规则：

1. 若某笔交易的发起者中涉及的地址不止一个，那么这些地址将都被归类到同一个人名下。
2. 若对于某笔交易，我们发现其交易的发起者在交易前后所接收和发送的比特币币值在当时的汇率下非常接近，同时剩余的一部分币值差则被保留时，说明剩余的币值差的拥有者与该交易的发起者大概率是同一个人。
3. 对于上述的这些残差，若这些残差被收集起来再次汇入某一账户并进行交易，则说明这些残差的拥有者与最终收集这些残差的接收者大概率是同一个人。

对于上述的三条规则，我们分别给出如下的解释：

1. 由于对每一个地址下残存的资金进行操作时，需要得到该地址的所有者的授权，而考虑到某笔交易在进行过程中，这些地址均是同时得到其授权并完成相关比特币的转移的，因此我们可以认为，这些地址的所有者是同一个人，并且该所有者几乎同时授权这些地址进行比特币的交易。
2. 考虑到比特币市场是一个巨大的类金融交易市场，那么在这个市场中的每一笔交易，其存在则必然应具有其合理性。参考金融市场上的高频交易，我们有理由认为，这种操作实际上是一种利用瞬间的价格差进行短期套利的行为，因而我们可以认为剩余的币值差的拥有者与该交易的发起者大概率是同一个人。
3. 在这种情况下，剩余的残差由于币值过小或者价格不是非常整齐，因而并不能满足一手交易所要求的特定单位的价格，因此残差的所有者需要将这些残差汇集起来并进一步的进行下一步的交易。因而我们可以认为这些残差的拥有者与最终收集这些残差的接收者大概率是同一个人。

在建立了如下规则后，我们对上述的 2018.4.23 比特币当天的全部交易记录(226201 笔交易，607424 个地址)进行再次的分析，最终发现具有性质 2 的交易共有 84531 条，占总交易数的 37%。而最终，我们得到的结果中，将 607424 个地址归类到 267861 个节点上。而从节点的度数分布而言，可以看出，去匿名化之后的度数分布，较去匿名前的度数分布相比，更加的符合幂律分布的规则。

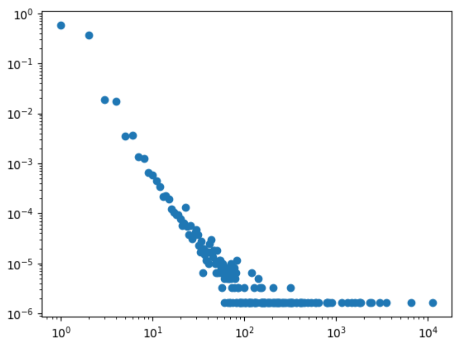

图 4 去匿名前的分布规律

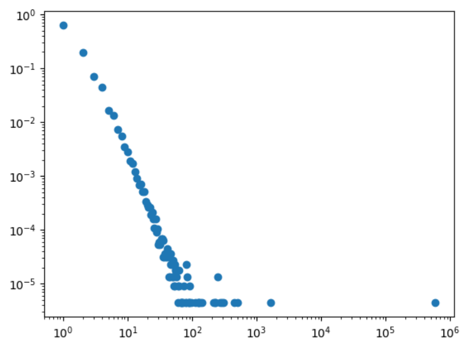

图 5 去匿名后的分布规律

5. 比特币节点度数分布的理论证明

在接下来的内容中，我们将建立比特币用户的度数分布的理论证明：
假设 k 为某交易者一天之间的交易数，则其当前的活跃交易数为α，当前时间窗口为 t，某交易赚钱的概率为 p，则有$P_{k,t}$为 t 时刻交易数为 k 的节点所占比例/某阶段交易数为 k 的概率。根据以上关系，我们有：

$$P_{k,t} = (1-p)^{\alpha k} P_{k,t-1} + [1-(1-p)^{\alpha(k-1)}] P_{k,t-1}$$

则有

$$\mathrm{P}[\text{交易数} \geq \mathrm{k}] = \frac{1}{t}\sum_{i=1}^{t} P_{k,i} = \frac{1}{t} S_{k,t}$$

考虑

$$P_{1,t} = (1-p)^{\alpha} P_{1,t-1}$$

且$P_{1,1} = 1$，则

$$S_{1,t} = \sum_{i=1}^{t} P_{1,i} = \frac{1-(1-p)^{t\alpha}}{1-(1-p)^{\alpha}}$$

$$S_1 = \lim_{t\to\infty} S_{1,t} = \frac{1}{1-(1-p)^{\alpha}}$$

考虑

$$S_{k,t} = (1-p)^{\alpha k} S_{k,t-1} + [1-(1-p)^{\alpha(k-1)}] S_{k,t-1}$$

有

$$S_k = (1-p)^{\alpha k} S_k + [1-(1-p)^{\alpha(k-1)}] S_k$$

则

$$\frac{S_k}{S_{k-1}} = \frac{1-(1-p)^{\alpha(k-1)}}{1-(1-p)^{\alpha k}}$$

即

$$S_k = \frac{1}{1-(1-p)^{\alpha k}}$$

即

$$\mathrm{P}[\text{交易数} \geq \mathrm{k}] = \frac{1}{t[1-(1-p)^{\alpha k}]}$$

利用 Taylor 公式展开，有

$$\mathrm{P}\left[\text{交易数} \geq \mathrm{k}\right] = \frac{1}{k(1-p)^{\alpha}t} + \frac{k-1}{2kt} + \mathrm{o}\left(\frac{(1-p)^{\alpha}}{t}\right)$$

考虑幂律分布的特征，我们实际上已经得到了

$$\mathrm{P}\left[\text{交易数} = \mathrm{k}\right] \sim power\ \ law\ \ distribution$$

的结论。

而对于存有匿名化的情况下，类似的，我们有

$$qP_{k,t} = q(1-p)^{\alpha k}P_{k,t-1} + \left[1-(1-p)^{\alpha(k-1)}\right]P_{k,t-1}$$

$$\mathrm{P}\left[\text{交易数} \geq \mathrm{k}\right] = \frac{q^{s-k}}{k(1-p)^{\alpha}t} + \frac{k-1}{2kt} + \mathrm{o}\left(\frac{(1-p)^{\alpha}}{t}\right)$$

其中 q 为用户选择更换地址的概率

注意到式中的非线性因子q^{s-k}，则应当有

$$\ln\frac{\mathrm{P}\left[\text{交易数} \geq \mathrm{k}，\text{匿名}\right]}{\mathrm{P}\left[\text{交易数} \geq \mathrm{k}，\text{非匿名}\right]} \sim (s-k)\ln q$$

即两者之比应与 k 成线性关系。

而实验结果如图 6 所示，也验证了这一点：

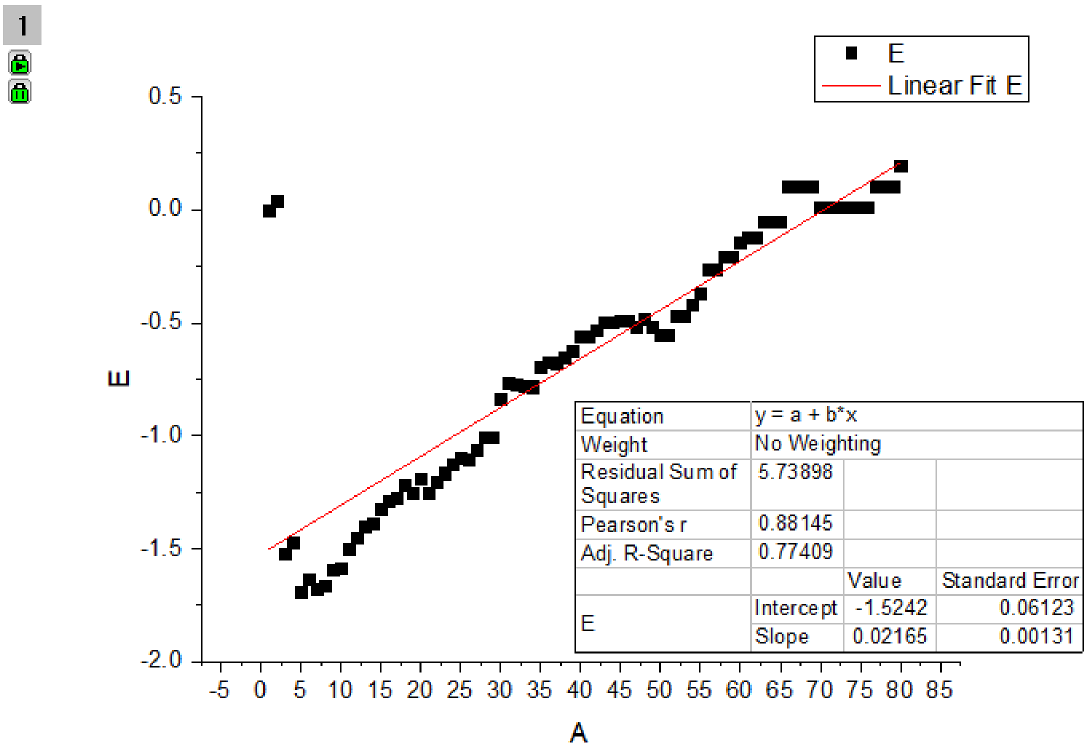

图 6 非线性因子的拟合

6. 结论

在本文中，我们通过利用二阶朋友的模型，并同时考虑到比特币交易市场所具有的特性，建立了一系列规则，从而对比特币的交易信息进行了相应的归类。此外，我们还从理论上说明了在理想情况下，比特币交易用户的度数应服从与幂律分布并给出了证明，而在匿名的条件下，我们则得出相关的地址更换使得幂律分布的结果中出现了一部分的非线性因数，并给

出了实验上的验证。