文章编号：2096-1472(2023)-04-57-06
DOI:10.19644/j.cnki.issn2096-1472.2023.004.013

基于复杂网络的以太坊重要账户识别研究

朱小栋，刘 欣

(上海理工大学管理学院，上海 200093)
✉zhuxd@usst.edu.cn; 540725290@qq.com

摘 要：区块链技术发展迅速、使用广泛，以太坊作为区块链2.0的代表，产生了大量的交易数据。为了研究以太坊交易中用户、矿池等相关主体的重要性及其在整个网络中的影响，构建了一种基于复杂网络理论的以太坊交易网络模型。首先，提出将度排序、K−shell、H−index和PageRank四种算法运用于以太坊网络节点的重要性排序。然后，分别进行静态和动态网络攻击，并通过两种评判指标判断攻击效果。理论分析结果表明，以太坊网络为无标度网络，少数节点在网络中具有关键作用，并且对网络进行动态攻击的效果好于静态攻击。最后，强调关键节点的设备维护对以太坊交易网络具有重要的安全意义。

关键词：区块链；以太坊；复杂网络；关键节点；网络攻击
中图分类号：TP309.2 **文献标识码**：A

Research on the Important Accounts Identification of Ethereum based on Complex Networks

ZHU Xiaodong, LIU Xin

(*Business School*, *University of Shanghai for Science and Technology*, *Shanghai* 200093, *China*)
✉zhuxd@usst.edu.cn; 540725290@qq.com

Abstract: With the rapid development and wide use of the Blockchain technology, Ethereum, as a representative of Blockchain 2.0, has generated a large amount of transaction data. In order to study the importance of users, mining pools and other related entities in Blockchain transactions and their influence on the whole network, this paper proposes to construct an Ethereum transaction network model based on complex network theory. First of all, four algorithms, namely, degree ranking, K-shell, H-index and PageRank, are applied to the importance ranking of Ethereum network nodes. Then, static and dynamic network attacks are carried out respectively, and the attack effect is judged through two evaluation indicators. The theoretical analysis results show that the Ethereum network is scale-free, a few nodes play a key role in the network, and the effect of dynamic attack on the network is better than static attack. Finally, it is emphasized that the equipment maintenance of key nodes has important security significance for the Ethereum transaction network.

Keywords: blockchain; Ethereum; complex network; critical nodes; network attack

1 引言(Introduction)

区块链(Blockchain)是近年全球信息科技领域最受关注的技术，它具有去中心化的核心优势，可以有效解决中心化机构中普遍存在的成本高、效率低和数据安全等问题[1]。区块链特有的去中心化特征大大减少了受到网络攻击时产生的危害，但是攻击者除了利用区块链系统本身设计或实现方面存在的漏洞攻击区块链，还可以通过攻击区块链周边的设施，例如交易所、数字钱包、矿场、矿池等获得利益。区块链技

基金项目：上海高校智库内涵建设计划(战略研究)项目“上海加强新动能培育和关键核心技术突破研究”(1022303001)；国家自然科学基金资助项目“产业互联‘智造’供需网的结构、演化及其动力学研究”(71871144)。

术本身的设计缺陷或其系统部署不当，并不是产生被攻击风险的主要原因，而是由于区块链应用开发、软件工具及其实践过程中缺乏必要的安全考虑或措施造成的。因此，找到关键的用户即可对该用户的外部工具等设施进行安全性建设，这对于整个区块链的安全建设具有重要作用。本文的研究重点为识别关键的用户(区块或节点)设施[2]。

本文针对区块链的安全问题进行研究，首次将复杂网络理论引入区块链研究中，研究人员将区块链或区块链的子集中发生的交易流表示为一个网络，其中节点是以太坊账户，即外部账户。在以太坊(Ethereum)场景中，一些加密货币转移、智能合约的创建或合约的调用都可以称为交易。区块链中记录的每笔交易都对应于在网络中创建的新连接。复杂网络提供了合适的建模，将区块链表示为一个复杂的系统分析其本质，通过运算得出相关特征值，即度、度分布，并使用多种节点重要度评估方法，如PageRank算法、K-shell算法和H指数算法，将节点按照重要度进行排序。本文基于以上基本原理找出区块链的重要节点，并对其安全建设提出建议。

2 背景(Background)

区块链是一种以区块记录交易的分布式账本[3]。每个区块都包含一组交易，交易可以记录任何类型的事件，这是因为区块链使用了P2P系统、加密技术、分布式共识方案等，并且使用者采取匿名的形式，从而使区块链具有公开、可追踪、抗篡改等属性[4]。以太坊作为区块链2.0的代表，是一种开源的基于区块链的软件平台。以太坊通过转换账户的余额并改变状态进行交易。状态表示所有账户的当前余额及其他数据，作为独立数据进行编码和维护。账户的类型有外部拥有账户和合同账户两种。交易中使用到的加密货币叫作以太币。使用这种加密货币，可以向其他账户进行付款(P2P支付)或者支付某种操作请求。

针对复杂网络的研究是由网络中“小世界”和“无标度”两个特征兴起而来的，WATTS等[5]利用演员合作数据发现了小世界特性，BARABASI等[6]发现万维网的无标度性质；在复杂网络中，挖掘发现网络的重要节点是一个核心问题，目前包含多种中心性指标，如度中心性、介数中心性[7]、接近中心性[8]、网络的核[9]、PAGE等[10]基于随机游走排序算法提出PageRank指标等。复杂网络理论对于实际网络的应用也取得重要成果。张彦超等[11]利用复杂网络理论构造了基于在线社交网络的信息传播模型，发现模型符合在线社交网络特性，并且该网络高度连通；BARRAT等[12]通过复杂网络理论研究航空交通关系，发现机场的航线数和机场的运力呈超线性关系；段文奇等[13]利用复杂网络研究全球贸易，发现国际贸易存在动态演化过程。

考虑到以太坊支付中具有P2P支付特征，结合加密货币交易网络结构，符合构成一个复杂网络的条件。本研究运用以太坊外部拥有账户的交易情况进行建模分析，旨在找到关键账户加强其自身安全性建设。利用复杂网络的相关研究方法进行分析，可以掌握以太坊网络结构、识别关键账户，并具有易操作、实验结果明显等优势。

3 研究理论和数据(Research theory and data)

3.1 以太坊数据及建模

3.1.1 以太坊数据源

区块链系统拥有三种原始数据类型：区块、交易回执和追踪数据。

区块数据直接存储在以太坊区块链中，每个区块由区块头和区块事务组成[14]。区块头包含一个区块的基本信息，包括矿工地址、时间戳等；区块事务构成区块的主体，每一个事务由From,To,Value,Input(发起账户，目标账户，值，输入)等组成。追踪数据本质上是一种内部合同，即资金从一个账户转移到另一个账户。

XBlock.pro是一个分享区块链数据集的共享平台，它收集了当前主流的区块链加密货币的相关数据，并对数据进行清洗和归类[15]。本文数据采集自XBlock.pro网站和文献[16]，通过运行以太坊的全部节点获得链上数据(From,To,Value)，截至2019 年3 月10 日，共收集5亿多个节点和38亿条交易数据，本研究采用了其中随机选择的107,452 个未标记节点和374,641 条交易链接。

3.1.2 交易网络建模

目前，区块链系统的交易模型分为两种，即以交易为中心的模型和以账户为中心的模型[15]，两者的交易模式不同，网络建模差别很大。简单来说，以交易为中心的模型区块链交易系统建模工作中，以交易作为网络的节点，将用户作为网络的边；而以账户为中心的模型区块链交易系统建模工作中，以账户作为节点，资金转移作为边。

本文的研究对象以太坊是一种基于账户为中心的交易模型的区块链系统。在相关研究中，有三种以太坊交易数据网络建模方式，分别是资金流网络、智能合约创建网络和智能合约调动网络[15]。在资金流网络中，外部账户和合约账户抽象为网络的节点，边表示资金流向；在智能合约创建网络和智能合约调用网络中，分别以外部账户和合约账户作为网络节点，与资金流网络不同的是，边分别表示合约的创建和合约的调用。

本研究为以太坊账户的关键节点识别，符合资金流网络

特征，故提出采用构建资金流网络作为本研究的区块链系统建模方法。账户作为网络中的节点v，当账户(节点v_i)与账户(v_j)之间存在交易或者资金流动时，则两者之间存在边e_{ij}。

3.2 以太坊节点复杂网络的构建

复杂网络记为图$G=(V,E)$。其中，V表示节点$v_1,v_2,\cdots,v_k$的集合，节点的数量记为$N=|V|$，$E=\{(v_i,v_j)|v_i,v_j\in V\}$表示边的集合。

3.2.1 节点重要性

(1)度(Degree)定义为连接到该节点v_i的边的数量，记为k_i。网络中所有节点的度的平均值称为网络的平均度，记为$<k>$。网络的平均度和节点之间的关系如式(1)所示：

$$<k>=\frac{1}{N}\sum_{i=1}^{N}k_i \tag{1}$$

本文提出度分布作为以太坊网络的一个重要统计特征。以太坊网络中节点的度分布可用分布函数$P(k)$描述，它表示整个网络中具有度值为k的节点在网络中所占的比例，即概率。

$$P(k_i)=\frac{n(k_i)}{N} \tag{2}$$

式(2)中，$n(k_i)$表示度为k_i的节点数量，通过度k_i的概率$P(k_i)$可以观察网络的节点分布情况。

(2)K−shell算法是一种基于网络全局结构的粗化分解方法，其判断节点重要性的根据是该节点在网络中所在的位置[17]。本文提出利用K−shell算法识别以太坊网络的重要节点。

图1展示了K−shell算法的分解示意图，具体的算法步骤如下：首先，删除网络中所有度$k=1$的节点及其连边，得到新的网络，再删除此类节点和连边(度为1)，直到网络中不再出现此类节点，此时将所有被删除的节点归为1−shell层[18]，并且这些节点的Ks值定为1。其次，重复k值为$2,3,\cdots,n$的上述过程，直到网络中所有节点均被分层、分配Ks值。在所有节点中，Ks值最大的节点所属的层称为网络的核心层，处于网络核心层的节点具有最大的影响力。

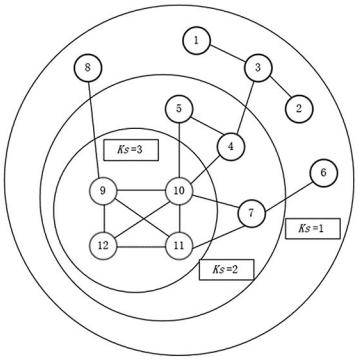

图1 K−shell分解示意图

Fig.1 Diagram of K−shell decomposition

在利用以太坊构建的复杂网络中，k值越大说明该节点越处于核心位置，即该账户的影响力越大。

(3)H−index原用来描述某位科研人员发表文章的影响力，定义为某名科研人员发表的所有文章中有h篇的引用次数都大于等于h，其他剩下的论文的引用次数都小于h[19]。

本研究中引入H指数研究以太坊交易网络影响力的评估。在以太坊交易网络中，H指数是指满足一个账户节点至少存在h个度大于h的邻居节点时的最大h值[20]。H指数值越大，说明该账户节点对于整个交易网络的影响力越高。

(4)PageRank算法又叫网页排名算法[21]，计算如式(3)所示：

$$PR(X)=\alpha\sum_{Y_i\in S(X)}\frac{PR(Y_i)}{n_i}+\frac{(1-\alpha)}{N} \tag{3}$$

在以太坊交易网络中，式(3)中$PR(X)$是指账户X所获得的PR值，$S(X)$是指所有指向账户X的账户的集合，$PR(Y_i)$是指账户X的入链账户Y_i的PR值，n_i指账户Y_i的出链数量，α是阻尼系数，一般情况下取0.85。

此算法最初用于评估网页的重要性排名，本文提出利用该指数表示以太坊网络中节点重要性排名。

3.2.2 评价标准

本文利用以上四种节点重要度评价算法得出排序结果，基于网络鲁棒性对算法排序结果进行评价。就以太坊网络而言，采用最大连通片相对规模与剩余连通平均规模量化移除节点后对网络结构与功能的影响[22]，以此分析网络结构、评价节点结构的重要性。

(1)最大连通片相对规模。将节点按照重要度评估计算法从大到小进行排序，删除一部分节点后，观察对网络极大连通子集的影响，计算如式(4)所示：

$$G=\frac{1}{N}\max_{1\leqslant k\leqslant M}|G_{k,o}| \tag{4}$$

式(4)中，$|G_{k,o}|$是指删除该部分节点后的网络连通子集的节点数($M\leqslant N$，M表示分裂后的子集数量)。G值随着节点的删除而降低的幅度越大，说明采用该方法攻击网络的效果越好[22]。

(2)剩余连通片平均规模。剩余连通片平均规模表示删除最大连通片以后，网络中剩余的连通子集所包含的节点数的平均值[23]，计算如式(5)所示：

$$\mu=\frac{1}{M-1}\left(\sum_{k=1}^{M}|G_{k,o}|-N\times G\right) \tag{5}$$

式(5)中，$|G_{k,o}|$指删除该部分节点后的网络连通子集的节点数，G为式(4)的最大连通片相对规模，M表示为分裂后的

子集数量。μ值随着节点的删除而降低的幅度越大，说明采用该方法攻击网络的效果越好。

4 实验和结果(Experiments and results)

4.1 数据处理

本文采用上述提到的真实以太坊交易数据进行实验，截至2019 年3 月1 日，共收集并筛选出374,641 条交易数据，将数据预处理后得到该网络中节点的总数为107,451 个，边的总数为123,740 条。图2是从网络中随机选出的1,000 个节点的网络拓扑图。

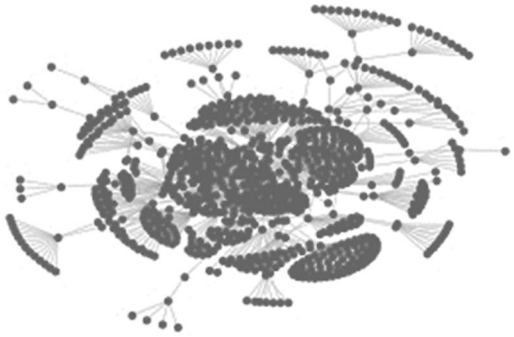

图2 N=1,000的拓扑示意图

Fig.2 Topology diagram of N=1,000

图2中，网络中心处的节点较为密集，边缘位置的节点较为稀疏。

(1)Degree。节点度的概率分布通过曲线实现，度分布概率曲线如图3所示，度的概率分布服从幂律分布。基于度的节点发布情况如表1所示。

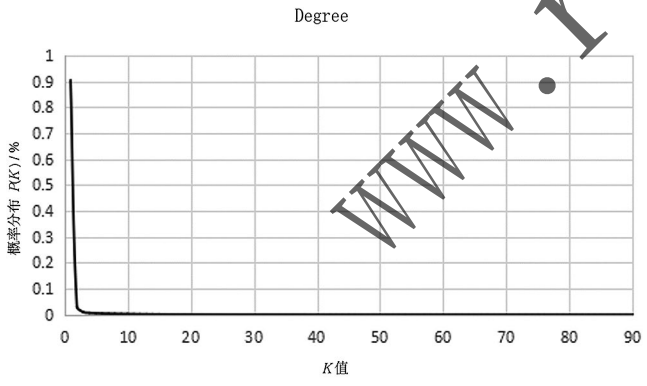

图3 度分布概率曲线

Fig.3 Degree distribution probability curve

表1 基于度的节点分布情况

Tab.1 Nodes distribution based on degree

K	节点数	概率分布*P*(*K*)/%
1	97,139	90.403
2—3	4,378	4.074
4—10	3,338	3.106
11—22	1,542	1.435
23—10,366	1,054	0.981

度为1的节点有97,139 个，占比为90.403%；度为2和3的节点分别为3,289 个和1,089 个，占比共为4.074%。将所有节点按照度从大到小进行排序，前0.981%的节点度在23—10,366，其中10 个节点度在1,000以上。

这表明在以太坊中，大部分节点的度很小，而少部分节点的度很大，即大部分的用户(节点)只和很少的用户交易，而少部分用户和很多的用户进行交易。这些度很大的节点可能是矿池节点或交易中心。

(2)K−shell。*Ks* 指标可以用来刻画节点的中心性，节点的核心数越高，表明节点的位置越靠近中心，核心数越低，表明节点的位置越靠近边缘。*Ks* 指数分布概率曲线如图4所示。

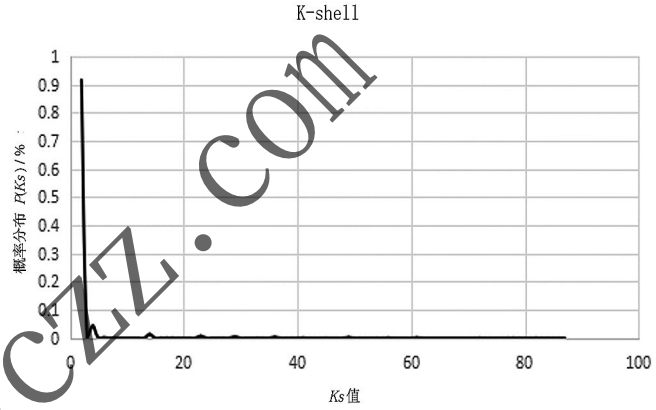

图4 K−shell分布概率曲线

Fig.4 Distribution probability curve of K−shell

基于*Ks* 核心数的节点发布情况如表2所示。

表2 基于K−shell的节点分布情况

Tab.2 Nodes distribution based on K−shell

Ks	节点数	概率分布*P*(*Ks*)/%
1	98,349	91.529
2—47	8,511	7.921
48—86	591	0.550

Ks 核心数为1的节点占比最高，节点数为98,349 个，占比为91.529%；占比第二高的*Ks*值为3，节点数为4,643 个，占比约为4.321%；*Ks* 值为13和22的节点分别有1,396 个和726 个，占比分别约为1.299%和0.676%。

在以太坊交易中，绝大部分节点都在网络的边缘，即其核心数为1；核心数在2—47的节点占比为7.921%，除了上边缘节点外的多数节点处于此情况；核心数在48—86的节点占比为0.550%。核心数最大的这些节点处于网络的中心，符合拓扑结构中的中心节点少、边缘节点多的特点。

(3)H−index。H−index用来描述节点的影响力排行。*H* 指数分布概率曲线如图5所示。

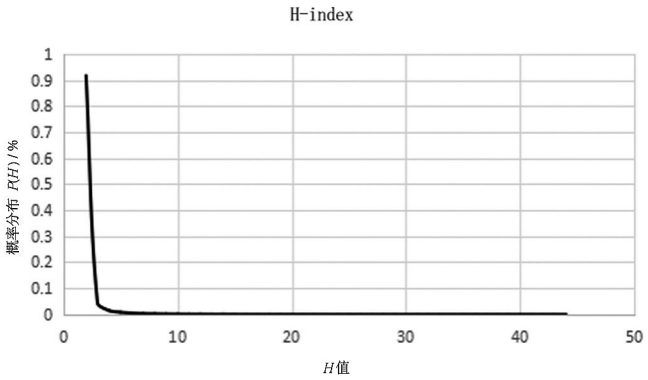

图5 H-index分布概率曲线

Fig.5 Distribution probability curve of H-index

基于*H*指数的节点发布情况如表3所示。

表3 基于H-index的节点分布情况

Tab.3 Nodes distribution based on H-index

H	节点数	概率分布*P*(*H*)/%
1	98,563	91.728
2—3	6,256	5.822
4—9	3,798	2.067
10—19	246	0.322
20—55	65	0.061

*H*指数在1—55范围，其分布概率同样服从幂律分布。*H*指数值为1的节点有98,563个，占比为91.728%；*H*值为2—3和4—9的节点数分别为6,256个和3,798个，占比分别为5.822%和2.067%；*H*值最大为55，则*H*值为20—55的节点，即大致可以看作本网络中该指标识别的重要节点，节点数为65个，占比为0.061%。基于*H*指数可知，在以太坊交易网络中，少数的账户节点对其他节点造成的影响力巨大。

(4)PageRank。*PR*指数代表节点在全局网络中的影响力，其绝大部分节点的*PR*值不尽相同，由此本研究中按照数值的数量级进行划分，基于*PR*值的节点发布情况如表4所示。

表4 基于PageRank的节点分布情况

Tab.4 Nodes distribution based on PageRank

PR	节点数	概率分布*P*(*PR*)/%
0.0469—0.0175	3	2.792×10^{-5}
0.00904—0.00105	32	0.0298
0.000946—0.000100	791	0.736
9.998×10^{-5}—9.995×10^{-5}	5,379	5.006
9.987×10^{-6}—2.778×10^{-6}	101,246	94.225

*PR*值在数量级为10^{-2}的节点只有3个，而数量级为10^{-6}的节点有101,246个，占据所有节点的94.225%。基于*PR*指数可知，极少量的账户会对以太坊交易网络形成全局性的影响，而大量的节点的影响力很小。

综上，对于以太坊交易网络，我们可以看到度值为1、*Ks*核心数为1、*H*指数为1和数量级为10^{-6}的节点数量占整个网络节点约90%。基于此，下一步实验中选取10%左右的节点作为重要节点，重点关注删除这些节点后对网络分别进行静态攻击和动态攻击的效果。

4.2 网络攻击模拟

基于以太坊交易网络，本文对度排序方法、K-shell算法、H-index算法和PageRank算法进行了比较分析。根据四种算法的排序结果，分别以静态攻击与动态攻击的方法移除排名靠前的一定比例的节点，模拟网络遭受蓄意攻击时最大连通片相对规模和剩余连通片平均规模的变化情况，从而分析网络结构。

(1)静态攻击效果。在静态攻击模式中，节点重要度指标不随着网络结构变化而变化，并保持与原始网络中各指标的值一样[22]。在模拟蓄意攻击网络对网络影响的实验中，最大连通片相对规模和剩余连通片平均规模情况如图6和图7所示。

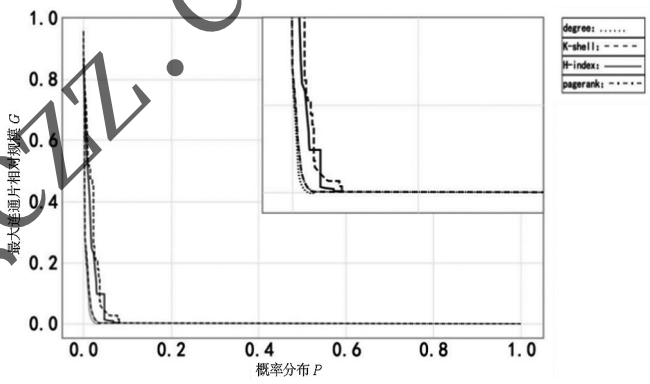

图6 静态攻击下最大连通片相对规模

Fig.6 The relative scale of the largest connected component under static attack

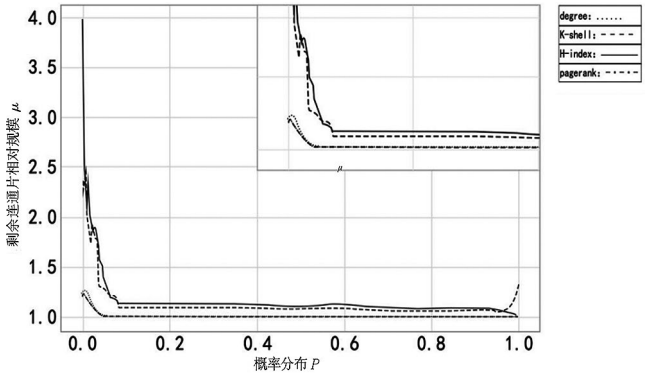

图7 静态攻击下剩余连通片平均规模

Fig.7 The average scale of residual connected component under static attacks

在本实验的以太网交易网络中，Degree和PageRank指标下的*G*值(最大连通片相对规模)下降幅度较大，这表明度排序方法和PageRank算法导致网络最大连通片相对规模变小的总

体趋势较为明显，即这两种识别节点重要性的方法更加适用于本研究的以太坊交易网络。观察图7可以看见，剩余连通片平均规模会急剧增大，其中在H−index算法中表现得最为明显，整个网络遭到严重的破坏。

分析图6可以发现，在蓄意攻击时，以太坊交易网络对于重要度高的节点的移除非常敏感。当少数(约10%)的中心节点被移除时，G 值立即减小，即一些较大的子集即刻被碎片化并从网络上剥离，网络结构破坏严重；通过图7可以发现，当10%左右的节点遭到攻击后，网络发生了崩溃。

本文研究的以太网交易网络中，我们可以理解为重要度排名前10%的节点对于整个网络的结构和稳定性都有着极为重要的意义，并且整个以太坊交易对于这类节点(重要度排名前10%的节点)的移除表现最敏感。

(2)动态攻击效果。在动态攻击模式中，每移除一个节点或一定比例的节点，节点的各个重要度指标需要更新一次[22]。最大连通片相对规模和剩余连通片平均规模情况如图8和图9所示。

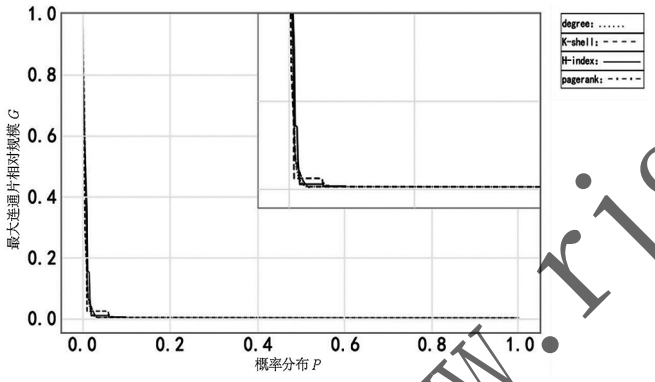

图8 动态攻击下最大连通片相对规模

Fig.8 The relative scale of the largest connected component under dynamic attacks

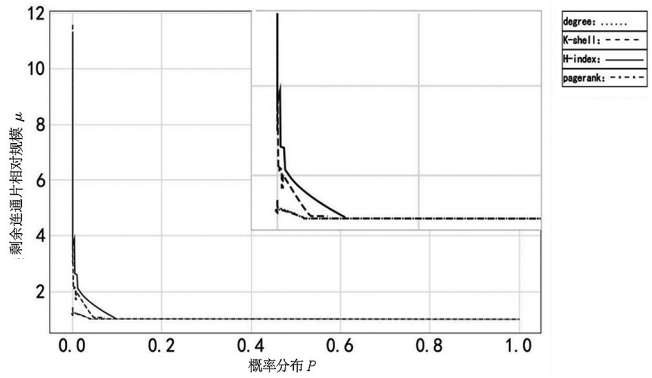

图9 动态攻击下剩余连通片平均规模

Fig.9 The average scale of residual connected component under dynamic attacks

观察图8可以发现，在最大连通片相对规模实验中，删除前约5%的节点后，G 值即达到最低值0，此外度排序、K−shell、H−index和PageRank算法对实验的结果影响没有特别大的差异，这表明以太坊交易网络中节点的重要性很容易通过这4 种算法识别且效果较好、误差不大。在剩余连通片平均规模实验中，PageRank算法获得的 μ 值波动最小，可见使用其算法识别的重要节点对于动态攻击的反应较小。

此外，静态攻击和动态攻击实验中对比两种攻击模式下的网络瓦解效果，发现动态攻击明显优于静态攻击。

5 结论(Conclusion)

区块链由于其去中心化的结构特征可以有效地减少网络攻击带来的危害，但是无法避免运行区块链时使用到的外部设备遭到恶意攻击。为了验证区块链交易中重要的节点对于整个网络具有重要的意义这一假设，本文以以太坊交易为例，创建复杂网络，通过度、K−shell、H−index和PageRank算法识别以太坊网络中的关键节点，并通过攻击网络实验分析得出结论：①以太坊网络是一个典型的无标度网络；②以太坊网络表现出对网络攻击的脆弱性，破坏相同数量的节点时，动态攻击效果优于静态攻击；③在以太坊网络中往往有少量(约10%)的节点在整个网络中占据了十分重要的地位，当这些节点被破坏，就会导致整个网络崩溃，因此需要强调这些节点外部设备的安全性建设。

近年来，越来越多的研究人员将目光从区块链原理转移到区块链的安全性研究中，本文提出将区块链结合复杂网络方法识别关键节点，此类节点对于整个网络的鲁棒性和结构稳定性具有重要作用，并呼吁研究人员加强对此类节点外部设备的安全建设，其具体实施的方法、流程将是相关人员下一步研究的方向。

参考文献(References)

[1] 袁勇,王飞跃.区块链技术发展现状与展望[J].自动化学报,2016,42(04):481−494.

[2] 魏松杰,吕伟龙,李莎莎.区块链公链应用的典型安全问题综述[J].软件学报,2022,33(01)324−355.

[3] 黄立波,王伟,徐彦军,等.基于区块链的数字结业证书管理系统及其性能评估[J].华东师范大学学报(自然科学版),2020(06):72−81.

[4] 郑佩娜.基于区块链技术的数字资产交易:案例分析视角[D].杭州:浙江大学,2018.

[5] WATTS D J, STROGATZ S H. Collective dynamics of 'small−world' networks[J]. Nature, 1998, 393(6684):440−442.

[6] BARABASI A L. Scale−free networks: a decade and beyond[J]. Science, 2009, 325(5939):412−413.

(下转第51页)