



本科生毕业论文（设计）

Undergraduate Graduation Thesis (Design)

题目 Title: _____

院 系
School (Department): 数据科学与计算机学院

专 业
Major: _____

学 生 姓 名
Student Name: _____

学 号
Student No.: _____

指 导 教 师(职 称)
Supervisor (Title): _____

时间: 年 月 日

Date: Month Day Year

说 明

1. 毕业论文（设计）的写作格式要求请参照《中山大学本科生毕业论文的有关规定》和《中山大学本科生毕业论文（设计）写作与印制规范》。
2. 除完成毕业论文（设计）外，还须填写三份表格：
 - (1) 表一 毕业论文（设计）开题报告；
 - (2) 表二 毕业论文（设计）过程检查情况记录表；
 - (3) 表三 毕业论文（设计）答辩情况登记表。
3. 上述表格均可从教务部主页的“下载中心”处下载，如表格篇幅不够，可另附纸。每份毕业论文（设计）定稿装订时应随同附上这三份表格。
4. 封三是毕业论文（设计）成绩评定的主要依据，请认真填写。

Instruction

1. Please refer to '*The Guidelines to Undergraduate Graduation Thesis (Design) at Sun Yat-sen University*' and '*The Writing and Printing Format of Undergraduate Graduation Thesis(Design) at Sun Yat-sen University*' for anything about the thesis format.
2. Three forms should be filled up before the submission of the thesis (design):
 - (1) Form 1: Research Proposal of Graduation Thesis.
 - (2) Form 2: Process Check-up Form.
 - (3) Form 3: Thesis Defense Performance Form.
3. All the above forms could be downloaded on the website of the Office of Education Administration. If there is not enough space in the form, please add extra sheets. Each thesis (design) should be submitted together with the three forms.
4. The form on the inside back cover is the grading sheet. Please fill it up before submission.

毕业论文（设计）成绩评定记录
Grading Sheet of the Graduation Thesis (Design)

指导教师评语
Comments of Supervisor:

成绩评定
Grade:

指导教师签名
Supervisor Signature : Date:

答辩小组意见
Comments of the Defense Committee:

成绩评定
Grade:

签名：
Signatures of Committee Members Date:

院系负责人意见
Comments of the Academic Chief of School:

成绩评定
Grade:

签名
Signature: 院系盖章
Stamp: Date:

表一：毕业论文（设计）开题报告
Form 1: Research Proposal of Graduation Thesis (Design)

论文（设计）题目 Thesis (Design) Title:	
(简述选题的目的、思路、方法、相关支持条件及进度安排等) (Please briefly state the research objective, research methodology, research procedure and research schedule in this part.)	
Student Signature: Date:	
指导教师意见 Comments from Supervisor:	
1.同意开题 2.修改后开题 3.重新开题 1.Approved(<input type="checkbox"/>) 2. Approved after Revision (<input type="checkbox"/>) 3. Disapproved(<input type="checkbox"/>)	
Supervisor Signature: Date:	

表二：毕业论文（设计）过程检查情况记录表

Form 2: Process Check-up Form

指导教师分阶段检查论文的进展情况（要求过程检查记录不少于 3 次）
The supervisor should check up the working process for the thesis (design) and fill up the following check-up log. At least three times of the check-up should be done and kept on the log.

第 1 次检查（First Check-up）：

学生总结

Student Self-summary:

指导教师意见

Comments of Supervisor:

第 2 次检查（Second Check-up）：

学生总结

Student Self-summary:

指导教师意见

Comments of Supervisor:

第3次检查 (Third Check-up):

学生总结

Student Self-summary:

指导教师意见

Comments of Supervisor:

第4次检查**Fourth Check-up**

学生总结

Student Self-summary:

指导教师意见 (Comments of Supervisor):

学生签名 (Student Signature):

日期 (Date) :

指导教师签名 (Supervisor Signature):

日期 (Date) :

指导教师意见 Comments of Supervisor:

**总体完成情况
(Overall Assessment)**

- 1、按计划完成，完成情况优 (Excellent):
- 2、按计划完成，完成情况良 (Good):
- 3、基本按计划完成，完成情况合格 (Fair):
- 4、完成情况不合格 (Poor):

指导教师签名 (Supervisor Signature):

日期 (Date) :

表三：毕业论文（设计）答辩情况登记表
Form 3: Thesis Defense Performance Form

答辩人 Student Name		专业 Major	
论文（设计）题目 Thesis (Design) Title			
答辩小组成员 Committee Members			
答辩记录 Records of Defense Performance:			
记录人签名 (Clerk Signature):		日期 (Date) :	

学术诚信声明

本人所呈交的毕业论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。本毕业论文的知识产权归属于培养单位。本人完全意识到本声明的法律结果由本人承担。

本人签名: _____ 日期: _____

Statement of Academic Integrity

I hereby acknowledge that the thesis submitted is a product of my own independent research under the supervision of my supervisor, and that all the data, statistics, pictures and materials are reliable and trustworthy, and that all the previous research and sources are appropriately marked in the thesis, and that the intellectual property of the thesis belongs to the school. I am fully aware of the legal effect of this statement.

Student Signature: _____ Date: _____

摘要

(* 中文摘要不少于 300 字。语言力求精练、准确，以 300—500 字为宜。

在摘要中简要描述你所解决问题的背景、问题的描述、难点与挑战性、研究方法、所提出的解决问题的具体方法、实验结果和最终结论。主要突出自己的工作，最后介绍论文的学术意义，与相关方法比较是否在某些方面有所提高。关键是让读者能从摘要中看出你在论文阶段完成的主要工作。

摘要应具有独立性和自含性，即不阅读论文的全文，就能获得必要的信息，摘要中有数据、有结论，是一篇完整的短文，可以独立使用，可以引用，可以用于推广。摘要的内容应包含与毕业论文同等量的主要信息，供读者确定有无必要阅读全文，也可供文摘等二次采用，摘要一般应说明研究工作的目的、实验方法、结果和最终结论等，而重点是结论和结果。

除非确实需要，否则在摘要中一般不要用图表、非公知公用的符号和术语。

摘要使用第三人称，被动语句；避免大量背景、必要性、意义的描述，可简要提及；不要过分使用形容词；*)

关键词：(*关键词 1；关键词 2；……最后一个关键词不打标点符号

关键词一般为论文中出现的重要的、描述主题的词；

应包含技术领域和问题领域两大类关键字；

尽量不要使用组合词；

中英文关键词不少于 3 个，不多于 5 个。*)

ABSTRACT

(* 英文摘要以 250—400 个实词为宜，严格使用英文标点符号。中、英文摘要意思要基本相同。在撰写完英文摘要后，请执行一次自动拼写检查，以减少英文拼写错误的可能性。*)

Keywords: (* Keywords 1; Keywords 2; 3~5 个，最后一个关键词不打标点符号
中英文关键词要严格对译。*)

目 录

第一章 引言.....	1
1.1 研究背景与意义	1
1.2 问题描述	3
1.3 本文的工作	3
1.4 论文结构简介	4
第二章 相关工作综述.....	5
2.1 EOS.IO 交易网络的研究	5
2.1.1 区块链技术的基本介绍	5
2.1.2 EOS.IO 系统的基本介绍.....	6
2.1.3 EOS.IO 相关研究的发展.....	8
2.2 基于区块链的网络科学研究	9
2.2.1 网络的定义表示与基础特征	9
2.2.2 基于区块链的网络研究发展	9
第三章 EOS.IO 交易网络性质分析基本框架.....	11
3.1 框架概述	11
3.2 框架详解	12
3.2.1 数据收集	12
3.2.2 网络构建	14
3.2.3 网络分析	16
3.2.4 总结应用	17
第四章 实验结果与分析.....	18
4.1 数据获取过滤与基本统计	18
4.2 网络构建	20
4.3 网络分析	21
4.3.1 货币转移网络 (MTG)	22
4.3.2 用户创建网络 (ACG)	25
4.3.3 合约授权网络 (CAG)	29
4.3.4 用户投票网络 (AVG)	31
4.4 总结应用	34
第五章 总结与展望.....	35
5.1 工作总结	35
5.2 研究展望	36
参考文献.....	37

致 谢.....	38
附 录.....	39

批注 [刘1]: 好像字体大小有问题
标题是三号宋体加粗
章: 四号黑体
其它: 小四号宋体

第一章 | 引言

批注 [刘2]: 格式

本章分为四个部分，首先阐述基于复杂网络的 EOS.IO 性质分析的问题来源、研究动机和意义，随后简要说明本文涉及研究的主要问题和挑战，接着概述本文的主要贡献、创新点和实验结果，最后简要介绍本文后续的章节安排和主要内容。

1.1 研究背景与意义

批注 [刘3]: 格式

自 2008 年比特币货币诞生以来，区块链技术逐步进入大众视野。由于具有着去中心化、不可伪造等特点，区块链技术不仅有可能彻底改造货币市场、金融服务等行业的每一个细节，甚至为所有领域提供重新配置的可能性。有许多企业家和开发者在不断尝试推广和革新区块链技术，以便使单一的区块链平台能够支持更为广泛的应用程序。

现如今，区块链技术已经步入以智能合约为核心的 2.0 时代，并且已经开始向以去中心化应用程序（decentralized application, DApp）为核心的 3.0 时代发展。区块链技术已经在 DApp 开发者中广泛流行，尤其是在创建智能合约方面。以太坊是目前最大的支持智能合约的平台，同时其也是继比特币之后使用最广泛的区块链系统。然而，由于其使用的共识机制是工作量证明（Power of Work, PoW）共识机制，在支持 DApp 的迅速发展方面逐渐显示出力不从心。对于布置在区块链上的应用来说，它们需要一个更为灵活的平台，可以支持每日数以千万的活跃用户、可以实现低延时和高效的交易确认，同时具有较高的时序性能和并发性能。

2018 年 6 月，一款名为 EOS.IO¹（Enterprise Operation System）的开源平台软件发布。与比特币和以太坊不同，EOS.IO 采用了唯一能满足上述区块链性能要求的去中心化共识算法——委托股权证明（Delegated proof of stake, DPoS）机制。在这种共识机制下，EOS.IO 做到了真正的秒级出块，同时极具扩展性，而且面对用户完全免费，由 DApp 开发者抵押付费。DPoS 共识机制还赋予了 EOS.IO 极为优秀的吞吐性能，使其可在单一线程上每秒处理超过 8000 条交易信息。所有这些都使得 EOS.IO 在上线后不久得到了迅速发展。根据报道，在 EOS.IO 平台上线仅三个月，其每日 DApp

删除了：漏洞修复

批注 [刘4]: 这个时序性能是啥？会跟低延时重复吗

批注 [刘5]: 引用？

¹ 本文用 EOS.IO 代指该平台软件，而 EOS 代指其对应的加密货币（俗称柚子币）

用户和交易量均超过了以太坊。另一方面，图 1-1 显示了近三个月来发生在 Dapp 上的交易额最多的三条链交易额波动图。从图中看出，即使到了今天，EOS 依旧拥有很大的 DApp 市场，远超于以太坊等链。

批注 [刘6]: 应该是 DApp，下面同

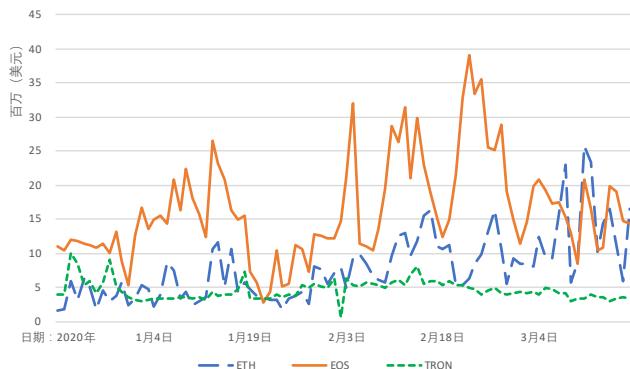


图 1-1 近 90 天 Dapp 交易额对比图

批注 [刘7]: 数据来源加脚注？

EOS.IO 发展至今，在工业界和研究界都获得了极大的关注。然而直到现在，学术界都还没有一个完整全面的对 EOS.IO 生态的了解。因此，本篇论文将致力于填补这个研究的空白，利用 EOS.IO 的真实数据进行网络构建，从复杂网络的角度深入分析 EOS.IO 的生态系统。该问题具有如下意义：

- 1) 本文将提供一个全面完整的针对 EOS.IO 生态的分析，通过挖掘网络特性，帮助学者或 Dapp 的开发者更好地理解 EOS.IO 的生态系统，我们的数据也将有利于后续的 EOS.IO 相关研究；
- 2) 我们将结论从数据角度上升为用户行为角度，将进行具体的案例分析从而可以探寻 EOS.IO 成功的意义以及是否真的存在“虚假繁荣”的现象；
- 3) 在分析过程中我们发现的异常行为可以为系统管理者提供更好的管理意见，同时促进区块链自身的安全监管；
- 4) 区块链数据的公开可用性以及不可篡改性为研究商业行为提供了巨大的商业价值，EOS.IO 作为一个新兴的商业系统，对其进行全面剖析对于促进价格检测、交易欺诈检测、智能合约软件开发等领域的发展有着巨大的好处。

批注 [刘8]: 对 EOS.IO 生态没有一个完整全面的了解

1.2 问题描述

本文研究的问题是如何正确并全面地运用复杂网络对 EOS.IO 做一个综合性分析，包括其转账系统、投票系统、合约授权系统、用户创建系统等。在本文的网络分析中，我们希望建立一个通用且完整的交易网络分析体系，从数据提取到数据分析，并结合每个系统的特性进行具体的分析，既具有普适性，又有独到的针对性。然而，在面对实际的网络，我们面临着以下的挑战：

- 1) 目前缺少针对区块链的通用数据提取工具，虽然许多区块链网站提供了一些 EOS.IO 数据，但它们的数据提取工具通常都不是开源的，通过接口进行数据获取的过程会采集缓慢且耗费大量内存；
- 2) EOS.IO 的操作分为三种类型，其中延迟操作和调用操作是在链上明文显示的，然而内联操作的具体信息无法从链上直接获取；
- 3) 由于 EOS.IO 秒级出块的特点，其每天的交易数据量非常庞大，如果想要得到一个较为全面的结论而减少偶然性，则所面临的数据量是非常巨大的，其中还会包含许多非必要的信息；
- 4) 虽然 EOS.IO 的账户是可读性的代码，然而由于其匿名化的特点，我们很难得到他们的真实身份，不利于度中心性的分析和异常社区的检测。

批注 [刘9]: 换成活动？

批注 [刘10]: 字符串

1.3 本文的工作

本文的工作主要围绕从复杂网络的角度对 EOS.IO 交易网络进行全面的性质分析展开。整个分析框架包括三个部分：数据提取与过滤、网络构建、网络分析。在数据提取与过滤部分，我们创新性地总结了一种针对 EOS.IO 的原始数据提取方法，并对数据进行了有效地筛选；在网络构建部分，我们建立了四个网络，分别是货币转移网络（Money Transfer Graph, MTG）、用户创建网络（Account Creation Graph, ACG）、合约授权网络（Contract Authorization Graph, CAG）与用户投票网络（Account Vote Graph, AVG）。其中账户投票网络 EOS.IO 的特色所在，也是第一次得到全面的分析；在网络分析部分，我们总体上对各个网络的度分布、直径、聚集系数、同配性等多个复杂网络属性进行深入分析，并结合每个网络的特点，进行针对性的具体分析。总的来说，本研究的贡献如下：

- 1) 据我们所知，本文是第一篇从网络角度对 EOS.IO 系统进行的全面分析，并且做

到了从数据上升到用户。从复杂网络出发使我们能够更清晰地呈现用户之间的关系与行为互动，增进学者和 Dapp 开发者对 EOS.IO 生态系统的了解；

- 2) 本文提供了一个完整的 EOS 数据集，其包括一定时间区间内完整且具体的链上数据，其中包括投票数据、智能合约数据、用户创建数据、加密货币数据等等，且数据集处理完善，利于未来的进一步探索；
- 3) 本文从网络特性中分析出许多 EOS. IO 存在的异常现象，这些现象将帮助管理者更好地管理和监管系统。

在本文中，我们通过网络构建和分析，发现了许多的有趣的现象。例如，EOS. IO 系统遭受了严重攻击，有十分严重炒内存现象；EOS. IO 依旧具有财富分布的阶级性，符合二八准则等。这些都有利于我们对 EOS. IO 的深入了解，证明了我们工作的有效性和意义所在。

1.4 论文结构简介

本文一共分为五章，后续内容安排如下：

第二章，对本文涉及的理论基础进行综述，主要包括 EOS. IO 交易网络的研究与基于区块链的网络科学研究；

第三章，详细地描述了本文提出的 EOS. IO 交易网络性质分析基本框架，并描述了数据收集、网络构建的过程，并详细说明我们将如何进行网络分析；

第四章，介绍了本文具体的实验内容，包括数据获取与过滤、网络构建、网络分析，并对结论进行深入的探讨和分析；

第五章，对本文的工作内容进行总结，并对未来的工作提出了展望。

参考文献与致谢置于文章末尾处。

第二章 相关工作综述

批注 [刘11]: 这一部分, 2.1.1 放到 2.2.1, 然后第二节和第一节互换顺序会不会更合适一点

本章将阐述本研究涉及的理论基础，并简要介绍与本文相关的研究工作。首先将简要介绍区块链、EOS.IO 系统的相关背景知识，并简要总结目前 EOS.IO 相关研究的发展。随后将引入网络科学基础，包括网络的定义及其图表示、相关基础特征等，并将具体介绍以区块链为基础的网络科学发展，通过优缺点总结与对比，为本文研究工作做铺垫。

2.1 EOS.IO 交易网络的研究

2.1.1 区块链技术的基本介绍

区块链是分布式数据存储、点对点传输、共识机制、加密算法等多种计算机技术在互联网时代的创新应用模式【白皮书】。区块链技术起源于一名化名为“中本聪”(Satoshi Nakamoto) 的学者在 2008 年发表的一篇奠基性论文《比特币：一种点对点电子现金系统》【1】，被认为是继大型机、个人电脑、互联网之后计算模式的颠覆式创新。

通俗来讲，区块链（blockchain）是一个分布式、去中心化的公共账本【<https://www.investopedia.com/terms/b/blockchain.asp>】，所有的交易信息都存储在区块（block）上，每个区块按照创建的时间顺序依次组成一条长链（chain），类似一个公开的数据库。每一个包含新信息的区块在添加到区块链上时都会被赋予一个独一无二的哈希码，同时，网络中其他每个节点都在进行监听、接收、验证这个新区块的添加信息，以实现全网同步。而一旦区块上链，其信息便基本不可能被篡改或删除，且任何人都可以对信息进行访问查看。因此，区块链具有不可伪造、公开透明、集体维护等特征，被很多大型机构称为是彻底改变业务乃至机构运作方式的重大突破性技术。

目前，区块链已经从以比特币为代表的虚拟货币 1.0 时代，发展到以以太坊为代表的智能合约 2.0 时代，而在未来，区块链将走进 3.0 时代，即区块链在金融行业之外的各行业都能应用发展。目前，区块链的应用已经延伸到物联网、智能制造、供应链管理、数字资产交易等多个领域，基于区块链的去中心化应用也在蓬勃发展。

2.1.2 EOS.IO 系统的基本介绍

(1) 什么是 EOS.IO

EOS.IO 创立于 2018 年 6 月 9 日，是一个致力于创建和部署智能合约和分布式应用程序的下一代区块链平台。作为一个区块链平台，EOS.IO 专为企业级用例设计，并为公有和私有的区块链部署而构建。

批注 [刘12]: 主网上线于

EOS.IO 的创立目标是建立一个安全透明的系统，适用范围广泛，不仅可用于本地及跨境货币交易管理，还可用于在线选举和投票等应用。作为一个区块链底层平台，EOS.IO 具有秒级出块、高吞吐、用户免费等特点，有更完善的智能合约生态。对分布式应用的开发者们来说，EOS.IO 为他们提供了更熟悉的开发环境，即他们可以使用自己喜爱的开发模式和多种编程语言。EOS.IO 的到来远远提升了用户体验，很多人评价“EOS.IO 打开了区块链 3.0 的大门”。

(2) EOS.IO 的交易与账户

交易是 EOS.IO 区块的主体。EOS.IO 区块中的每一个交易都是由操作（action）组成的，操作是一个基本原子单位，一个交易内可以包括一个或多个操作，例如用户 A 想给用户 B 转账 10 EOS，这个交易可以分为 10 个操作，每个操作转账 1EOS。而每一个操作都是一次智能合约的调用，会根据合约名和相应函数做出的行为。例如，在 EOS.IO 中，每一笔转账都需要调用“eosio.token”合约的“transfer”函数，而每一个新用户的创建都需要调用“eosio”合约中的“newaccount”函数。总的来说，EOS.IO 的每一个交易都是由合约调用实现的，一个区块里包含多个交易，一个交易内包含多个操作。

EOS 的操作可以分为三类，调用操作、内联操作和延迟操作。调用操作即指用户的对智能合约的调用，内联操作发生在合约内部或者合约之间的操作，延迟操作主要用于启动延迟交易。

在 EOS.IO 中，用户的主要标识是可读的用户名，而不是公钥。账户可以拥有许多不同的权限，不同的权限与 EOS 网络上的不同操作相关联。每一个合约的调用都需要相应的用户进行授权才能发生，每个权限对应一对公私钥。权限一般分为 owner 权限和 active 权限，owner 是最高权限，可执行所有的操作。而 active 可执行除更改

owner 以外的所有操作，负责日常操作。

(3) EOS.IO 的共识机制与投票机制

在 EOS.IO 生态系统中，区块的产生和验证由被称为“块生产者(block producers)”的特殊节点执行。而这些生产者由 EOS.IO 利益相关者投票产生。基于 EOS.IO 的区块链使用委托股权证明机制 (DPoS) 来选出活跃的块生产者，授予他们权限在网络中进行有效区块的签署确认。

相比于比特币所使用的工作量证明 (Proof of Work, PoW) 共识机制，委托股权证明机制共识² (简称 DPoS 共识) 无需浪费算力资源争夺记账权。其拥有 21 个超级节点，出块顺序由超级节点决定，网络连接延迟较低的超级节点之间可以相邻出块，每个超级节点会一次连续生产 6 个区块以减少网络延时的影响。当超过三分之二的超级节点确认了交易后，此交易便不可逆转。通俗来说，如图 2-1 所示，DPoS 共识系统采用“股份投票”的方式决定谁来生产区块，而 PoW 共识系统采用电脑运算的方式决定谁是下一个区块生产者。在 DPoS 共识机制中，每个货币持有者都拥有投票权，21 个超级节点只要作恶就会被投票出局，并且有很多候选节点等待接班，因此反而具有相对低的中心化风险。采用 DPoS 共识机制的 EOS.IO 可以实现 0.5 秒的出块速度，1 秒实现区块的不可逆转。而也正因为如此迅速的出块速度，EOS.IO 的区块链在如此短的时间内就成长为迄今为止最长的一条区块链。



图 2-1 (不是终板！会重新画)

由于共识机制的特点，EOS.IO 的一切在治理方面都类似于一个民主的系统，每

² DPoS 共识现已发展为 BFT-DPoS 共识机制，本文的介绍都是基于 BFT-DPoS 共识

个决策都是由投票产生的而不是个人决策。每个 EOS 代币持有者都可以进行投票，投票过程记录在区块链网络中，这也就避免了欺诈或暗箱操作的嫌疑。

2.1.3 EOS.IO 相关研究的发展

目前，由于 EOS.IO 独有的特性和发展前景，其相关研究逐渐火热起来。其独有的生态环境和广泛的用户群体，吸引了越来越多的学者的目光。

(1) EOS 研究综述

Xu Brent 等学者在综述论文【】中对 EOS 系统的体系结构、性能以及经济模式进行了较为全面的分析。在结构上，作者对 EOS 系统的运行体系进行了全面的拆解剖析，包括账户、共识机制、智能合约、运行环境等；在性能上，作者使用白块（white-block）区块链测试框架对 EOS 系统的交易吞吐量、容错度、安全性进行了全面的测试；在经济模式上，作者主要分析了 EOS 代币系统以及相关经济行为等。

该论文是针对 EOS 系统的一个底层深度分析，主要为大家提供一个全面了解 EOS 工作体系的渠道。然而，该论文基本没有涉及数据层次的分析，主要视角集中在底层基本的层次架构，对于其具体的用户行为和生态行为涉及甚少，研究角度过于宏观。

(2) EOS.IO 特性分析

Huang Yuheng 等学者在文献【】中将 EOS.IO 的分析提升到了数据层面。通过收集数据集，作者对 EOSIO 区块链及其相关的去中心化应用进行了大规模的测量研究。结合数据，作者对 EOS 系统进行了描述性统计分析，包括资金转移、账户创建、合约调用，并且着重分析了账户之间的关系和行为相似性来识别僵尸账户和欺诈活动，最后也对 EOS.IO 的安全问题进行了描述。

该论文是对 EOS.IO 系统的一个较为综合性的分析论文，角度多样。然而，首先对于占据 EOS 生态系统很重要一部分的投票系统没有进行分析，其次虽然分析看上去较为全面，但是很多地方都是浅尝辄止，仅仅陈述数据表面的情况而没有进行进一步深度挖掘其特性与本质。

2.2 基于区块链的网络科学研究

2.2.1 网络的定义表示与基础特征

在现代社会中，人们生活在一个充满各种各样复杂网络的世界中，如社交网络、生态网络、交通网络等。网络由若干节点和连接这些节点的链路构成，表示诸多对象及其相互之间的联系。对网络的结构和性质进行分析建模，有利于我们深入了解网络结构与网络功能之间的相互关系和影响。

研究复杂网络的一种共同的语言就是图（graph）。其提供了一种用抽象的点和线表示各种实际网络的统一方法，将错综复杂的网络抽象为由点和线组成的结构，使我们更好地透过现象看本质。一般来说，一个具体的网络可以抽象为由点集 V 和边集 E 组成的图 $G=(V, E)$ ， E 中每条边都有 V 中一对点与其对应，且用 $|V|$ 表示网络中节点的个数， $|E|$ 表示网络中边的个数。用计算机实际分析网络的性质时，一般会采用邻接矩阵或者邻接表的数据结构来进行数据存储。若定义 $A \in \mathbb{R}^{|V| \times |V|}$ 为网络的邻接矩阵，则矩阵的第 i 行第 j 列的元素 $A_{ij} = 1$ 表示边 $(v_i, v_j) \in E$ ，反之若 $A_{ij} = 0$ 则表示边 $(v_i, v_j) \notin E$ 。

现实社会的网络通常具有复杂的拓扑结构，因此我们很难凭肉眼观察出一个网络的结构特征，而衡量结构的特征指标就尤其重要。从网络结构来看，首先我们关心的是网络中的节点是否连在一起，即衡量网络的连通性。基于连通性，许多拓扑性质都有了意义。如在刻画复杂网络“小世界”性质时，我们需要考量网络的平均路径长度与聚类系数。对于网络出入度均匀程度，我们可以用泊松分布或更适用于无标度网络的幂律分布来刻画。对于多个网络，可以进一步考虑包含更多结构信息的高阶拓扑特性来衡量网络之间的相似性与模块度，例如条件概率、余平均度和相关系数等，以及其他社区检算法。

批注 [刘13]: 加引用

批注 [刘14]: 社区检测，但是社区检测算法不算特性吧？

批注 [刘15]: 这一部分只写了一篇论文，有点不恰当，可以参考一下我在 blocksys 论文的 related work

2.2.2 基于区块链的网络研究发展

与传统的数据库相比，区块链技术天然地具有防伪造、公开等特点。数据的开放性与真实性对复杂网络研究有着重要意义，公有区块链的开放性为交易数据挖掘提供天然的有利条件。

陈厅等学者的文献[25]是第一篇对以太坊进行的系统性网络研究，也是第一篇较

批注 [刘16]: 前面的论文作者用英文这里用的中文

为全面的从复杂网络角度对区块链数据进行的分析。作者首先通过运行以太坊客户端收集了以太坊交易数据，再利用网络分析技术对以太坊的三个重要活动进行建模与分析，分别是资金转移、合约创建、合约调用，并获得了以太坊新的观察和见解。作者利用复杂网络的传统统计特征进行分析，如聚类系数、同配性系数、皮尔逊系数、连通图个数等。

在该论文，作者提出了一个新的收集交易数据的方式，并从以太坊数据中获得了很多有趣的结论和启发。该论文是第一篇较全面的用网络分析以太坊的论文，对我们 的工作也有很大的启发。当然，该论文依旧存在分析不够深入的问题，例如在分析各 类统计特征时，没有从数据特性上升到每个用户的行为，以至于无法使读者深入理解 以太坊生态。

第三章 EOS.IO 交易网络性质分析基本框架

本章将主要介绍 EOS.IO 交易网络性质分析框架。首先将简述该框架的流程图，分部分进行框架概述。随后针对框架中的每一部分所采用的方法进行详细叙述。

3.1 框架概述

这一节我们将简述本工作提出的 EOS.IO 交易网络性质分析基本框架。如图 3-1 所示，该框架包含如下四个主要部分：

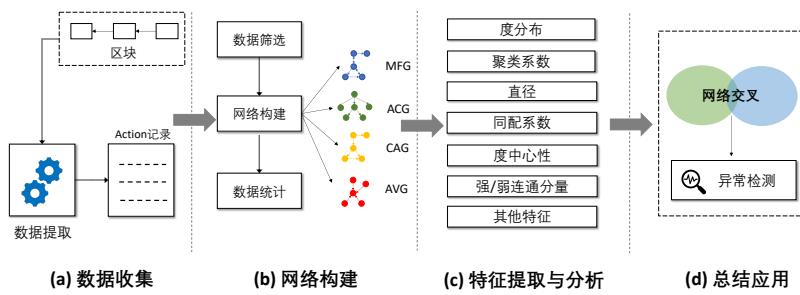


图 3-1 EOS.IO 交易网络性质分析框架图

- 1) 数据收集。根据事先确定好的时间区间，使用针对性的方法分别收集区块信息、交易收据、内外部操作信息，并整合为以操作为基本单元的操作记录。
- 2) 网络构建。利用上一步收集到的原始数据，根据不同网络的需求过滤出所需数据，指定节点与边的意义分别构建货币转移网络（Money Transfer Graph, MTG）、用户创建网络（Account Creation Graph, ACG）、合约授权网络（Contract Authorization Graph, CAG）与用户投票网络（Account Vote Graph, AVG），并针对构建后的网络进行描述性统计量的统计；
- 3) 特征提取与分析。针对构建的每一个网络，分别提取其网络特征，其中包括可视化其度分布，并计算其聚类系数、直径、同配性、度中心性等特征，找寻其强/弱连通分量等，并对每个网络联系其实际意义寻找其他特征。对找到的每个特征，联系实际进行深刻理解。
- 4) 总结应用。针对前文构建的网络和特征分析，总结其带来的实际价值，并通过多网络合并对比分析其中可能存在的异常行为和可能的应用

批注 [刘17]: 以操作为基本单元整合为

3.2 框架详解

本节我们将针对框架中的每一部分对我们实际采用的方法进行详细阐述，并进行一些专有名词解释。

3.2.1 数据收集

本小节将详细介绍我们从 EOS.IO 区块链中获得原始数据(raw data)的具体过程。图 3-2 描述了一个典型的 EOS.IO 交易流，即在从第 N 个区块到第 N+1 个区块的过程中，每个区块链对等节点的 EOS 虚拟机 (EOSIO virtual machine) 执行的一系列操作。我们从这个过程中收集三种类型的区块链数据，即区块信息、交易收据与操作记录。接下来我们会结合图 3-2，分别描述这三种不同类型的数据收集方法。

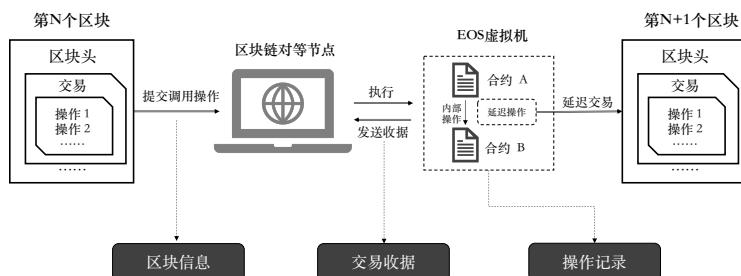


图 3-2 从 EOS.IO 交易流中收集原始数据示意图

(1) 区块信息

正如第二章 2.1.2 节所介绍的一样，EOS.IO 的区块信息是直接存储在区块链上的。每一个区块的信息都由两部分组成，一部分是存储着区块的基础信息的区块头，包含块生产者、时间戳等信息；一部分是交易与操作，每个区块可包含多个交易，每个交易可包含多个操作，每一个操作都是一次合约调用，其中包含着合约名、调用函数、函数参数与授权者等信息。然而需要注意的是，只有调用操作和延迟操作会被显式地存在在区块上，内联操作无法从区块上直接获得其具体信息。

目前，EOS.IO 的开发团队为用户提供了一个名为 Nodeos 的后台程序用于同步主链上的数据。当用户想同步数据时，其可以选择在初始块启动 Nodeos，或从一些备

份服务器（例如 EOS Amsterdam）下载快照备份的区块文件，再在指定的区块启动 Nodeos。通过比较，我们选择了效率更高的后者，激活指定插件并从 Nodeos 的 RPC (Remote Procedure Calls) 接口获取区块信息。

然而，只凭区块信息无法达到我们全面分析的目的。其不仅不包含内联操作，对于交易执行的具体信息也无法获取（例如交易执行时是否有其他的合约被调用等），因此我们还需要更多的数据。

(2) 操作记录

操作记录 (Action trace) 数据本质上是 EOS 虚拟机在生成每个操作时所产生实时数据。在操作记录中我们可以收集到有关于内联操作、延迟操作的详细信息，再与之前收集的区块信息相结合就可以得到关于交易的完整信息，例如调用了哪个合约，使用了哪个函数等。

操作记录只有在交易执行的过程中才得以被记录。因此，我们需要重演所有的交易，并在这个过程中收集操作记录。EOS.IO 开发团队提供了 state_history_plugin 插件和 mongo_db_pulgin 插件，可以将操作记录直接加入到数据库中。然而由于这些插件花费了多余的时间来与数据库引擎交互，会降低交易重演的速度。因此，通过比较，我们选择了速度更快的由 Xblock-EOS 提出的 history_file_plugin 插件，并且该插件可以直接将操作记录存为 json 格式，更方便我们下一步数据处理。通过该插件，我们使用 Nodeos 重演指定时间区间的交易并获取了完整的操作记录。

(3) 交易收据

批注 [刘18]: 和操作记录描述顺序换一下?

在 EOS.IO 区块链上，交易收据 (transaction receipts) 是记录交易的执行结果的，即说明该交易是否被成功执行。此外，交易收据还包含一个交易的资源 (CPU、内存等) 消耗信息。交易收据的收集能够使我们从所有操作中筛选出执行成功的那些操作，去除不必要的数据。

交易收据的收集和操作记录的收集方法很相似。我们依旧选择激活 history_file_plugin 插件，使用 Nodeos 将一个时间区间内所有的交易收据存储为 json 格式。

通过以上三种数据的整合，我们获取了一个一短时间之内完整的交易数据集，其

不仅包括每个区块的基本信息，对于每个交易的执行情况、每种不同类型的操作、每种操作内部所涉及的合约授权合约调用，也均整理完全。基于此数据集的分析工作也将更加全面和完整。

3.2.2 网络构建

货币转移、账户创建、账户投票、合约授权是发生在 EOS.IO 上的四大主要活动，本节将主要介绍我们如何运用数据来构建这四种不同的网络来全面理解 EOS.IO 生态，并对每个网络组成部分的实际意义进行具体阐述。

(1) 货币转移网络 (MTG)

定义货币转移网络 $MTG = (V, E, \omega)$, V 是所有节点的集合, E 是所有边的集合, ω 是映射到每条边上的权重。在本网络中, V 中每一个节点都代表着一个账户, E 的每条边都是一组有序的节点对且 $E = \{(v_i, v_j) | v_i, v_j \in V\}$, 边的顺序代表着钱财的转移。而 $\omega: E \rightarrow \mathbb{R}_+$ 赋予了每条边一条权重，在这里代表着转账的总金额（单位为 EOS）。例如，对于边 (A, B, t) , 代表着用户 A 向用户 B 总共转账了 t EOS, 这个总金额可能是由一个或者多个交易实现的。综上，货币转移网络 (MTG) 是一个带权有向图。

批注 [刘19]: 斜体

在建立过程中，我们通过为节点添加边来进行网络构建。对于一条转账信息来说，如果发送者和接受者之间没有边，则我们在他们之间添加一条权重为转账值的边。如果已有边，则将这条边的权重值再加上本条信息所对应的转账值。至此我们可以实现边的合并和权重值的分配。

(2) 用户创建网络 (ACG)

在 EOS.IO 中，每一个账户都对应一个用户，用户创建网络即是一个用户创建另一个用户的过程。定义用户创建网络 $ACG = (V, E)$, V 是所有节点的集合，每个节点代表一个用户。 E 是所有边的集合，依旧为有向边，边的顺序代表用户的创建过程，即边 (v_i, v_j) 代表用户 v_i 创建了账户 v_j ($v_i, v_j \in V$)。用户创建网络 (ACG) 是一个有向图。

由于每个账户都只能被创建一次，所以用户创建网络不含权重，且每条边均不可逆，具体性质我们将在第四章 4.2 节再做具体陈述。

(3) 合约授权网络 (CAG)

在 EOS.IO 中，每个合约调用的操作都需要进行用户授权。由于调用每个合约所用的函数涉及参数不同，我们无法统一地获得每个合约调用的主体用户，但是我们可以直接地得到每个合约调用操作的授权者信息。通过比较统计我们发现，绝大部分的合约所对应的主体调用用户和授权者是保持一致的。为了表述精确，我们在这里称此网络为合约授权网络，但是其在一定程度上也可以代表合约调用网络的性质。在此网络中，我们只考虑非系统合约的数据信息。

定义合约调用网络 $CAG=(V, E, \omega)$ ，与前文类似， V 、 E 、 ω 分别代表节点集合、边的集合以及每条边的权重。对于集合 E ，每条有向边 (v_i, v_j) 代表用户 v_i 授权合约 v_j ，即进行一次合约调用。由于 EOS.IO 中不对用户和合约做具体区分，每一个用户账户在制定了合约之后就变成了合约账户，因此 V 中的每个账户既有可能是真实的用户，也有可能是一个智能合约。一个合约通常会有很多用户向其授权，而用户基本不会被授权，因此由边的定义我们可以得出，一个入度很大的节点代表一个智能合约账户，而一个基本没有入度只有出度的节点代表一个真实的用户账户。每条边的权重代表授权总次数，例如对于边 (A, B, x) ，其意为用户 A 共授予 B 合约 x 次权限使其执行合约的具体操作。

对于合约授权网络的构建，我们依旧采用向节点添加连边的方法。对于一次授权调用来说，我们检查如果授权者和合约之间没有边，则进行边添加且设权重为 1。如果已经存在一条边，则为这条边的权重增加 1。如此我们成功创建了合约授权网络。

(4) 用户投票网络 (AVG)

在用户投票网络中，每一个账户都是一个真实的用户，用户通过投票来选出块生产者或进行其他决议。定义用户投票网络 $AVG=(V, E, \omega)$ 。节点集合 V 中的每一个节点代表一个真实用户；边集 E 是有向边的集合，每一条边代表着投票关系，即 (v_i, v_j) 代表用户 v_i 将票投给了用户 v_j ；每条边的权重 ω 代表投票总次数。例如对于边 $(A, B,$

v), 其意为用户 A 共授予 B 合约 v 次权限使其执行合约的具体操作。用户投票网络采取与合约授权网络相似的方法构建。

3.2.3 网络分析

本节将主要介绍我们在分析网络时使用的方法以及衡量网络特性时所采用的一些指标, 描述这些指标具体的实际意义等。

首先对于每个网络, 我们都会先进行网络可视化以对这个网络有较宏观的认知。由于每个网络数据过大, 我们采用抽样可视化的方法。对于货币转移、合约授权、用户投票三种网络, 我们直接进行随机取样。而对于用户创建网络, 由于其是一个以“eosio”为根的树状网络, 随机取样很容易破坏整个网络的连通性。因此我们采用了动态规划的并查集算法(Union-Find), 保证采样的每个点都拥有一条连接到“eosio”的通路, 这样即可尽可能地保持网络结构不变进行网络宏观观察。

其次, 对于每个网络, 我们都计算了一些复杂网络指标用以全方面衡量网络的性质。其中, 对于每个网络通用的指标包括:

- 1) 度分布 (Degree distribution)。在网络中, 与某节点相连接的边的数目称为该节点的度。度分布是对节点度的总体描述, 指的是网络中每个节点的概率分布。对于一个网络的度分布, 我们可以用泊松分布或者幂律分布来拟合, 从而对网络的度分布有较准确的认知。一般泊松分布存在在随机网络中, 幂律分布则广泛存在在真实网络中;
- 2) 聚类系数 (Clustering coefficient)。与一个节点相连的所有点成为这个节点的邻接点, 聚类系数用于描述一个点的临界点之间相互连接的程度, 也就是描述网络中节点集聚的程度。对于带权图, 我们一般计算加权聚类系数作为考量指标;
- 3) 直径 (Diameter)。在网络中, 定义两个节点之间的距离是他们之间所有路的长度的最小值。则一个网络的直径指网络中任意两个节点间距离的最大值。网络的直径可以直观反应网络的大小。由于一个网络不一定连通, 所以在计算一个网络的直径时通常先找到其连通分量们, 再找到连通分量的最大直径和最小直径作为原网络的考量指标;
- 4) 同配系数 (Assortativity coefficient)。网络的同配性用于考察网路中度值相近的节点是否倾向于互相连接。如果总体上度大的节点倾向于互相连接, 则称这个网络

批注 [刘20]: 所有的联通分量

是同配的，对同配系数大于 0。如果总体上度大的节点倾向于连接度小的节点，则称这个网络是异配的，对同配系数小于 0。同配系数本质上是一个基于度的皮尔森相关系数。

- 5) 度中心性 (Degree Centrality)。在网络中我们一般认为，一个节点的度越大则这个节点在网络中越重要，则我们认为这个节点的度中心性越高。研究度中心性较高的节点的具体信息有助于我们把握网络的主要社区，进而分析其相关特征。
- 6) 强/弱连通分量 (Strongly/Weakly connected components)。有向网络中，如果两个节点可以通过路径互相到达，则这两个节点强连通。如果有向图网络中的每两个节点都强连通，则称这个网络是一个强连通网络。有向网络的最大连通子网络即称为强连通分量。弱连通分量的定义与此相似，但对于两个顶点，其仅需保证至少一方可以到达另一方即可，并不要求双向。强连通分量可用于描述有向网络节点之间的交互关系，弱连通分量则用于考量有向网络的连通性。

以上均为对于每个网络都需要考察的较通用的网络指标。针对每个网络，我们还会考察其具体的指标，比如对于货币转移网络，我们又考量了 EOS 币交易额的累积分布，以研究 EOS.IO 生态系统中的贫富差距等。

3.2.4 总结应用

本节我们将通过网络结合的方式探寻本文的应用价值，例如异常检测部分。在此之前我们已经对单个网络进行了分析，而异常检测将不能只基于某个网络，多网络结合才能探寻生态系统中存在的实际问题。在实际操作中，我们将进行网络一对一结合或以对多结合，通过网络字段进行一一对应从而得出结论，具体内容将在本文第四章 4.4 节再进行详细说明。

批注 [刘21]: 这个总结是以前论文里的吗，怎么感觉跟这一章不太对应

第四章 实验结果与分析

本文在第三章中描述并总结了我们使用的 EOS.IO 交易网络性质分析框架，并分阶段介绍了我们使用的方法。本章将详细描述我们根据这些方法基于数据所做的实验并进行具体分析。本章分为四个小节，分别为数据获取过滤与基本统计、网络构建、网络分析以及总结应用。在每一小节，我们都会得到网络的一些性质和结论。

4.1 数据获取过滤与基本统计

在本文第三章的 3.2.1 小节，我们详细阐述了我们所采用的数据收集方法。由于 EOS.IO 秒级出块的特性，其每天产生的数据量非常庞大，如果时间区间选取过长会导致实验效率十分低下，但如果时间区间过短又会导致无法观察网络演化过程，网络特征将展现出偶然性。因此，在实验中，通过对比和推敲，我们最终选定并提取了 EOS.IO 主网前 1500 万个区块的数据，其包括了自 2018 年 6 月 6 日主网正式上线后三个月的完整数据。

前文提到，原始数据有三个部分综合而成：区块信息、交易数据与操作记录。综合得到的原始数据是以操作（action）作为基本单位，每个操作内既包含所在区块的基本信息，亦包括所在交易的基本数据。然而此数据集依旧过于庞大和冗杂，为了提升后续处理效率，我们对原始数据进一步进行过滤和筛选。

首先，原始数据会存在许多失效的操作。一方面可能是因为区块生产者本身原因，另一方面也可能是被主链丢弃的区块。因此第一步我们先筛选出所有成功执行的交易（操作）。随后，我们根据每个网络所需的数据，筛选出对应字段进行网络提取。过滤后，原始数据与各个网络的数据字段对应说明见表 4-1。

表 4-1 原始数据与各网络数据字段对应说明

网络名称	字段	字段说明
各网络通用	txId	交易 ID
	block_num	区块号
	block_time	区块时间 ³

³ 区块时间可作为操作时间

(续表 4-1)

网络名称	字段	字段说明
货币转移网络 MTG	from	转账发出者
	to	转账接受者
	quantity	转张金额 (单位: EOS)
用户创建网络 ACG	crator	创建者
	name	创建的新账户名称
合约授权网络 CAG	authorizations	授权者
	contract_name	合约名
用户投票网络 AVG	voter	投票者
	producers	被投票者

批注 [刘22]:

批注 [刘23]: 困惑的地方: 和下面标黄的几个, 有时加 s
有时候不加

原始数据的基本单位是操作, 我们针对每个网络分别统计了在这个时间段的总操作数量。此外, 我们亦分别统计了本时间段内每个网络内联操作的数量和占比, 具体统计信息见表 4-2。在 EOS.IO 中, 调用操作是由用户发出的合约调用命令, 而内联操作是智能合约内部, 或发生在智能合约之间的操作, 由另一个操作或操作本身直接触发。内联操作在总操作数量中的占比在一定程度上反映了用户在生态中的直接参与度。

表 4-2 各网络内联操作数量与占比

网络名称	内联操作数量	总操作数量	内联操作占比
MTG	4064790	8622288	47.143%
ACG	2860	302038	0.947%
CAG	2515729	331528545	0.759%
AVG	4	129804	0.003%

由表可以看出, 各个网络的内联操作占比不尽相同。除了货币转移网络的两种操作占比类似, 其余三种网络都有明显的差异。对于货币转移网络来说, 调用操作基本为用户主动的转账, 而内联操作可以是用户在进行一些赌博等游戏, 由智能合约自动进行转账。内联操作占了很大一部分, 也体现出智能合约在 EOS.IO 中应用广泛。

用户投票网络 (AVG) 的内联操作极少, 即几乎所有的投票都是由用户主动发起。这证明, 至少在 EOS.IO 前期, 投票操作基本都是公开透明显示在链上,

并且是用户主动进行的操作。另一方面，这也显示用户对这种 EOS.IO 独有的投票系统的配合。

用户创建网络（ACG）的内联操作也占有了很少一部分比例。其内联操作主要体现在例如“signuposeos”等 Dapp 会使用智能合约为用户提供收费注册的服务。此类 Dapp 在初期数量还较少，但是呈现出上升趋势，体现出 EOS.IO 生态逐渐繁荣。

合约授权网络（CAG）也拥有极少的内联操作。合约被授权执行或合约调用都需要用户的权限，所以绝大多数的操作均为调用操作，仅有一小部分是发生在智能合约内部进行自我调节等。

4.2 网络构建

在本节，针对每个网络，我们使用过滤整合后的数据进行网络构建，将具有相同顶点的边进行合并，并将他们的重复次数作为边的权重（其中货币转移网络的边权重是总转账金额），本文在第三章的 3.2.2 节有我们构建这四种网络的详细方法介绍。构建完成后各网络属性的统计值如表 4-3 所示，网络中的每一条边在原始数据中都是一个操作的体现。

通过对比表 4-3 的网络总边数和表 4-2 的总操作数量可以看出，货币转移网络和用户创建网络的这两个属性是完全一致的，但是合约授权网络与用户投票网络的数值稍有差别。这是因为一个合约的调用可能需要两个授权者共同授权，这两个授权者的权限可能是同等的（均为 active 权限）也有可能一高一低（一个 active 权限，另一个 owner 权限），但是都是真实的授权者，因此我们在网络构建时将一个操作拆分为两条边，分别对应每个授权者。而对于用户投票网络，在一个操作中一个用户可以给多个生产者同时进行投票，因此我们在处理时也将其拆分为多条边，这就造成了总边数和总操作数量在数值上的差异。

表 4-3 各网络属性统计值

网络名称	总边数	合并后边数	节点数
MTG	8622288	1370813	204841
ACG	302038	302038	302039
CAG	331530705	126918	35479
AVG	1519464	439154	28769

接下来我们将针对表 4-3 中每个网络属性的特点分别进行分析。

四个网络中结构最特殊的是用户创建网络 (ACG)。由于 EOS.IO 系统本身的机制，即先由系统用户创建用户，再由创建出的其他用户进行下一代的用户创建，因此 ACG 整个网络呈现树状结构，而根节点即为系统账户 “eosio”。换句话说，“eosio” 是所有其他用户的“老祖宗”，每个用户都有一条能够通向 “eosio”的单向路径。另外，由于每个账户只能被一个“父亲”节点创建，因此每个账户都只有一个入度（除根节点 eosio 外）。从表 4-3 中可以看出，合并后的边的数目与和合并前的边数目保持一致，这也验证了整个网络是弱联通的，且用户没有被重复创建。而节点数目刚好比边数小 1，因此也验证了 ACG 理论上的树形结构。

与 ACG 不同的是，货币转移网络 (MTG) 中的用户是可以互相转账或者多次转账的。合并后的边数大约变成了原边数的六分之一，证明账户之间多次转账的现象极为广泛。

对于合约授权网络 (CAG)，我们可以看出，合并后的边数和合并前的边数相差甚远，即在这个时间区间内有某些合约被同一个授权者重复授权（重复调用）非常多次。通过对原始数据进行排查，我们发现在原始的 33153070 条边中，有 316579248 条边所连接的合约账户是一个名为 “blocktwitter”的账户。我们进一步查阅原始数据并搜寻相关资料，我们发现这样的调用并没有实际意义，却因占用了大量的资源导致 EOS.IO 系统遭遇了 CPU 危机，使其他用户无法正常进行转账等操作。因此我们猜测这可能是一个攻击账户。

对于用户投票网络 (AVG)，从表中我们也发现，合并后的边数与合并前的边数差距很大，表明有一些账户经常重复为某一固定账户投票。不过这并不能武断地认为系统中存在不公平的刷票现象。从另一个角度想，这可能也引向另一种我们更希望看到的现象，即有些超级节点十分有能力且公正，受到了 EOS.IO 平台用户的广泛肯定，因此会不断地获得投票。在下一节 4.3.4 中我们还会具体分析这一现象。

4.3 网络分析

在本小节，我们将针对我们建立的四个网络，分别进行多角度的性质分析，以全面理解 EOS.IO 整个生态系统。对于每个网络，我们都会首先进行网络可视化以拥有一个对网络的宏观认识。之后，我们会对每个网络进行在本文 3.2.3 节

介绍的通用指标计算与分析。对于每个网络，我们还针对其网络特性设置了针对性的特殊属性分析。

4.3.1 货币转移网络（MTG）

首先我们对网络数据进行抽样可视化以获得直观认识。由于网络数据过大，我们从合并后的网络数据中随机选取了 10000 条边，生成的可视化图如图 4-1 所示。

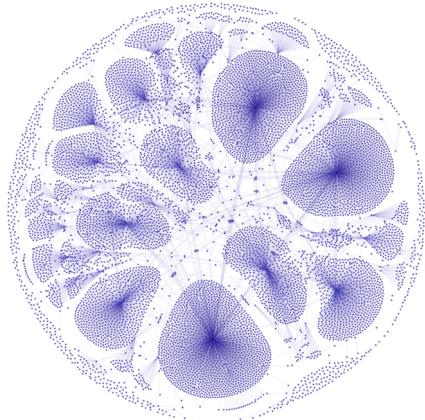


图 4-1 货币转移网络抽样可视化

MTG 的可视化网络有明显的社区结构。整个网络存在着数个大度节点，每个节点都类似一个社区中心，与周围节点进行频繁的交易互动。这些节点可能是 Dapp 中管理转账的节点或是大型交易所。此外，网络中还有较多度很小的节点，我们推测这些节点是个体用户，这些个体用户会互相进行交易，或与大型交易所进行交易。这些大量个体用户的存在也体现了由于 EOS.IO 系统的转账手续费为 0 导致的用户低入驻门槛。

(1) 网络度分布

图 4-2 是我们统计的 MTG 的总度、入度、出度分布图，即拥有不同度数的节点数量分布图。对于 MTG，网络中的度代表交易次数，入度代表账户收到转账的次数，出度代表账户发起转账的次数。为了便于观察分布特点，图中的横纵

坐标都进行了取对处理，即图 4-2 是双对数曲线图。从图中可以看出，MTG 的每个度分布图都符合幂律分布，即大度节点多，小度节点少。因此我们也在图中画出了 $y \sim x^\alpha$ 的拟合线（红色直线），由于是双对数坐标轴所以拟合线在图中呈直线， α 是幂律的指数， α 越大说明度分布越具有多样性。从幂律分布的图中也可得出与可视化相似的结论，即少数几个节点发生了大部分的交易，这些节点可能是作为交易所或 Dapp 收款转账方存在。

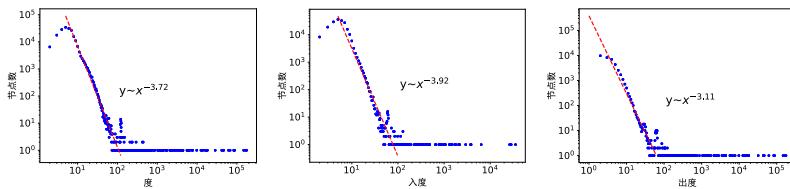


图 4-2 MTG 度/入度/出度双对数分布图

(2) 网络统计特征描述

接下来，我们针对 MTG 网络进行通用的网络统计特征计算并给出分析。每个指标的具体含义和计算方法已于第三章 3.2.3 节进行了详细说明，计算结果记录在表 4-4 中。

表 4-4 MTG 统计特征数值表

MTG	聚类系数	同配系数	最大直径	最小直径
	0.259	-0.338		
	强连通分量数量	最大强连通分量节点数	弱连通分量数量	最大弱连通分量节点数
	149536	1	1	204841

首先，MTG 的聚类系数是一个相对来说很大的值（0.259），这表明当账户 A 和账户 B 都和账户 C 有转账关系时，A 和 B 之间也很有可能有转账关系，即网络中存在许多类似的三角结构。同配系数为负，说明交易次数多的节点倾向于跟交易次数少的节点进行交易，而这也正对应了网络可视化中，每个大型的交易所或 Dapp 转账方都会与一群个体用户进行连接进行交易。

MTG 的最大强连通分量 (SCC) 节点数为 1，这是出乎我们意料的。这说明，EOS.IO 中，至少在此很长一段时间区间内，不存在用户之间相互转账的现象。这和 EOS.IO 的机制亦有关系，其是通过每个人持有币的数量进行的资源分配。

删除了：直径啊啊啊啊啊啊啊啊啊啊啊啊

对于富人来说，其没必要自毁系统，对于穷人来说，其不具备攻击系统的资源能力。因此没有用户通过“刷单”来扰乱系统秩序。

另一方面，MTG 的弱连通分量只有一个就是它本身，说明整个网络都是弱连通的，这其实是因为我们所取得时间区间是连续的，从主网刚刚启动开始就进行了数据统计，一切 EOS 币最初都是由系统账户“eosio”进行层层分发的，所以下续的转账交易也将基于这些通过层层递进拿到了 EOS 币的用户，因而形成了整个网络都是连通的局面。

(3) 其他特征描述

除了通用的网络指标，我们针对货币转移网络还进行了其他独有的特征分析。

A. 交易金额累积分布图

我们计算并画出了 MTG 的交易金额累积分布图。图中的节点(x,y)的含义为在网络中有占比 y 的用户总交易金额小于 x。图 4-3 描绘了计算结果：

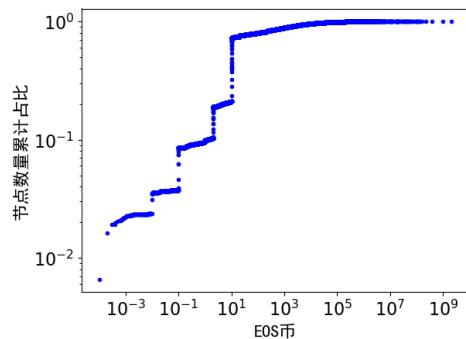


图 4-3 MTG 交易金额累积分布图

从图中我们可以看出，节点数量的累计占比是呈阶梯状上升的。如果我们将交易金额看做是用户财富的象征，则说明 EOS.IO 存在着鲜明的财富等级。一小部分人拥有着绝大部分的财富，大约 95% 的节点仅转移了 100 个或以下的 EOS 币。这表明尽管 EOS.IO 系统是公平的，但依然逃脱不了少数人手中掌握多数财富的现象。当然，EOS.IO 其独有的零手续费等系统机制使手握少数财富的穷人也依然可以平等地进行交易并拥有良好的交易环境。

B. 度中心性分析

在 MTG 中，我们统计并输出了前五个度最大的节点，即分析 MTG 的度中心性，探讨实体对 MTG 的作用和影响。表 4-5 展示了我们的统计结果。同时，由于区块链匿名性的特点，我们无法直接得到这些账户的身份。通过或是广泛搜索，或是在 ACG 中寻找“父亲”节点，或是研究原始数据中其交易的备注，备注，我们对他们的身份进行了大致的推测，结果也在表 4-5 中展示。

设置了格式：字体颜色：红色

表 4-5 MTG 度最大的前五个节点统计分析表

节点名称	度数	推测身份
pumpdumpwars	175686	Dapp 账户
eosio	163954	系统账户
eosballoonxx	148210	Dapp 账户
gm2tkmbyguge	86811	个体账户
guytiobzguge	86278	个体账户

和我们最初猜测的很相似，具有大度的节点很可能是交易所或某个 Dapp 的主要账户之一，“pumpdumpwars”和“eosballoonxx”两个账户验证了我们的想法。他们都各自关联一款 EOS.IO 平台上的游戏，且有着频繁的转账，我们推测其为对应 Dapp 的主要账户之一，可能是独有的转账账户，也可能是总的负责账户。“eosio”是系统的账户，最初的 EOS 币都是由其分发的。此外还有两个我们无法准确判断身份的两个账户，其创造于主网刚刚上线的那天，并且在之后几乎每天都有频繁的交易，我们推测其是类似 EOS.IO 的初始加盟者，即个体账户。

4.3.2 用户创建网络（ACG）

接下来我们分析用户创建网络的基本性质，首先我们对其进行抽样可视化。由于用户创建网络的特殊性，我们使用了动态规划的并查集算法（Union-Find）来保证抽样后的每一个节点都拥有到达“eosio”的通路。图 4-4 展示了我们抽样了 8000 条边进行可视化的结果。

如图所示，ACG 网络是一个树状网络，根节点为“eosio”，绝大多数节点都是由其直接创建，并分出自己的分支。同时，ACG 亦有明显的社区结构，除“eosio”外，亦有许多其他度较大的节点创建了一大部分的账户。

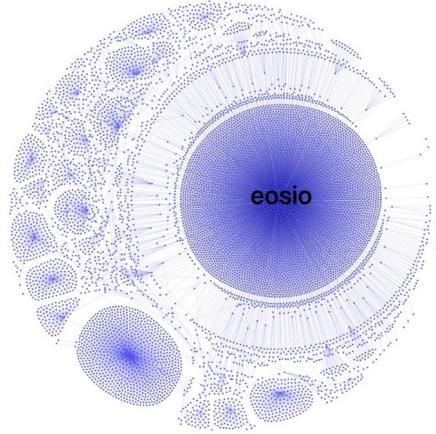


图 4-4 用户创建网络抽样可视化

(1) 网络度分布

基于可视化，我们猜测用户创建网络的度分布也是符合幂律分布的。图 4-5 显示了我们度分布的可视化结果。用户创建网络中的出度指一个账户创建新用户，入度指一个账户被创建。由于每个账户都只能被创建一次，即节点的入度都是 1 或者 0（对于“eosio”来说），因此我们在这里不做入度的相关分析。

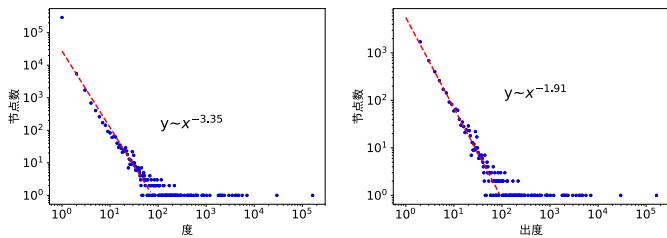


图 4-5 ACG 度/出度双对数分布图

从双对数分布图看，我们验证了我们的猜想，即 ACG 网络度分布复合幂律分布，即有一小部分节点有非常大的度，他们是创建用户的主体。在做了 $y \sim x^\alpha$ 拟合后，我们发现度分布的拟合直线斜率要远大于出度分布，这是因为度分布将入度考虑了进去，而除了“eosio”外，其他用户都有且仅有一个创建者，因此会导致度为 1 的节点数量极多。

(2) 网络统计特征描述

我们对用户创建网络进行了通用的网络统计特征描述如表 4-6 所示。同样因为网络的入度基本都是 1 或 0，我们在这里不做同配系数的分析与考虑。

表 4-6 ACG 统计特征数值表

ACG	聚类系数	同配系数	最大直径	最小直径
	0	/		
	强连通分量数量	最大强连通分量节点数	弱连通分量数量	最大弱连通分量节点数
	302039	1	1	302939

由于不存在两个被创建的账户之间互相创建的关系，因此其聚类系数为 0，即不存在这样的三元组。

同理，因为两个用户之间不能互相创建，所以最大强连通分量（SCC）的节点数为 1，强连通分量数量刚好和网络的节点数相等。弱连通分量（WCC）只有一个且节点数也与网络的节点数相等，即整个网络是一个弱连通网络，这和我们之前得出的结论相符。

(3) 其他特征描述

A. 用户创建时间演化分析

对用户创建网络，我们进一步想分析其创建用户的时间分布走向趋势，以及系统账户和其他用户在创建账户时的差别，我们作出了如图 4-6 的用户创建时间演化分析气泡图，其中橙色的代表系统账户“eosio”的创建，蓝色的代表其他所有用户，横坐标为日期，纵坐标代表在这一天创建的账户数量，同时创建的账户数量还会直观地反应在气泡大小上。

首先，从趋势上，我们看到在 2018 年 6 月 20 号之前，即在 EOS.IO 主网上线后的大概十天内是一个用户创建的高峰，随后便进入了稳定期，且数据分布较为随机。值得注意的是，系统账户“eosio”只在主网上线的第一天进行必要的用户创建，且第一天只有其进行活动，随后的账户创建的任务便都由其他用户所担任。可以说，系统账户在随后的 EOS.IO 用户迅速增长和 Dapp 的繁荣发展中起到的只是一个初期的引导作用。

删除了：用户创建网络的直径为。

设置了格式：突出显示



图 4-6 用户创建时间演化分析

B. 度中心性分析

我们对 ACG 进行了度中心性分析，除已知的“eosio”外，我们又统计并输出了其余度数最大的五个节点，即创建用户数量最多的六个节点。对于他们的身份，我们也通过搜索引擎、原数据观察等方法进行了推测，统计分析结果如表 4-7 所示。

表 4-7 ACG 度最大的前六个节点统计分析表

节点名称	度数	推测身份
eosio	163941	系统账户
eosdacstaked	29582	Dapp 账户
haytqnjug4ge	7047	Dapp 账户
prochainfree	5583	Dapp 账户
gy4dkmjzhege	4688	个体账户
1freeaccount	4676	Dapp 账户

我们通过相关 Dapp 搜索或新闻报道发现，表中大部分账户都与 Dapp 相关联，有的是创建的去中心化社区，有的是类似“糖果盒”的生态平台等。在 EOS.IO 机制中，用户可以在 Dapp 平台直接进行账户注册，并且用户免费由 Dapp 抵押付费，因此 Dapp 几乎支撑起整个用户创建网络，这也从另一方面说明了 EOS.IO 上 Dapp 的繁荣发展。

4.3.3 合约授权网络 (CAG)

接下来我们分析合约授权网络的性质。注意由于在绝大多数情况下，合约的授权者就是参与这个合约调用的主体用户，所以我们在分析时也会站在合约调用的角度进行分析。注意 EOS.IO 的用户和合约之间没有明显的界限，为了表述方便，我们将授权者描述为用户，将被授权者描述为合约。图 4-7 是我们随机抽取了 10000 条边进行可网络可视化结果。

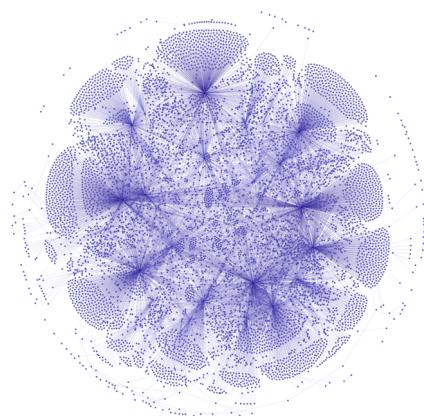


图 4-7 合约授权网络抽样可视化

从图 4-7 中可以看到，与前两个网络相比，合约授权网络的社区结构没有那么明显。虽然也存在着一小部分社区群体，即这一小部分用户只调用这一个合约，但是总体来说用户对合约调用的随机性很高，也体现了 EOS.IO 初期用户在多种智能合约或 Dapp 中不断尝试的过程。

(1) 网络度分布

同样，我们对合约授权网络进行了度分布可视化。在 CAG 中，入度是指每个智能合约得到授权进行执行，而出度是授权者为合约授予权限。

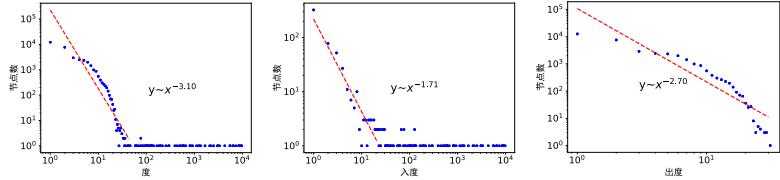


图 4-8 CAG 度/入度/出度双对数分布图

从图中看, CAG 的度分布依旧复合幂律分布。对于入度来说, 这意味着大多数合约只会被一小部分的用户调用, 这意味着并不是所有的合约都是被用户广泛所知并使用的。而对于出度来说, 其分布图并没有很长的尾巴, 即出度的大小并没有那么悬殊, 也就是说, 至少在 EOS.IO 初期, 大多数用户只会选择使用几个合约, 展现了他们对系统还没有太过熟悉。而即使是授权调用合约数最多的用户, 其授权合约的数量也并不多。当然, 这也和初期非系统合约的智能合约数量较少有关。

批注 [刘24]:

(2) 网络统计特征描述

同样的, 我们对 CAG 进行了通用的网络统计特征计算与分析。表 4-8 展现了我们的计算结果。

表 4-8 CAG 统计特征数值表

CAG	聚类系数	同配系数	最大直径	最小直径
	0.086	-0.160		
强连通分量数量	最大强连通分量节点数	弱连通分量数量	最大弱连通分量节点数	
35462	1	233	35086	

首先, 聚类系数非常小基本接近于 0, 说明几乎没有三元组的存在, 即如果账户 A 同时授权调用了合约 B 和合约 C, 则这两个合约之间几乎不会进行互相调用, 这可能的原因是智能合约基本不会复杂到需要多次调用才得以实现结果。

同配系数是负数证明大度节点更偏好于连接小度节点。直径

设置了格式: 非突出显示

最大强连通分量节点数是 1, 说明不存在两个合约之间互相授权的现象, 但是这并不能排除有授权给自身的现象。在观察原始数据时, 我们经常发现有合约其授权者和合约本身是使用的同一账户的情况。这种情况在 EOS.IO 系统的初期比较多, 而到了后期, EOS 用户逐渐增加, 这一现象由于不方便用户管理而逐渐

消失。最大弱连通分量的节点数为 35086，占整个网络的 98.89%，也就是说虽然弱连通分量很多，但是整个网络的大部分都是连通的。

(3) 度中心性分析

在这里我们不做合约授权次数的统计，是因为有类似前文“blocktwitter”这种攻击型合约的例子。出于对合约活跃性的考量，我们选取了 CAG 网络中度最大的五个节点，或者说拥有最多用户的合约进行了统计分析，同时我们通过 eospark【】等网站探寻了其合约具体用途，统计结果见表 4-9。

批注 [刘25]: 这个加脚注就好了

表 4-9 CAG 度最大的前五个节点统计分析表

节点名称	度数	合约用途
ridlridlcoin	9741	进行 RIDL 币的转账操作
prochaintech	9366	负责 PRA 般若（一个数字广告生态系统）的转账等功能
poormantoken	8789	Poorman 代币的转账、确认等操作
wizznetwork1	8776	一个聊天软件账户的转账功能
trybenetwork	8617	一个代币化知识共享网络的转账、抵押等多种操作

从表 4-9 中我们可以看到，在 EOS.IO 平台上已经逐渐兴起了许多社区或代币。基于 EOS.IO 系统的种种优点，他们使 EOS.IO 的生态系统不断多样化全面化，各个拥有不同功能的小型系统正在繁荣发展，同时满足具有不同用户的需求。

4.3.4 用户投票网络 (AVG)

最后我们探讨用户投票网络的具体性质。图 4-9 为我们抽样了 10000 条边后可视化的结果。从直观上我们发现，几乎每个类似超级节点的大度节点都具有一小簇只为它投票的小度节点，当然，我们也不能武断地认为这其中的操纵行为。另一方面，绝大多数用户都是随机投票的，具有很大的不确定性。

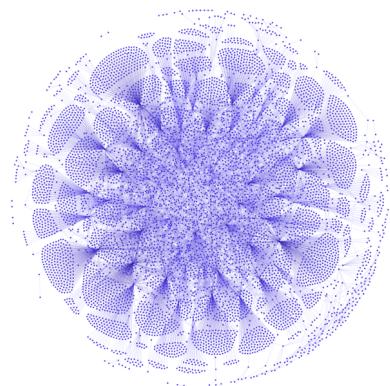


图 4-9 用户投票网络抽样可视化

(1) 网络度分布

图 4-10 是我们对用户投票网络进行的度分布可视化图。对于用户投票网络来说，入度是自己获得的投票，出度是自己投出的票。

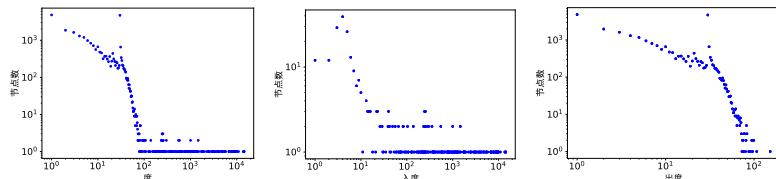


图 4-10 AVG 度/入度/出度双对数分布图

从图中我们可以看到，用户投票网络并没有严格地符合幂律分布，因此我们也没有进行指数拟合。但是，其仍然具有与幂律分布类似的性质特点。对于入度来说，能够获得极多票数的账户占据非常小的比例，这些都是能够担任超级节点责任的账户。对于出度来说，大部分节点的投票数量都很少，一次投很多票的账户占非常小的比例。由于投票时需要进行 EOS 币抵押，因此这和贫富差距也有一定的关联。

(2) 网络统计特征描述

表 4-10 为我们对用户投票网络进行的通用网络特征分析。

表 4-10 AVG 统计特征数值表

AVG	聚类系数	同配系数	最大直径	最小直径
	0.066	-0.221		
	强连通分量数量	最大强连通分量节点数	弱连通分量数量	最大弱连通分量节点数
	28750	1	3	28766

用户投票网络的聚类系数几乎接近于 0，即网络中几乎不存在三元组结构，也就是说当一个用户给两个节点投票后，这两个节点几乎不会再互相投票，这可能和超级节点之间的竞争有关。同配系数为负，即度大的节点偏好向度小的节点投票，即竞争力相对较少的节点，而不会投给同样拥有很多票数的另一度大节点。

直径

最大强连通分量节点数为 1，说明不存在互相投票刷票的情况。最大弱连通分量的节点数只比网络的总节点数少 3，证明整个网络的绝大部分都是弱连通的。而对于剩余的几个独立节点，其可能是创建了投票的操作但是并未进行实质性的投票，但是总体来说，用户投票网络中的用户都积极使用了自己投票的权利。

设置了格式：非突出显示

(3) 用户投票时间演化分析

为了观察用户投票数随时间的变化趋势与走向，我们作出了如图 4-11 的用户投票网络时间演化图。

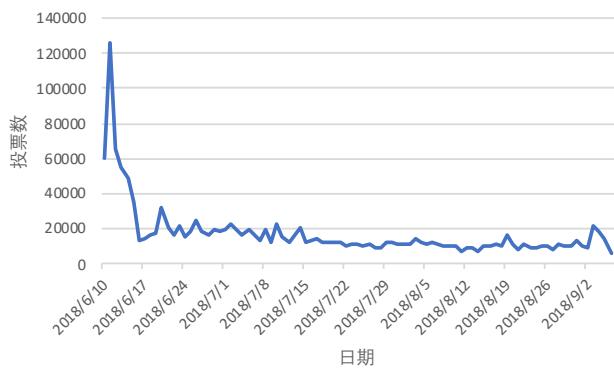


图 4-11 用户投票网络时间演化图

从图中我们可以看到，在 EOS.IO 主网上线第一周，投票数达到了一个峰值，而后下跌到一个稳定值，最后在每天一万票上下波动。初期的投票顶峰是因为接

照规则，投票率需达到 15% 主网才会启动，以选出 21 个超级节点。因此在初期，用户们为了早日激活主网，投票率一度很高。主网激活后，投票数日益下降，且数量与初期相差甚远，因此除非投票人集中力量为一个节点刷票，否则初期超级节点的地位几乎难以撼动。

4.4 总结应用

前文我们针对每个网络从多个角度进行了性质分析，并得出了一些针对性的结论。在本节，我们将探寻应用价值，通过将构建的网络相结合来寻找其中存在的一些异常现象。

由于用户创建网络（ACG）中包含这一个时间段中网络中创建产生的所有用户，我们将从 ACG 开始入手。本章的表 4-3 显示，在 EOS.IO 的前 1500 万个区块中一共有 302039 个用户。而在我们将这些用户与其他三个网络中的用户一一比对时，我们发现这个庞大的数字下的不寻常之处。统计显示，有 89301 个用户自创建以来没有进行过任何活动，占总数的 29.57%。虽然在这其中可能有部分新创建的账户在以后，即我们统计的时间区间以外，有着其他操作，但这些毕竟只占少数，总体上仍显示出在这表面繁华背后的沉寂。

用户投票决策一直是 EOS.IO 的一大特点之一，然而我们将用户投票网络与用户创建网络结合发现，实际参与了投票的用户仅占全部用户的 9.52%，所以在 EOS.IO 主网上线初期，经历数天的时间才选举出 21 个超级节点并进行主网激活。在这其中可能包含着新用户对机制的不了解，或对投票安全性的担忧。

在之前针对货币转移网络的分析中，我们发现了 EOS.IO 中存在着极不均衡的贫富差距现象，同样证明这一现象的是通过综合表 4-3 和 4-5 我们发现，MTG 中度数最大的五个点几乎涉及了网络中一半用户的转账操作。此外，我们将用户创建网络与货币转移网络相结合，发现在 MTG 中，有 178051 个用户与自己的“父亲”（即创建自己的账户）有着交易，占总体的 86.92%。而有 5308 个用户与自己的“儿子”（即自己创建的用户）有过转账记录，占总体的 2.59%，占有创建用户记录的用户群体的 9.44%。虽然之前我们得出了货币转移网络中不存在互相转账的情况，但是根据这个数据，我们可以想象一个用户创建了多个子用户，并且子用户将自己得来的资产不断转移给父用户，或者父用户将自己得来的资产分给子用户。而这无疑是一种异常的情况。

第五章 总结与展望

本章将对本文的研究结论与研究创新点进行总结，并指出本文的局限性，对未来的研究工作提出展望。

5.1 工作总结

本文基于复杂网络，对 EOS.IO 区块链交易网络性质进行了完整且系统地分析。首先我们通过使用针对性的方法分别收集 EOS.IO 相关的区块信息、交易收据、内外部操作信息，进行数据处理与过滤工作，随后针对 EOS.IO 系统上四大活跃生态，即货币转移、用户创建、合约授权与用户投票，分别构建了网络。针对每个网络，我们不仅分析了其基本的网络统计特征，还根据网络的特性提出并分析了其独有的性质以获得更全面的了解。此外，我们还将多网络进行结合分析，以寻找其生态系统存在的异常情况，对管理者给予启示。

通过分析，我们总结四大生态分别具有以下的性质：

- 1) 货币转移：EOS.IO 的零手续费等设置促进了个体用户对系统的参与和融入，网络中存在许多互相转账的三元组促进了交易网络的繁荣，同时 EOS.IO 与资源挂钩的系统使网络中不存在相互转账的情况。然而网络中依旧存在很严重的贫富差距问题，并且存在着许多父用户和子用户转账的问题，其可能是用于为所在 Dapp 刷量而存在；
- 2) 用户创建：用户数量众多，系统合约在这种繁荣中只起着初期的引导作用，Dapp 账户支撑起了绝大部分用户的创建。然而存在着大约四分之一的“沉默”用户自创建以来从未进行过任何生态系统上的活动；
- 3) 合约授权：EOS.IO 的生态系统繁华，智能合约使用广泛，且其拥有许多各种功能的小型系统可以满足用户的不同需求，然而存在着有人利用合约授权进行资源消耗从而进行系统攻击的现象；
- 4) 用户投票：超级节点之间存在着较为激烈的竞争，参与投票的用户基本都拥有自己的人选，几乎没有投空票的情况，然而从整体上来说用户对于投票系统的参与度并没有很高。

批注 [刘26]: 活动？

批注 [刘27]:

据我们所知，本文是第一篇从网络角度对 EOS.IO 进行的全面分析，并且从简单的数据上升到了用户行为与系统生态，此外也是第一篇涉及投票系统的分析。

我们还处理并提供了完整的数据集，对提取方法也做了详尽的描述。本文所进行的异常行为探讨也将为管理者或参与者提供一定的启示。

5.2 研究展望

目前对 EOS.IO 的系统研究还处在初级阶段，本文从网络结构的角度进行了全面系统地分析且对上升为对用户行为和生态的理解。然而，本文仍然有拓展和改进的空间：

- 1) Dapp 在 EOS.IO 上线后得到了全面的发展，且基于 EOS.IO 的 Dapp 发展迅速猛烈，成为整个生态系统不可缺少的一部分。因此，一些异常行为也可以针对不同类型的 Dapp 进行分析，例如其是否具有排行榜刷榜现象等等；
- 2) 在分析中我们发现，资源分配、买卖与使用作为 EOS.IO 独有的体系，也具有一定的分析价值，和体系中的各个环节息息相关。我们可以针对这个角度进行探寻，并探索其中的异常行为，例如是否有内存价格操纵现象等；
- 3) 以太坊作为早期支持 EOS.IO 的一大区块链平台，和 EOS.IO 有许多相似的地方，我们可以将两者进行对比，从机制上或是数据统计结果上，使读者对 EOS.IO 的认识更加全面。

参考文献

致 谢

(* 致谢，单独占一页。对在论文工作中对你有支持、有帮助的单位或个人表示感谢。 *)

附 录

毕业论文成绩评定记录

指导教师评语:

成绩评定:

指导教师签名:

年 月 日

答辩小组或专业负责人意见:

成绩评定:

签名(章):

年 月 日

院系负责人意见:

成绩评定:

签名(章):

年 月 日