星云指数黄皮书

星云研究院

2018 年 6 月
版本号:1.0.1

目录

1 概要

区块链技术带来的“去中心化”理念正在被用于越来越多的场景。作为区块链技术的起源，比特币已经证实了去中心化对于数字资产的非凡意义；更进一步的，以太坊证实了去中心化对于分布式应用的重要性；越来越多的区块链项目正在探索去中心化这一理念在更多场景及应用下的价值。

不难发现，“去中心化”理念的背后，是区块链系统中的开放性（Openness）与匿名性并存。

然而，区块链系统的这种特性在一定程度上造成了价值衡量体系的缺失 [1]。这反映在两方面。首先，由于区块链系统的匿名性，很难推断多个属于不同账户的数据和资产是否属于同一个用户，这导致了区块链系统中不能构建类似 HTTP Cookie [2] 的机制，也很难通过传统的数据分析技术从不同的角度分析用户特征；另一方面，区块链系统的开放性又使得其面临着很强的操纵挑战，价值衡量体系很容易受到各种针对性的操纵攻击，这不同于任何封闭的、独立的价值衡量体系。

我们认为有效的价值衡量体系是区块链生态能够繁荣发展的基础，价值衡量体系的缺失或无效必然会限制整个区块链行业的发展。

首先，随着协作规模不断变大，并且对效率的要求不断升级，我们需要一个价值衡量体系来为区块链系统以及区块链系统上的应用、数据和账户的价值提供可判断的量化标准，否则要么因为无法量化评估而影响效率，要么因为评估失当引发不公平甚至导致失控。

其次，区块链尚处于发展阶段，区块链上大量数据及资产的价值等着去发现。有效的价值衡量体系将使得冰山下的部分得以露出，催生新兴应用甚至领域出现。比如区块链上的借贷和征信类服务、数据搜索和个性化推荐、原生跨链交易和数据交换等，价值衡量体系将使这些领域突破瓶颈。

最后，生态建设需要有效的激励和健康的发展方向。有效激励的基础就是有效的价值衡量体系。如果没有价值衡量体系，甚至价值衡量体系是歪曲的，那么就会导致激励机制失效，整个区块链系统不可避免地走向灭亡。

综上，一个区块链价值衡量标准需要具备三个特点：

- 真实性 一个好的价值衡量标准应该能够准确反映出区块链经济系统的特征，这样才能在相应的领域具有足够的公信力；
- 公平性 价值衡量标准为相应的激励提供了依据，因此，这一依据必须足够公平，

才能防止作弊或操纵带来的“劣币驱逐良币”现象；

- 多样性 需要使用数据及数字资产价值的场景可能是多种多样的，其使用方式及对应的激励方式不尽相同，因此相应的价值衡量标准既不能脱离应用场景，亦要满足前述的真实性及公平性。

星云指数（Nebulas Rank）将是一个满足以上三个特点的区块链价值衡量标准。

为了体现真实性，我们参考了诸多指标，最终我们定义星云指数为：衡量账户地址对于区块链这一经济系统的贡献度。

本质上来说，区块链作为一个经济体，并不违背经典货币理论。我们认为区块链系统之上的加密数字货币应该具备基本的货币属性，并且加密数字货币的价值源于其流通性。因此，加密数字货币的交易记录是衡量加密数字货币这一经济体的有效数据来源。更进一步的，我们认为每个账户发起的每一笔交易都在一定程度上增加了加密数字货币的流通性。微观角度看，每个账户的交易行为都最终反映在了区块链系统的价值中；从宏观角度看，我们将所有账户地址的星云指数的和定义为整个区块链系统的经济总量。

为了验证星云指数设计的有效性，我们在基于以太坊的链上数据中计算了所有账户的星云指数之和，并与 Coinmarketcap.com 中同期的以太坊市值进行了对比。我们的对比表明了二者具有很强的正相关性（0.84），即星云指数既能够在微观层面衡量每个账户对经济系统的贡献，亦能在宏观层面反映整个经济系统总量的变化。

为了保证公平性，我们设计了能够有效抵抗操纵的计算函数。并论证出了星云指数在抵抗操纵方面可达到的性能。

在星云指数的理论基础上，为了满足多样性的需要，我们将星云指数分为核心星云指数（Core Nebulas Rank）和计算基于核心星云指数的扩展星云指数（Extended Nebulas Ranks）两部分。

核心星云指数针对区块链中不同账户对于区块链系统的贡献度给出了计算方法。其计算基于两个参考因素：其一，账户在一定时期内的资产中值；其二，账户在一定时期内的出入度衡量。

扩展星云指数则基于核心星云指数来构建，针对区块链生态中各应用可能需要的价值尺度给出了不同的计算方法，以便更符合不同场景的实际需要。并举了几种扩展星云指数的计算方法作为参考，例如：如何根据核心星云指数对智能合约进行排名；如何将星云指数拓展到多个维并给予不同的权重等。

本黄皮书除了给出理论论证，还解决了几个星云指数落地时必须面对的问题，例如星云指数是否上链，星云指数的计算如何更新等。对星云指数实际落地给出了具体的工作方向。

特殊提示：本星云指数黄皮书作为专项讨论星云指数的黄皮书，对星云技术白皮书（2018 年 4 月发布的 1.02 版本） [3] 中星云指数相关章节进行了大幅度的升级和拓展。相对于一年前的概念论证，经过一年的深入思考与实际验证，我们有信心和能力设计出更为严谨的算法，并对星云指数的更多实际细节问题提供明确的解决方案或方向。为了方便阅读，我们将使用实线框高亮解释技术白皮书提及过、并且在本黄皮书中有升级的相关技术点。

2 背景及相关技术

本章主要介绍区块链发展背景以及相关技术。鉴于目前的区块链生态之上价值衡量尺度的缺失，我们进一步讨论了经典排名算法在区块链领域应用的情况以及其缺陷。

2.1 区块链发展现状

2008 年 10 月，中本聪（Satoshi Nakamoto）公开发表比特币白皮书 [4]。比特币作为区块链的太初应用，践行了其作为“一个去中心化电子现金系统”的初衷。比特币的产生不依赖于任何机构，而是根据特定算法，依靠大量计算产生，保证了比特币网络分布式记账系统的一致性。

通过特定的脚本语言，我们可以利用比特币实现第三方支付交易、高效小额支付（efficient micro-payments）等功能。此后涌现出许多以比特币为参照的试验品，在提供基本的货币属性外，作出了更多的尝试。例如早期的域名币（Namecoin） [5] 提供了一种去中心化的域名系统 DNS。以及基于“货币染色（Colored coins）”的开放资产项目（OpenAssets） [6]，其本质都是模仿比特币，利用可追溯性，又复制了一份智能资产。

很遗憾的是，比特币脚本语言的设计存在很多缺陷，如仅支持较少指令，且并不具备图灵完备性，这使得其应用场景受限。

随着区块链技术研究的不断深入，涌现了更多后继者，尝试拓展和添加更多与应用程序相关的功能。其中最令人瞩目的实现是以太坊（Ethereum） [7]。以太坊突破性地提供了图灵完备的智能合约（Smart Contracts），从而大幅拓展了应用场景。

智能合约是区块链系统中可以用技术手段来强制执行的合约，以太坊智能合约运行在以太坊虚拟机（Ethereum Virtual Machine）上，以太坊虚拟机不受任何实体控制，通过共识算法来验证合约本身及其输出的完整性。

基于以太坊的智能合约，人们得以开发能实现复杂功能的分布式应用（DApp）。

除了基本交易功能外，DApp 为众多领域提供了解决方案，如投票、众筹、借贷、知识产权等。

以太坊成功拓展了区块链的可能性，但以太坊缺少价值衡量标准，导致潜在杀手级应用的落地和推广存在困难。

对于支持智能合约的区块链系统，其账户通常包括外部账户（Externally owned account，EOA）和智能合约账户，对于这两类账户目前尚缺少合理的评价指标。同时，在诸多交易以及智能合约的调用过程中，隐藏着难以估计的信息。后者相比传统交易数据，往往具有更多维度，因此也无法使用传统价值衡量标准评估。

其实，早在 2015 年，Chris Skinner 便提出了“价值网络（value web）”的理念 [8]，其中提到，一个价值经济系统（Value ecosystem）应包括价值交换（value exchanges）、价值存储（value stores）以及价值管理系统（value management systems）三部分，缺一不可。同时 Chris 也指出，对于比特币等数字加密货币来说，价值网络的衡量相比传统社会价值有着明显不同，挑战更大。

2.2 图中节点的排名算法

由于引入了智能合约，当前以以太坊为代表的新一代区块链项目不仅仅是电子货币交易平台，而是在此之上建立了复杂庞大的经济体系。尚不存在一个合理的方式去评估链上实体（例如用户地址）的价值。例如，目前我们无法知晓哪些实体对整个区块链生态贡献较大？又应如何衡量这类贡献？

在此，需要先提到一个大家比较熟悉的传统互联网领域价值衡量标准：PageRank 算法 [9]。作为 Google 早期的核心算法，PageRank 设计初衷是用于解决链接分析中网页排名问题。随着国内外学者的深入研究，PageRank 算法被广泛应用于其它方面，例如学术论文的重要性排名、网络爬虫、关键词句抽取，以及基于 PageRank 的社交用户影响力排名研究等等。

学术界已将 PageRank 算法应用于区块链的研究，比如 Fleder、Kester、Pillai 等人 [10] 使用 PageRank 来帮助发现感兴趣的比特币地址，并分析这些地址的活动。但他们的主要方法仍然是人工主观分析，PageRank 只起到辅助作用。

作为诞生于互联网 2.0 时期的经典排名算法，PageRank 算法应用于在线社会网络影响力评估存在局限性。

此后涌现出了一些其它在 PageRank 算法基础上进行改良的研究，其中较为著名的包括 LeaderRank 算法 [11]，它是 PageRank 的一种拓展形式。在 PageRank 中，每个节点都有相同的随机跳转概率，而 LeaderRank 是对跳转概率简单有效的改进。通过在网络中添加背景节点和加权的双向链接，可以使得不同节点具有不同的随机跳

入和跳出概率。

而 LeaderRank 也存在一定局限性，其只考虑了节点之间的关系（即网络结构），通过迭代得出最后影响力排名，缺乏对用户行为的衡量。

需要指出的是，PageRank 类排名算法无法应对女巫攻击（sybil attacks）[12]，女巫攻击是指攻击者通过创建大量的假名标识来破坏对等网络的评价系统，从而获得虚假的高重要性评分。

和星云指数最相关的项目是 NEM [13]，不同于比特币的 Proof-of-Work 以及以太坊的 Proof-of-Stake 共识策略，NEM 设计了 Proof-of-Importance 共识机制，其中排名算法 NCDawareRank [14] 利用了网络拓扑的社群效应。Proof-of-Importance 使用 SCAN [15] [16] [17] 作为社群聚类算法。虽然社区结构在交易网络的确存在并且可以帮助应对欺诈节点，却无法保证同一个实体对应节点一定可以映射到相同社群，因此利用社区划分的结果会提供一定的可操纵空间。

2.3 对抗操纵

提升可信性，即具备抵抗操纵的能力，是核心星云指数最重要同时也是最具挑战的目标。

Hopcroft 等人发现，在存在恶意操控的情况下，PageRank 无法有效衡量用户的影响力 [18]。Zhang 等人指出，在社交网络中，即便建立了节点的影响评价指标，攻击者仍然能够有效削弱其他非僵尸用户的影响力 [19]。针对 PageRank 算法的典型女巫攻击——环形攻击（two-loop attack）就是一例。以女巫节点 s 发起到目的节点 v_j 以及从 v_j 返回到 s 的 Random walk，尽可能多地访问其它节点，从而增加 v_j 的访问概率。

这是因为从本质上而言，PageRank 类算法依赖于网络拓扑结构对用户进行排名，而在对称网络（symmetric Network）下，恶意操纵者非常容易通过构建镜像网络来获取同样甚至更高的影响力 [20] [12]。

在区块链生态中，部分恶意操纵的手段通常有以下几种:

1. 环形转账，攻击者沿环形拓扑，让同一笔资金不断流过对应的边，以提高边权；
2. 向其它任意账户转钱，提高出度，并且提高资金流出的传播性；
3. 控制多个账户形成独立分支，伪造中心节点；
4. 频繁同权威交易所账户交易，多次在交易所账户中取入取出同一笔资金，获得较好的网络结构位置。

因此我们在设计核心星云指数时，需要考虑到上述情况，来保证核心星云指数的公平性。

3 区块链经济模型

加密数字货币无论是作为交易媒介还是智能资产，都被赋予了经济意义，因此一个合理的经济模型有助于我们构建区块链之上的价值衡量标准，这也是核心星云指数的目标。本章先介绍了加密数字货币的数学表示，然后将加密数字货币应用在简单但广为认可的货币模型上，并且我们在其中引入核心星云指数作为重要参数。

3.1 加密数字货币表示

作为加密数字货币，其与传统经济体的一个重要差别在于所有交易都是可追踪的，这为我们定量研究每一笔交易对整个经济系统的影响提供了强力的数据支持。

一般地，一个加密数字货币系统可以描述为一个二元组 $(\mathcal{L},\mathcal{U})$，其中 $\mathcal{L}$ 为账本系统，$\mathcal{U}$ 为数字货币的用户集合。账本系统 $\mathcal{L}$ 可以被描述为一个三元组，即

$$\mathcal{L} = (\mathcal{A}, \mathcal{D}, \mathcal{T}) \tag{1}$$

其中 $\mathcal{A}$ 为账户的集合，$\mathcal{D}$ 为初始状态下各个账户余额的集合，$\mathcal{T}$ 为交易记录的集合，每条交易记录为一个四元组，即

$$\mathcal{D} = \{a \to d, a{\in}\mathcal{A}, d{\in}R^*\} \tag{2}$$

$$\mathcal{T} = \{(s, t, w, \tau)\} \tag{3}$$

其中 $a \to d$ 表示对应账户 a 下的余额 d（d 为大于 0 实数，在初始状态下余额为 0 无法创建账户）。s 为交易的发起地址，t 为交易的目的地址，w 为交易的金额，τ 为交易的时间。

一个账户受控于相应的用户，该用户能够发起交易，我们记为

$$u \ \mathtt{dom}\ a. \quad u \in \mathcal{U}, a \in \mathcal{A} \tag{4}$$

一个用户可以控制多个账户，即

$$A(u) = \{\forall a \in \mathcal{A} : u \ \mathtt{dom}\ a\} \tag{5}$$

而一个账户仅能被一个用户所控制，即

$$\forall u_1, u_2 \in \mathcal{U} : A(u_1) \cap A(u_2) = \phi \tag{6}$$

需要注意的是，上述模型是任意加密数字货币系统的合理简化。我们在模型中未区分链上数据或链下数据、未引入成交价格、未引入智能合约的调用等。此外有一个特殊情况是中心化交易所的账户。通常来说，一个交易所账户会被分配给多个用户，每一个用户使用不同的地址进行交易，在交易所内的交易由交易所在中心化的数据库中进行记录，而不会记录在链上。这就意味着我们无法仅仅通过链上数据获取交易所内的交易记录。但是可以在交易所的配合下获取相应的数据，进一步将交易所账户映射为不同的多个账户，从而使用上述的模型进行描述。

3.2 加密数字货币模型

尽管数字货币和传统的商品货币以及法币存在较大区别，但经典货币理论仍然具有指导价值，可以借鉴。加密数字货币作为新型经济体的通行货币 [21]，承载了货币的属性，具有三种货币所具有的职能：交换媒介、计价单位，以及价值储藏手段。

在此我们建立一个简单且经典的货币模型来帮助理解星云指数的物理意义。

首先，我们需要给出加密数字货币生态系统中衡量“流通性”的指标。

需要加以区别的是经济学中经常出现的另一个概念：流动性（liquidity）。流动性用以说明一种资产兑换为交换媒介的容易程度。由于货币本身是经济中的交换媒介，所以货币是最具流动性的资产。

在星云技术白皮书 [3] 中，我们使用了“流动性”一词。然而，“流动性”一词缺乏严格的定义，即使在经济学中，这一涵义也是十分广泛的。例如，在《新帕尔格雷夫金融学辞典》中，对流动性解释的专门词条包括了完全不同的三个方面。兰德尔 · 克罗兹勒 [22] 指出，在过去六个月里，有 2795 篇独立的文章谈到了流动性，但流动性是什么含义，大概有 2795 种不同的说法。在本黄皮书中的“流动性”，统一特指货币流通速度（velocity of money），即指单位货币在一定时期内的周转 (或实现交换) 次数。

我们用货币流通速度 [23] 表示加密数字货币的周转速率，即单位数字货币在一定时期内（在本文中为一天）的周转次数，用 V 表示。根据经典的货币数量论，数量方程式表示如下：

$$M \times V = P \times Y \tag{7}$$

其中 M 表示整个经济系统中的货币数量，V 是货币流通速度，P 表示物价水平（用每单位经济产量的货币单位来衡量，所以货币价格是 $\frac{1}{P}$），Y 表示真实经济产量（真实 GDP）。该方程式说明，货币数量乘以货币流通速度等于产品的价格乘以产量。

对于货币总量 M, 星云链类似以太坊，星云生态环境中总货币量保持着稳定增长（目前星云币的增发比例暂定为 4%）。不同于比特币，后者的发行货币总量最终将稳定在 2100 万。货币流通速度 V 可以描述为一段时间内流通的货币总量与当时货币总量的比值。因此式 7可以进一步写为

$$(M + \Delta m) \times \frac{\sum_{(s,t,w,\tau)\in\mathcal{T}} w}{M} = P \times Y \tag{8}$$

其中，Δm 为增发的货币量。

对于物价水平 P，无论是古典货币理论还是新凯恩斯理论，都始终认同货币供给与需求最终决定货币价值的观点。从长期来看，物价总水平会调整到使货币需求等于货币供给的水平。

然而短期内，物价总水平本身并不能使货币供求平衡。在健康的经济系统中，物价水平增长速度往往小于货币总量的增长速度。通过适当增加货币供给（也可描述为降低利率），在物价 P 增长的同时也保证物品与服务需求量 Y 的增加。另一方面，同时需要对物价水平的增速进行一定的控制，以保证用户不会因为过强的增长预期而保持持有数字货币的状态，从而导致流通速度的降低。

对于真实经济产量 Y，经济学家通常用真实 GDP 表示，即“某一既定时期一个国家生产的所有最终物品与服务的市场价值”。我们认为加密数字货币的价值来源于其流通性，即每一笔流通都在一定程度上为整个经济总量做了贡献。也就是说，当一笔实际的交易发生时，既在一定程度上增加了加密数字货币的流通性，也在一定程度上增强了人们对数字货币的认可及信心。因此，我们认为式 8中的 Y 来源于每一笔交易，考虑到经济系统的主体为账户，也可以认为 Y 来源于各个账户上发生的交易，即

$$Y = \sum_{a \in \mathcal{A}} \mathcal{C}(a) \tag{9}$$

其中 $\mathcal{C}(a)$ 表示账户 a 对整个经济产量的贡献，即星云指数。

数字货币的发展依赖于社区的发展，因此，我们认为量化社区中每个账户对于经济总量的贡献，为正确的激励提供了必要的基础。基于此，经济系统可以对账户产生明确的激励（例如星云技术白皮书中的贡献度证明机制（Proof of Devotion, PoD)），亦可能产生不确定的激励（例如，搜索引擎中由于星云指数的影响而对搜索结果的排序产生的影响）。不同于传统货币理论，在加密数字货币中，直接的、原生的激励是增发货币的发放。

4 核心星云指数

核心星云指数用于衡量一段时间内用户对整个经济体的贡献。精确地计算这一指标是十分困难的，因此，我们使用了近似算法。在该近似算法中，我们考虑了两个重要的因素，即账户持有的币龄和账户在交易网络中的位置信息。在稍后的评测中，我们将证明这种近似算法的有效性。

我们使用一段时间内链上的交易记录作为核心星云指数的数据来源。对于一段时间 $[t_0 - T,\ t_0]$ 内的交易记录，可以描述为集合

$$\Theta(t_0) = \{(s, t, w, \tau) \mid t_0 - T \leq \tau \leq t_0 \ \wedge \ w > 0 \wedge s \neq t\} \tag{10}$$

基于 $\Theta(t_0)$，我们可以构造有向加权图，其中节点为账户地址，从节点 s 到节点 d 的有向边是一次交易，边的权值为 w，边的时间为 τ。

对于账户 $a \in \mathcal{A}$，其核心星云指数 $\mathcal{C}(a)$ 的计算基于 $\Theta(t_0)$，即

$$\mathcal{C}(a) = \Omega(\beta(a)) \times \Psi(\gamma(a)) \tag{11}$$

其中 $\beta(a)$ 为一段时间内账户 a 持有资产的中值；$\gamma(a)$ 为账户 a 在一段时间内的出入度指标。

> 相比星云技术白皮书 [3] 中对于星云指数的计算方法，我们做了如下改动：
> 1. 在构造交易图时取消了采取最高 K 个交易额作为权值；
> 2. 取消了 LeaderRank 中通过计算背景节点有向边权重来获取重要性评分的方式。
> 首先，我们在计算出入度指标 β 时采用了去环处理，因此已经可以抵抗环形转账攻击，并同时保留边的强度信息。对于存在同构图的拓扑结构，PageRank 等对称函数（包括 LeaderRank）被证明无法有效抵抗女巫攻击 [20]。因此在本黄皮书中我们没有采用类拓扑排名策略。在 §4.3中我们构造的非对称计算函数 21将可以有效降低伪造低收入节点的收益。

下面，我们将分别考虑 11中的三个问题：资产中值 $\beta(a)$ 的计算、出入度指标 $\gamma(a)$ 的计算，以及函数 Ω 及 Ψ 的选择。

4.1 资产中值 $\beta(a)$

对于时间段 $[t_0 - T,\ t_0]$，区块链系统中存在 n 个区块，记为

$$B_0, B_1, \ldots, B_n$$

其中 B_i 为 B_{i+1} 的父块。对于账户 $a \in \mathcal{A}$，其在每个区块结束后，其相应的账户余额为

$$d_0^a, d_1^a, \ldots, d_n^a$$

上述序列按从小到大排序后可以得到

$$d_{(0)}^a, d_{(1)}^a, \ldots, d_{(n)}^a$$

其中 $d_{(i)}^a < d_{(i+1)}^a, 0 \leq i \leq n-1$，由此，可以得到

$$\beta(a) = \begin{cases} d_{(k)}^a & \text{for} \quad n = 2 \times k, k = 1, 2, 3, \ldots \\ (d_{(k)}^a + d_{(k+1)}^a)/2 & \text{for} \quad n = 2 \times k + 1, k = 1, 2, 3, \ldots \end{cases} \tag{12}$$

资产中值一定程度上代表了“币龄”，即账户中需要至少持有该资产一半以上的时间。

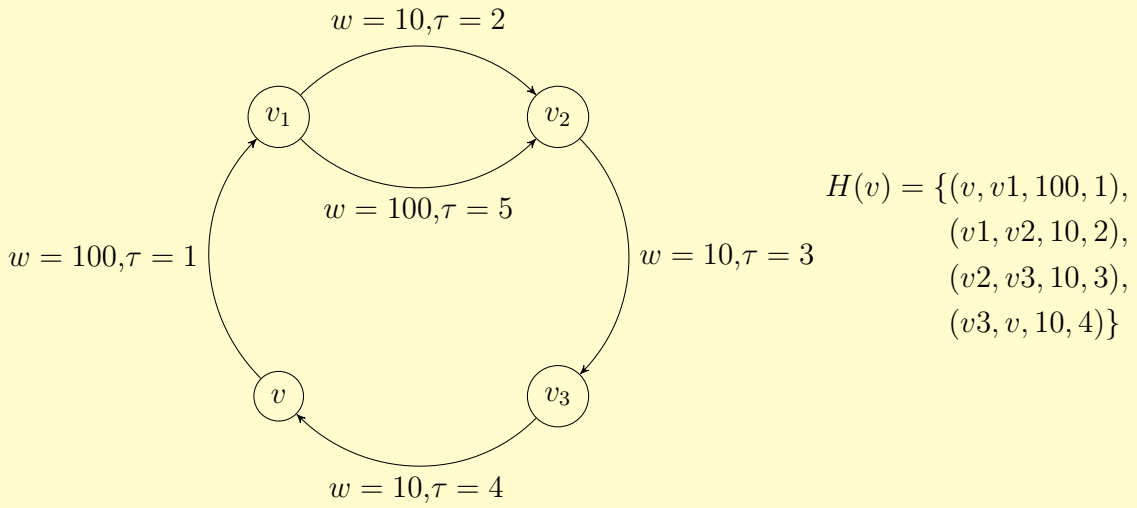

图 1: 交易图中的交易环

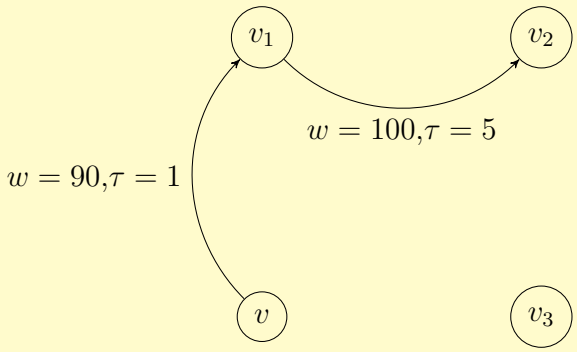

图 2: 图1去掉交易环后的交易图

4.2 出入度指标 $\gamma(a)$

考虑到恶意操控者会利用“环形转账”行为提升自己的出入度指标，因此对于出入度指标的计算首先需要对交易图进行“去交易环”处理。交易环（forwarding loop）是指一组按时间顺序进行的交易行程的环路。一个交易环在节点 v 开始并结束，是交易图中边的集合，记为 $H(v)$，即，

$$H(v) = \{(v, v_1, w_1, \tau_1), (v_1, v_2, w_2, \tau_2), \ldots, (v_i, v_{i+1}, w_i, \tau_i), \ldots, (v_n, v, w_{n+1}, \tau_{n+1})\}$$

其中，$\forall 1 \leq i \leq n : \tau_i \leq \tau_{i+1}$。如图 1所示，包含了一个交易环，注意，其中 $(v_1, v_2, 100, 5)$ 并不包含在交易环中。

在找到交易环后，需要进行去交易环处理。假设系统中存在 n 个交易环，按照

交易环在交易图中出现的顺序记为

$$H^1(v_1), H^2(v_2), \ldots, H^n(v_n)$$

其中，$H^i(v_i)$ 中交易金额最小的交易为 $(s_m^i, t_m^i, w_m^i, \tau_m^i)$，即

$$\forall (s^i, t^i, w^i, \tau^i) \in \mathcal{T} : w^i \geq w_m^i$$

然后，需要依次将 $H^i(v_i)$ 中所有的交易减去相应的最小交易量 w_m^i，如果新的交易量为 0，则移除该交易，即

$$\mathcal{E}((s,t,w,\tau), w_m) = \begin{cases} (s,t,w-w_m,\tau) & \text{if} \quad w \neq w_m \\ \phi & \text{if} \quad w = w_m \end{cases}$$

$$\Theta'(t_0) = \Theta(t_0) - H^i(v) \cup \{\mathcal{E}(t), t \in H^i(v_i)\} \quad i = 1, 2, \ldots, n \tag{13}$$

如图2所示，为图1去掉交易环之后的交易图。

记节点 v 的转入金额为 $p(v)$，则

$$p(v) = \sum_{(s_i, v, w_i, \tau_i) \in \Theta'(t_0)} w_i \tag{14}$$

同理，节点 v 的转出金额有

$$q(v) = \sum_{(v, t_i, w_i, \tau_i) \in \Theta'(t_0)} w_i \tag{15}$$

由此，对于节点 v，其出入度指标 $\gamma(v)$ 为

$$\mathcal{G}(v) = (p(v) + q(v)) \cdot e^{-2\sin^2(\frac{\pi}{4} - \arctan\frac{q(v)}{p(v)})} \tag{16}$$

该函数图形如图3所示。

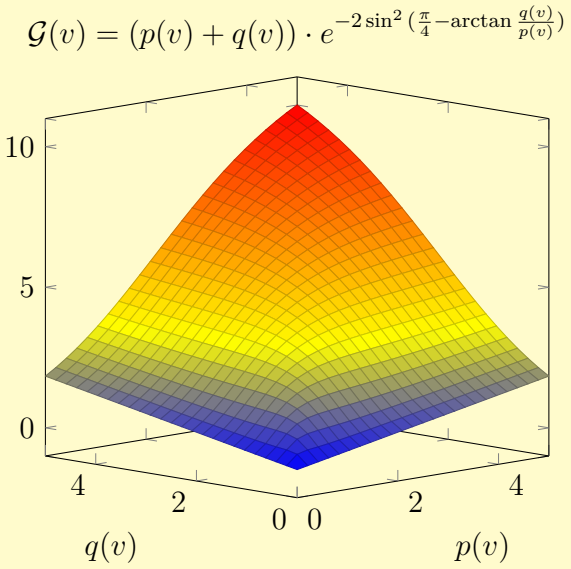

图 3: 出入度计算函数曲线

$$\gamma(v) = (\frac{\theta \cdot \mathcal{G}(v)}{\mathcal{G}(v) + \mu})^{\lambda} \tag{17}$$

其中 θ, μ, λ 为待定的参数。

4.3 Wilbur 函数

考虑到不同的使用场景及不同的性质，星云指数的计算是十分复杂的，然而，我们可以总结出一般意义上的星云指数计算函数的特征。

在此我们将星云指数的计算函数命名为 Wilbur Function[1]，记为 $f(x)$，其中 x 为星云指数需要参考的因素，可以为持有的余额、币龄或账户的出入度。$f(x)$ 需要满足两个性质：

特征 1. 对于任意大于 0 的两个输入变量 x_1，x_2，其计算函数之和小于其和的计算函数。

$$f(x_1 + x_2) > f(x_1) + f(x_2) \quad x_1 > 0, x_2 > 0 \tag{18}$$

[1]女巫攻击（Sybil Attack）命名的来源于上世纪七十年代的美国系列电影《Sybil》，剧中患有人格分裂症的小女孩名叫 Sybil，而治疗她的心理医生则名叫 Dr. Cornelia Wilbur。

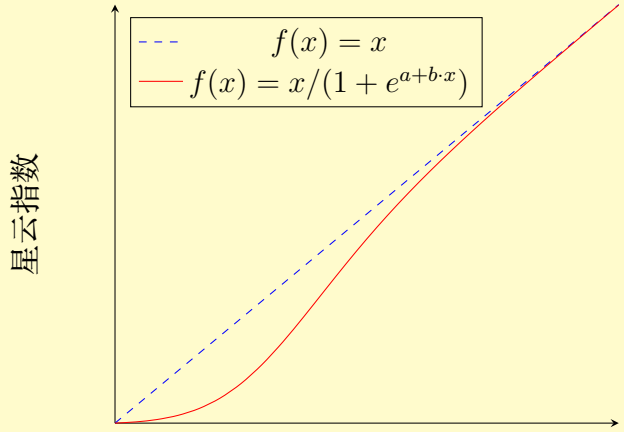

图 4: 星云指数计算函数曲线

特征 2. 当任意大于 0 的两个输入变量 x_1，x_2 趋近于无穷大时，其计算函数之和趋近于其和的计算函数。

$$\lim_{x_1\to\infty, x_2\to\infty} f(x_1 + x_2) = f(x_1) + f(x_2) \quad x_1 > 0, x_2 > 0 \tag{19}$$

上述特性保证了在给定交易行为的情况下，用户通过控制多个账户实现该交易行为的收益小于通过单一账户实现交易的收益，同时当用户资产足够大时，其拆分账户交易的损失可以忽略不计。

满足上述两个性质的函数有很多，在此，我们仅给出一个满足上述性质的函数，该函数的图形如 图4所示。

$$f(x) = x/(1 + e^{a+b\cdot x}) \quad a > 1, b < 0 \tag{20}$$

证明：详细见附录 A

综上，式 11可以进一步写为

$$\mathcal{C}(v) = \frac{\beta(v)}{1 + e^{a+b\cdot\beta(v)}} \cdot \frac{\gamma(v)}{1 + e^{c+d\cdot\gamma(v)}} \tag{21}$$

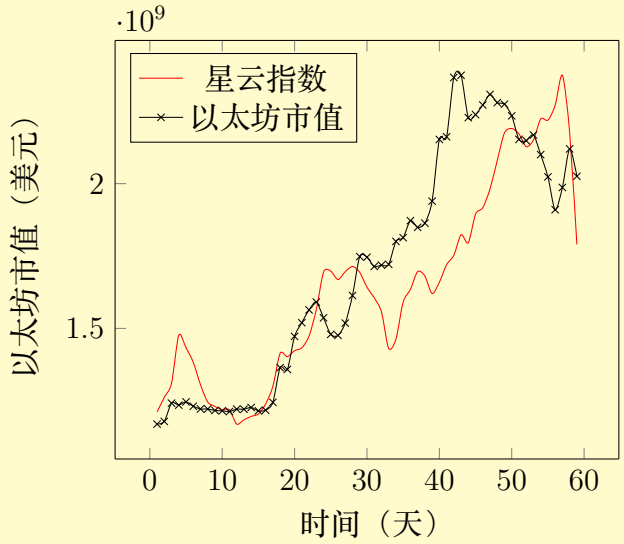

图 5: 以太坊之上的星云指数及以太坊市值

其中，a, b, c, d 为待定的参数。

为了验证该函数的有效性，我们根据以太坊中的数据，计算了以太坊中地址的星云指数，并根据其市值变化计算了两者的相关性。测试数据包括了以太坊从 2017 年 5 月 1 日起至 2017 年 6 月 30 日（3629091 区块至 3955158 区块）的所有交易记录，以及 ETH 日均价格（对美元）和日均总交易量 [24]。

图5表示了以太坊的市值与星云指数随时间变化趋势，如图所示，其中黑色标记实线为通过 ETH 日均交易量以及 ETH 日均价格计算得出的以太坊市值（对美元），而红色实线则是根据公式 21计算所有以太坊用户的星云指数之和。

可以看出，星云指数能够有效反映以太坊市值变化，二者的相关系数（Correlation coefficient）为 0.84427，p 值（p-value）为 $4.48 \times 10^{-17} < 0.001$。说明函数 11能够有效表示用户对整个经济体的贡献，即体现出星云指数的真实性。

5 核心星云指数如何抵抗操纵?

本节分析核心星云指数应对操纵的情况，即核心星云指数的公平性。

“操纵”指的是攻击者作出特定行动以获得最高的利益。攻击者的行动空间是利用自己和盟友所能控制的资产和账户进行转账操作。其中，转账操作的金额不超过攻击者拥有的资产数目；转账的发起方是其能够控制的账户，包括攻击者及其盟友创建的账户，以及愿意提供资产中转服务的服务商账户等等。攻击者能够获得利益一般由其知晓私钥的账户的评分决定。如果有多个这样的账户，一种简单的情况是，攻击者的

利益正比于这些账户的得分总和。当然，先前提到的服务商账户的私钥不受攻击者掌控。

本节的分析基于上述的行动空间和简单情况下对攻击者利益的定义。首先，我们讨论针对单个账户的分数提高上限；然后，我们分析攻击者多个受控账户的分数提升上限；最后，引入合谋行为，讨论多个攻击者联合操纵的结果。

5.1 提升单个账户的分数

为了提升单个账户的分数，根据式21，分数与账户上的资产数目以及出入度指标正相关。账户上的资产数目，即 β 存在上限，即不高于攻击者的资产数目，记为 β_0。而出入度指标 γ 反映了交易量的大小，这意味着攻击者需要尽可能增加单个受控账户的交易量。

增加交易量分为两个方面：增加入度和增加出度。增加出入度需要两个账户参与，除了需要提升分数的账户之外，另一账户有两种情况：受控账户和非受控账户。如果是非受控账户，增加出入度意味着与其他人进行交易，此类情况将在 §5.3讨论，本小节不考虑这种情况。另一种情况是攻击者还可以无条件地向陌生人账户转移资产以增加出度，此类行为的代价较大，亦不予考虑。由此可以判断通常情况下，攻击者的行为主要是增加自己所控制的账户之间的交易。由于攻击者所控制的账户的资产有限，而评分的时间长度也有限，因此该账户的出入度之和具有上限，并由攻击者的资产数目决定。

综上我们考虑与受控账户交易的情况，由于我们在 §4.3提出的计算函数 21，攻击者拆分资产交易的收益会被降低。因此攻击者会采取交易量最大策略，即设法将所拥有的所有资产转入该账户，并随后全部转出，由于存在去环算法，攻击者的资产无法在当前时间段内再次转入。此时的出入度之和为 $\gamma = 2\beta_0$。此时的分数为 $\mathcal{C} = \frac{2\beta_0^2}{(1+e^{a+b\cdot\beta_0})(1+e^{c+2d\cdot\beta_0})}$。

假如攻击者采用了线下交易等方式，将资产完全平移到其它账户并再次转入目标账户，出入度之和的上限是转移次数乘以资产数目，由于评分时间段是有限的，转移次数的上限是常数，因此 γ 的上限是 $2T\cdot\beta_0$，其中 T 为表示评测时间段长度的常数，因此此时分数的上限是 $\mathcal{C} = \frac{2T\cdot\beta_0^2}{(1+e^{a+b\cdot\beta_0})(1+e^{c+c\cdot d\cdot\beta_0})}$。

5.2 提升多个账户的分数（女巫攻击）

女巫攻击（Sybil Attack）是指攻击者通过创建大量假名标识来破坏对等网络的信誉系统，使用其获得虚假的高重要性评分 [25]。

对等网络上的实体是能够访问本地资源的一块软件。实体通过呈现身份在对等网络上通告自身。多于一个标识可以对应于单个实体。换句话说，身份到实体的映射是多对一的。对等网络中的实体为了冗余，资源共享，可靠性和完整性而使用多个标识。在对等网络中，身份用作抽象，使得远程实体可以知道身份而不必知道身份与本地实体的对应关系。默认情况下，通常假定每个不同的标识对应于不同的本地实体。实际上，许多身份可以对应于相同的本地实体。对手可以向对等网络呈现多个身份，以便出现并充当多个不同的节点。因此，对手可能能够获得对网络的不成比例的控制水平，例如通过影响投票结果 [26]。

这里假设攻击者的收益是所有受控账户的分数之和。考虑上一小节的针对单个账户的分数提升策略，这里将此策略重复施用于多个账户：从一个账户出发，攻击者将自己资产的一部分转出当前账户并且转入下一个账户，最终形成链式的流。这样的情况下，由于核心星云指数对资产的计量要求不小于最终数目的资产在账户上停留不少于一半的时间，攻击者无法令所有账户的 β 变成其拥有的全部资产数目。另一种策略是平均地将自己的资产分到所有账户上，假设流长度为 N，即有 N 个账户，则每个账户上的 $\beta = \frac{\beta_0}{N}$。出入度的分析与 §5.1相同，$\gamma$ 上限为 $K \cdot \beta$，其中 $K = 2 \cdot N$ 为常量。因此攻击者所拥有的所有账户分数之和的上限是:

$$\mathcal{C} = N \cdot \frac{K\frac{\beta_0^2}{N}}{(1+e^{a+b\cdot\frac{\beta_0}{N}})(1+e^{c+K\cdot d\cdot\beta_0})} = \frac{K\beta_0^2}{(1+e^{a+b\cdot\frac{\beta_0}{N}})(1+e^{c+K\cdot d\cdot\beta_0})} \tag{22}$$

5.3 联合操纵

联合操纵的结果和一个攻击者拥有了原先两个攻击者的资产数目的情况相同，因此可以通过分析单个攻击者资产增加的后果来分析合谋的结果。

6 核心星云指数的实现

核心星云指数的完整实现不在本文的讨论范围，此处我们仅讨论对于核心星云指数的实现方面的一些关键问题。

6.1 是否上链?

核心星云指数给出了每个账户对经济总量的贡献度，正常来说，每个节点都可以根据历史区块信息计算指定账户对经济总量的贡献度。然而，我们是否需要周期性的将一定间隔时间内的星云指数上链呢?

我们认为并不需要将星云指数上链，这出于以下两个方面的考虑：

- 链上不适合存储如此大量的数据，即使对于 IPFS，Genaro 之类 [27] [28] 的以数据存储为目标场景的公链，也不适合周期性的存储所有账户的核心星云指数；
- 对于核心星云指数的计算会影响出块速度，核心星云指数的计算复杂度较高，如果将计算结果上链，将很大程度上影响出块及验证速度，导致整个系统的 TPS 降低。

综上，我们认为每个节点可以根据需要自行计算核心星云指数。

然而，在节点自行计算核心星云指数的情况下，一个重要的问题是如何保证核心星云指数计算的可信性。例如，节点任意篡改核心星云指数的计算结果，从而根据错误的核心星云指数给出指定的激励。对于关键性的应用，我们认为需要在各个节点对涉及核心星云指数的计算结果进行校验，以保证结果的公平性。而对于非关键性的应用，对核心星云指数的使用依赖于应用本身，是否对核心星云指数进行校验取决于应用。

在节点自行计算核心星云指数的情况下，另一个重要的问题是节点考虑到能耗问题，而拒绝计算核心星云指数。对此，我们认为可以引入可信的核心星云指数服务，避免各个节点上的重复计算，对于该服务的使用，可以免费使用，也可使用次数计费。具体的实现及服务细节不在本文讨论范围内。

6.2 核心星云指数的更新

我们深知，核心星云指数是和加密数字货币的生态紧密相关的，随着生态的不断变化，核心星云指数的计算也需要不断更新，尤其是其中的各个参数。如何快速地更新核心星云指数的计算非常关键。对此，我们将通过星云原力来保障核心星云指数计算的更新迭代。

我们会更新区块结构，新的区块结构中将包含核心星云指数的算法及参数（以 LLVM IR 形式），星云虚拟机（NVM）作为算法的执行引擎，从区块中获得核心星云指数的算法及参数，并执行算法，在节点内获得账户的核心星云指数。

在算法或参数需要更新时，我们将和社区一起协作，让新的区块中包含入最新的算法及参数，从而保证整个更新过程的及时性及平滑性，亦避免了可能到来的分叉。

7 扩展星云指数

核心星云指数用来衡量一个账户地址对于整个加密数字货币经济总量的贡献。对于星云链规划中的贡献度证明（PoD）及开发者激励（DIP）是十分重要、且符合其应用场景的。然而，我们注意到，很多场景需要不一样的价值衡量尺度，为此，除了核心星云指数，我们还设计了扩展星云指数。扩展星云指数基于核心星云指数，以保证在不同应用场景下，能够持续激励星云生态的发展。

7.1 针对智能合约的扩展星云指数

生态中智能合约排序是十分重要的，一方面能够帮助用户发现更高质量的 DApp，另一方面，能够激励开发出高质量应用的开发者，使整个生态健康稳定的发展。

对智能合约的排序，基于两个事实：账户地址对智能合约的调用，和智能合约之间的调用。我们首先将账户地址对智能合约的调用看作账户地址向智能合约分摊自己对经济总量的贡献，从而使得每个智能合约有了初始的分值。然后将智能合约之间的调用看作有向无环图，使用 Page Rank 对每个智能合约计算最终的星云指数。

7.2 多维的扩展星云指数

我们同时注意到一些应用需要多个维度的数据，以便对链上数据的相关性进行计算。例如基于区块链的广告系统，需要在多个维度对需要投放的广告及用户进行相关性计算。在这种场景下，扩展星云指数是多维的，即表示为一个向量，核心星云指数是其中的一个维度。

扩展星云指数是多维度的，除了核心星云指数外，其它维度均依赖于具体的应用场景。这些维度的使用同样依赖于具体的应用场景。但并不冲突的是，其计算方法亦可参考本文中核心星云指数的计算方法。

我们通过智能合约的扩展星云指数这样一个真实的应用场景，描述了扩展星云指数的一种实现方式，给出了相应的价值衡量尺度，我们还给出了多维的扩展星云指数，为更多应用场景下的价值衡量尺度提供了可能性。

8 未来工作

星云指数的目标是为区块链提供必要的价值衡量尺度，而不仅仅是从账户地址等实体对经济总量贡献的角度对链上数据进行衡量，因此，未来仍有许多工作。在此，我们简要地给出部分我们将持续关注的工作。

- 跨链的星云指数 可以预见到，跨链的数据转移将成为不可避免的应用需求。例如跨链的数据交互、数字资产转移等，这需要对不同链上的价值作出衡量。开发者将 DApp 从一个链迁移到另一个链，该 DApp 在一个链上星云指数如何在另一个链上反映出来等，这些都需要星云指数能够对不同链的价值提供统一的衡量尺度；
- 更多经济总量的贡献指标 星云指数的基础是经济总量的贡献，然而，区块链的发展离不开社区，社区的增长对经济总量的贡献是不可忽视的，如何衡量个体或组织对社区增长的贡献，并将其反映到星云指数中，具有不可忽视的现实意义。

参考文献

[1] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: characterizing payments among men with no names," in Proceedings of the 2013 conference on Internet measurement conference, pp. 127–140, ACM, 2013.

[2] "Http cookie." https://en.wikipedia.org/wiki/HTTP_cookie.

[3] "Nabulas Technical White Paper." https://nebulas.io/docs/NebulasTechnicalWhitepaper.pdf. Accessed: 2018-04-01.

[4] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[5] "Namecoin." https://namecoin.org.

[6] "Openassets protocol." http://github.com/OpenAssets/open-assets-protocol.

[7] V. Buterin et al., "Ethereum white paper," 2013.

[8] "Forget fintech –welcome to the valueweb." http://thefinanser.com/2015/02/forget-fintech-welcome-to-the-valueweb.html/.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," tech. rep., Stanford InfoLab, 1999.

[10] M. Fleder, M. S. Kester, and S. Pillai, "Bitcoin transaction graph analysis," arXiv preprint arXiv:1502.01657, 2015.

[11] Q. Li, T. Zhou, L. Lü, and D. Chen, "Identifying influential spreaders by weighted LeaderRank," Physica A: Statistical Mechanics and its Applications, vol. 404, pp. 47–55, 2014.

[12] A. Cheng and E. Friedman, "Manipulability of pagerank under sybil strategies," 2006.

[13] "NEM Technical Reference." http://nem.io/NEM_techRef.pdf. Accessed: 2017-08-01.

[14] A. N. Nikolakopoulos and J. D. Garofalakis, "NCDawareRank," Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13, no. February 2013, p. 143, 2013.

[15] X. Xu, N. Yuruk, Z. Feng, and T. A. Schweiger, "Scan: a structural clustering algorithm for networks," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 824–833, ACM, 2007.

[16] H. Shiokawa, Y. Fujiwara, and M. Onizuka, "Scan++: efficient algorithm for finding clusters, hubs and outliers on large-scale graphs," Proceedings of the VLDB Endowment, vol. 8, no. 11, pp. 1178–1189, 2015.

[17] L. Chang, W. Li, L. Qin, W. Zhang, and S. Yang, "pscan: Fast and exact structural graph clustering," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 2, pp. 387–401, 2017.

[18] J. Hopcroft and D. Sheldon, "Manipulation-resistant reputations using hitting time," in International Workshop on Algorithms and Models for the Web-Graph, pp. 68–81, Springer, 2007.

[19] J. Zhang, R. Zhang, J. Sun, Y. Zhang, and C. Zhang, "Truetop: A sybil-resilient system for user influence measurement on twitter," IEEE/ACM Transactions on Networking, vol. 24, no. 5, pp. 2834–2846, 2016.

[20] A. Cheng and E. Friedman, "Sybilproof reputation mechanisms," in Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, pp. 128–132, ACM, 2005.

[21] M. Swan, Blockchain: Blueprint for a new economy. O'Reilly Media, Inc., 2015.

[22] R. S. Kroszner, "Liquidity and monetary policy," 2007.

[23] R. Selden, "Monetary velocity in the united states," 1956.

[24] "CoinMarketCap." `https://coinmarketcap.com/`.

[25] D. Quercia and S. Hailes, "Sybil attacks against mobile users: friends and foes to the rescue," in INFOCOM, 2010 Proceedings IEEE, pp. 1–5, IEEE, 2010.

[26] Wikipedia contributors, "Sybil attack — Wikipedia, the free encyclopedia," 2018. [Online; accessed 25-June-2018].

[27] "Ipfs." `https://ipfs.io/`.

[28] "Genaro." `https://genaro.network/en/`.

附录 A　证明

A.1　特征1证明

证明. 对于任意 $x_1 > 0$，$x_2 > 0$，有

$$\begin{aligned} f(x_1 + x_2) &= \frac{x_1 + x_2}{1 + e^{a+b\cdot(x_1+x_2)}} \\ &= \frac{x_1}{1 + e^{a+b\cdot(x_1+x_2)}} + \frac{x_2}{1 + e^{a+b\cdot(x_1+x_2)}} \\ &= \frac{x_1}{1 + e^{b\cdot x_2} \cdot e^{a+b\cdot x_1}} + \frac{x_2}{1 + e^{b\cdot x_1} \cdot e^{a+b\cdot x_2}} \end{aligned}$$

在公式21中，有 $b < 0$，因此 $0 < e^{b\cdot x_1} < 1$，$0 < e^{b\cdot x_2} < 1$，进一步地，

$$\frac{x_1}{1 + e^{b\cdot x_2} \cdot e^{a+b\cdot x_1}} > \frac{x_1}{1 + e^{a+b\cdot x_1}} = f(x_1)$$

$$\frac{x_2}{1 + e^{b\cdot x_1} \cdot e^{a+b\cdot x_2}} > \frac{x_2}{1 + e^{a+b\cdot x_2}} = f(x_2)$$

即

$$f(x_1 + x_2) > f(x_1) + f(x_2)$$

□

A.2　特征2证明

证明. 对于任意 $x_1 > 0$，$x_2 > 0$，有

$$\begin{aligned} f(x_1 + x_2) - f(x_1) - f(x_2) &= \frac{x_1 + x_2}{1 + e^{a+b\cdot(x_1+x_2)}} - \frac{x_1}{1 + e^{a+b\cdot x_1}} - \frac{x_2}{1 + e^{a+b\cdot x_2}} \\ &= \left(\frac{x_1}{1 + e^{b\cdot x_2} \cdot e^{a+b\cdot x_1}} - \frac{x_1}{1 + e^{a+b\cdot x_1}}\right) \\ &\quad + \left(\frac{x_2}{1 + e^{b\cdot x_1} \cdot e^{a+b\cdot x_2}} - \frac{x_2}{1 + e^{a+b\cdot x_2}}\right) \end{aligned} \tag{23}$$

这里用函数 $g(x_1, x_2)$ 表示左边部分，$h(x_1, x_2)$ 表示右边部分，即：

$$g(x_1, x_2) = \frac{x_1}{1 + e^{b \cdot x_2} \cdot e^{a+b \cdot x_1}} - \frac{x_1}{1 + e^{a+b \cdot x_1}} \tag{24}$$

$$h(x_1, x_2) = \frac{x_2}{1 + e^{b \cdot x_1} \cdot e^{a+b \cdot x_2}} - \frac{x_2}{1 + e^{a+b \cdot x_2}} \tag{25}$$

因此 (23)对于 x_1 和 x_2 的极限可以表示为：

$$\lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} [f(x_1 + x_2) - f(x_1) - f(x_2)] = \lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} g(x_1, x_2) + \lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} h(x_1, x_2)$$

其中

$$\begin{aligned} g(x_1, x_2) &= \frac{x_1}{1 + e^{b \cdot x_2} \cdot e^{a+b \cdot x_1}} - \frac{x_1}{1 + e^{a+b \cdot x_1}} \\ &= \frac{x_1 \cdot e^{a+b \cdot x_1} \cdot (1 - e^{b \cdot x_2})}{(1 + e^{b \cdot x_2} \cdot e^{a+b \cdot x_1}) \cdot (1 + e^{a+b \cdot x_1})} \\ &< \frac{x_1 \cdot e^{a+b \cdot x_1} \cdot (1 + e^{a+b \cdot x_1})}{(1 + e^{b \cdot x_2} \cdot e^{a+b \cdot x_1}) \cdot (1 + e^{a+b \cdot x_1})} = \frac{x_1 \cdot e^{a+b \cdot x_1}}{1 + e^{b \cdot x_2} \cdot e^{a+b \cdot x_1}} \\ &< \frac{x_1 \cdot e^{a+b \cdot x_1}}{1 + e^{a+b \cdot x_1}} = \frac{x_1}{1 + \frac{1}{e^{a+b \cdot x_1}}} \end{aligned}$$

对 $\frac{x}{1+\frac{1}{e^{a+b \cdot x}}}$ 求极限，根据洛必达法则，

$$\begin{aligned} \lim_{x \to \infty} \frac{x}{1 + \frac{1}{e^{a+b \cdot x}}} &= \lim_{x \to \infty} \frac{1}{(e^{-a-b \cdot x})'} \\ &= \lim_{x \to \infty} \frac{1}{-b \cdot e^{-a-b \cdot x}} \end{aligned}$$

在公式21中，有 $b < 0$，因此 $\lim_{x \to \infty} -b \cdot e^{-a-b \cdot x} = \infty$，因此，

$$\lim_{x \to \infty} \frac{x}{1 + \frac{1}{e^{a+b \cdot x}}} = 0$$

根据 A.1，有 $g(x_1, x_2) > 0$，因此根据夹逼定理有：

$$\lim_{\substack{x_1\to\infty\\x_2\to\infty}} g(x_1,x_2)=0$$

同理，可以求得：

$$\lim_{\substack{x_1\to\infty\\x_2\to\infty}} h(x_1,x_2)=0$$

因此，

$$\lim_{\substack{x_1\to\infty\\x_2\to\infty}} [f(x_1+x_2)-f(x_1)-f(x_2)]=0$$

□

附录 B　修订记录

- 1.0 正式版本发布。
- 1.0.1 修改 §4.3中特征1和特征2以及相关证明中部分数学符号描述，避免引起歧义。