去中心化声誉网络：一种自然社会组织的框架

Pavel Kudrna
Prague, Czechia
Draft v1 — March 2026

Abstract

正义感——即坚信错误会得到纠正、功绩会得到回报的信念——是社会最强大的凝聚力量。当中心化的治理机构因为强制执行一项权利的成本超过该权利本身的价值而无法提供这种正义感时，社会凝聚力便会瓦解。当前的制度结构无法解决相当一类纠纷，其失败方式与中央计划在资源配置上的失败在系统上如出一辙：都源于信息不对称、反馈回路缺失，以及决策者与其决策后果之间的脱节。本文提出一种声誉社交网络（Reputation Social Network，RSN）：一个建立在去中心化标识符（Decentralized Identifiers，DID）之上的去中心化、不可腐蚀、抗审查的通信层，使匿名化的参与者能够发布、验证和消费有关现实世界行为的声誉信息。其核心机制建立在三条设计原则之上：（1）一种非对称的成本结构，其中读取声誉数据成本低廉，而写入则需要可验证的付出；（2）一种基于一致性哈希的非确定性验证者选择协议，通过不断升级的迭代成本对激进主张进行定价；以及（3）一种由市场驱动的声誉权威模型，其中私营验证者以其累积的声誉为其所背书主张的准确性作担保。由此产生的架构无需中央协调即可催生出涌现式的精英体制，并使得从现有的国家管理体系进行渐进且可逆的迁移成为可能。

1 引言

1.1 中心化信任的问题

现代治理建立在一个根本假设之上：中心化机构能够大规模地在陌生人之间充当信任的中介。法院裁决纠纷。登记处认证所有权。监管机构强制执行标准。这些机构是在信息稀缺、通信缓慢、将权力委托给中央机构是唯一可行的协调机制的时代设计出来的。然而，中心化治理会因与中央计划相同的结构性原因而失败：信息不对称、反馈回路缺失，以及决策者与其决策后果之间的脱节。

举例来说，当制度性调解的成本超过纠纷本身的价值时，这一假设便告失败。设想一名租户发现其租住的公寓缺少法律要求的电气安全证书——这是一种对生命构成直接危险的状况。该租户依法退出合同的权利是明确无疑的。然而，通过法院系统强制执行这一权利的成本，却超过了押金加应赔损失的金额，即便计入可能的法定赔偿也是如此 [1]。理性的应对是吸收损失并退出。

这并非病态的边缘案例。它是相当一类纠纷的默认体验——在这类纠纷中，伤害是真实的，但金钱利益却低于制度性强制执行在经济上变得合理的门槛。其结果是一个在结构上偏袒不良行为者的系统：那些以低于此门槛的规模伤害他人的人可以肆无忌惮，因为寻求正义的成本落在了受害方身上。

上述例子取自作者所处的法律环境——捷克大陆法系。在其他法律传统中——判例式的普通法、习惯法、混合法系——正义会以不同的方式、不同的程度失效，取决于该管辖区在文化和程序上的具体特征。读者很可能会从自身所处的语境中辨识出类似的例子。具有结构普遍性的是这一模式：价值低于经济上合理的强制执行门槛的纠纷，最终以损失由受害方承担而告终。RSN 并不承诺完美的正义——它只是削弱不良行为者在制度性强制执行不划算时所享有的结构性优势。

1.2 为何现有方案力有未逮

去中心化身份（DID）框架 [2] 提供了用于自主主权身份管理的密码学机制，但主要聚焦于身份验证，而没有涉及行为声誉这一更广泛的问题。

灵魂绑定代币（Soulbound Tokens，SBT） [3] 抓住了声誉不可转让这一洞见，但依赖于存在可扩展性约束的区块链基础设施，且没有解决支配信息录入的经济激励问题。诸如功绩代币（例如 Liberland）之类的相关概念秉持相似的理念，但仍受限于特定的管辖框架。

去中心化自治组织（DAO） [4] 通过智能合约实现治理，但复制了金权政治的动态，其中影响力与资本成正比。二次方投票 [5] 部分缓解了这一问题，但限制了实际采用。DAO 还面临一个根本性的预言机问题：智能合约无法在没有可信外部来源的情况下可靠地验证现实世界的状态，而这一问题在区块链架构内没有通用的解决方案。

现有的系统中，没有任何一个能够将去中心化、抗审查、匿名化、可验证的准入成本以及涌现式共识融为一体。

1.3 贡献

本文作出以下贡献：

1. 定义了**声誉社交网络（RSN）**，一种扩展 DID 框架以支持双边声誉交换的去中心化协议。
2. 一种基于一致性哈希 [6] 的**非确定性验证者选择协议**。
3. **昂贵激进主义原则**，通过三条相互叠加的成本渠道，根据主张与网络共识的距离对其定价。
4. 一种具有非对称激励的**可信权威模型**，其中虚假背书的代价灾难性地远高于正当收费。
5. 一条穿越并行财政基础设施的**渐进迁移路径**。

2 设计原则

RSN 架构受五条不可妥协的设计原则约束。

延续与演化。系统在长度不定的过渡期内与现有机构共存。过渡在每一个阶段都必须是可逆的。

自愿与责任。参与是自愿的，但自愿与责任相绑定：一个承担起某项制度职能控制权的参与者，同时也承担起随之而来的义务。

多样性与文化中立。共识从双边互动中涌现，而非来自系统设计者强加的预设规则。

韧性与抗审查。审查网络的成本必须超过容忍它的成本。只有全面的极权主义——以毁灭性的政治和经济代价——才能压制该系统。

共识与执行。系统将人类对琐碎报复、怨恨和复仇快感的倾向纳入其中，作为承重的核心特性。不当行为不被禁止；它被定价。

3 系统概览

3.1 声誉社交网络

RSN 是一个去中心化的点对点通信层，参与者在其中交换关于现实世界行为的可验证信息。其决定性属性包括：去中心化且不可审查的基础设施；通过 DID 实现的匿名化参与；非对称的成本结构（读取廉价，写入昂贵）；密码学可验证性；以及**标准支持**——为通过网络传播的信息、为权威提供的服务（主张组合、背书、见证服务）、以及为策略格式（包括评估交易对手声誉、评估策略不匹配风险的规则，即声明行为与实际行为之间的差异）提供机器可读的格式。

3.2 架构组件

RSN 由四个主要组件构成（图 1）：

1. **DID 层。**带有声明规则、声誉元数据和网络路由的参与者身份。
2. **主张协议。**用于发布关于现实世界事件断言的结构化格式。
3. **验证网络。**使用一致性哈希指派验证者的去中心化协议。
4. **声誉聚合层。**生成人类可读声誉摘要的服务。

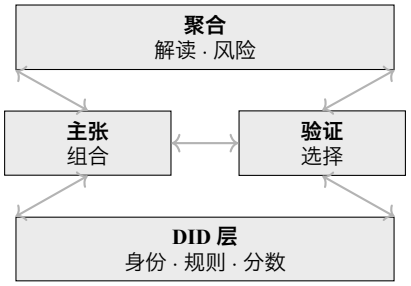

Figure 1: RSN 四层架构。

3.3 参与者角色

一次验证交易最多涉及六个不同的 DID 角色（图 2）：**发起者**（DID_I，创建并提交主张）、**主体**（DID_S，主张所针对的 DID——可能与发起者重合）、**权威**（DID_A，收费为主张背书；也可能提供见证服务——可选）、**见证者**（$DID_{W_{1..n}}$，通过挑战码隐身监督验证者诚信——可选）、**验证者**（DID_V，通过算法选出以发布主张），以及 **RSN**（已验证主张所发布的网络）。任何角色都可以委托给另一个 DID（第 7.4 节）。权威通常将背书与见证服务捆绑，但这两项功能在逻辑上相互独立。

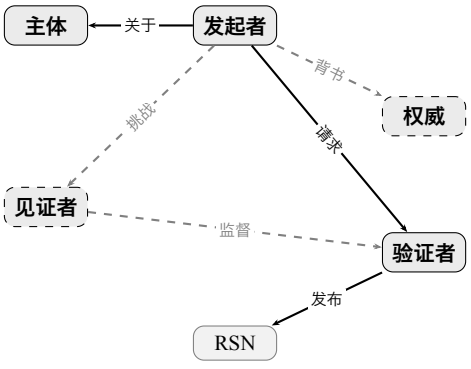

Figure 2: 验证交易中的 DID 角色。虚线 = 可选（权威、见证者）。

3.4 不可腐蚀性：四种力量

RSN 的不可腐蚀性并非源于单一机制，而是源于四种并行的力量。任何一种都不足以独立成事；只

有它们的组合才使腐蚀的企图在经济上变得非理性。

不可预测的目标。每条主张的验证者都通过一致性哈希（第 5.3 节）非确定性地选出。攻击者无法预先锁定某个特定验证者进行行贿，因为验证者的身份是主张内容和先前迭代的函数，而非发起者的选择。

声誉杠杆——上升慢，下降快。声誉只能通过持续一致的行为逐步获得。然而，它却可能因一次欺诈性背书而灾难性地丧失（第 6.2 节）。这种不对称意味着多年累积的声誉可能被一次不诚实的背书摧毁；最优策略始终是拒绝可疑主张。

公开承诺。每个 DID 持有者都公开声明其策略——一套他们承诺遵守的机器可读规则。声明与实际行为之间的偏离是可被检测的，并作为一种比底层违规更严重的声誉过失被定价（第 7.3 节）。因此，伪善比直接的不诚实代价更高。

网状攻击成本不对称。审查或破坏网络稳定的成本随网络的规模和密度非线性增长，而容忍它的成本保持恒定。要使审查有利可图，中心化行为者就必须在整个网络范围内实施——这需要与任何可以想象的收益都不成比例的措施（第 10 节）。

4 去中心化身份模型

4.1 具有特殊属性的 DID

RSN 扩展了 W3C DID Core 规范 [2]，增加了：**声明规则**（机器可读的行为承诺）、**声誉元数据**（聚合的行为分数），以及**网络路由**（可变的洋葱网关，与稳定的环位置相分离）。

4.2 主张生命周期

一条主张经历六个阶段（图 3）：组合、可选的权威背书、通过一致性哈希进行验证者选择、验证决定（接受，或拒绝并迭代）、发布，以及所有各方的声誉更新。

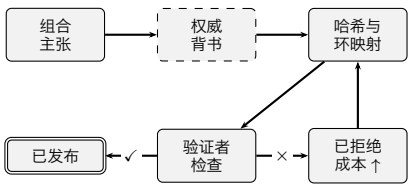

Figure 3: 带迭代回路的主张生命周期。

4.3 与 W3C DID Core 的关系

RSN 的身份模型是 W3C DID Core 规范 [2] 的一个真超集。所有 RSN DID 都是有效的 W3C DID。RSN 特有的扩展被编码为专用 DID 方法规范中定义的 DID 文档属性，与诸如 ION [7] 和 Ceramic [8] 之类的现有基础设施兼容。

5 信息验证与传播

5.1 信息录入的成本

RSN 的核心经济原则是读取与写入之间的不对称。对于处于 DID 网络中、并拥有与其声誉相称访问权的参与者而言，读取声誉数据的边际成本趋近于零——新加入者必须逐步赢得更广泛的访问权(参见反吸血)。写入——理解为声誉在网络中的持久物化，而非一次性的技术动作——需要跨三个维度的可验证累积投入：**时间**（用于一致行为、履行承诺和培养关系的多年人生）、**精力**（投入于诚实交往、维护合作关系和长期可信度的努力），以及**金钱**（对自身名声的投资——职业发展、工作质量、对所造成伤害的补偿）。声誉无法即刻购得——它是终身累积的产物，系统只是使其可见，并可在不同情境间转移。协议的一次性技术成本（密码学签名、验证者费用）与这一底层投入相比微不足道。

5.2 社交图谱与联系人深度

RSN 本质上是一个社交网络：每个 DID 添加允许建立连接的联系人（其他 DID）。这些联系人又有各自的联系人，如此递推。算法验证者搜索在可配置的关联联系人深度 h 范围内运作（例如 $h = 3$: 一个人的直接联系人、他们的联系人，以及联系人的联系人）。这种架构不需要单一的全局区块链——它天然有利于社区/集群的涌现，并与其他社区相互交叠。每个 DID 参与者都必须发布一份**策略**——一套支配如何评估传入请求的机器可读规则。策略违规会作为负面痕迹被记录在网络中。

5.3 算法验证者选择

验证协议采用一致性哈希 [6]（图 4）。验证是一项收费服务：验证者收费运作，形成一个由竞争推动定价、速度和可靠性的市场。

步骤 1。将主张文档与上一次迭代的哈希结果（首次迭代时为 $\varnothing$）拼接后进行哈希：

$$\text{pos} = f_h(\text{claim} \| \text{prev\_hash}) \tag{1}$$

该算法必须包含所有先前请求和响应的完整路径——而不仅仅是上一次迭代的哈希。每个验证者的完整响应（接受、拒绝、报价、理由）都被追加到不断增长的文档中，并在每一步通过密码学签名可验证。

步骤 2。哈希被映射到一致性哈希环上的 N 个位置，均匀分布（例如，对于 $N = 4$: $+0°, +90°, +180°, +270°$）。从每个位置出发，顺时针搜索选出最近的 DID 作为验证者候选（图 5）。N 是一个可变参数，可能取决于当前活跃 DID 的百分比、一天中的时段，或网络历史响应性。

步骤 3。验证者接受（发布，以声誉作担保）或拒绝（带着拒绝哈希返回步骤 1）。当某个节点声明在线却不响应时，下一次请求必须包含来自全部 N 个先前验证者候选的所有响应。

反吸血。目标 DID 可为来自声誉甚微的新 DID 的查询设置费用或访问限制。这一分级机制（并非二元）迫使新加入者在自由消费他人数据之前，通过真实活动建立声誉。

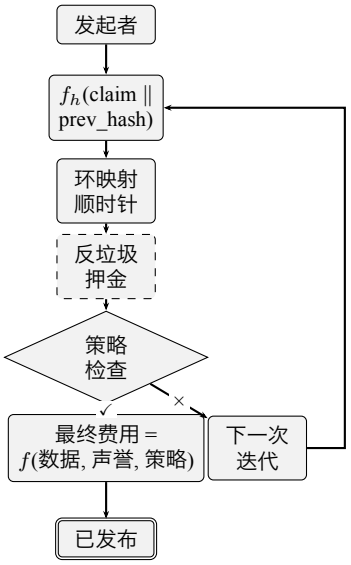

Figure 4: 验证者选择协议。反垃圾押金为可选，且可以退还。最终费用仅在接受时支付。

每次拒绝都会将验证者的完整响应（包括理由和条件）追加到主张文档，扩展其内容。从扩展后的文档中，非确定性地计算出一个新哈希，映射到不同的环位置并选出不同的验证者。记录完整路径是强制性的——发起者无法预测或影响下一个验证者。如果某个验证者尽管声明了兼容的策略却忽略请求，见证者（若在场）会记录此事，发起者可在最终发布时附上见证证据。

5.4 见证者协议

见证者角色通过一种密码学挑战码协议，为验证者诚信提供隐身的问责：

步骤 W1。请求者签署验证请求并将其发送给 N_w 个见证者——来自请求者社交网络的联系人，或收费运作的专业见证服务提供者。

步骤 W2。每个见证者 w_i 签署整个已签署的请求，保留原始签名 σ_i，并计算挑战码 $c_i = h(\sigma_i)$。挑战码被追加到请求中。

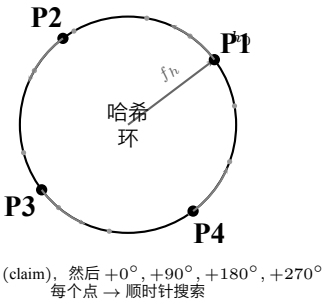

Figure 5: 多点选择（N=4）。哈希 h_0 设定偏移量；从 h_0 均匀分布 4 个点。

步骤 W3。现在携带 N_w 个挑战码的请求被发送给验证者。验证者观察到这些码，但无法确定见证者是谁，也无法确定这些码是否对应真实的签名。

步骤 W4（诚实的验证者）。如果验证者按其声明的策略作出响应，挑战码保持不透明。见证者永不暴露。不发布额外数据。

步骤 W5（策略违规）。如果验证者违反其策略（忽略请求、响应前后不一），请求者持有见证者的原始签名 σ_i。这些可作为配套数据发布：任何人都能验证 $c_i = h(\sigma_i)$，证明该见证者当时在场。请求者可以独立发布——验证者已在其响应中签署了挑战码。

虚张声势属性。请求者可以用随机值替换部分或全部挑战码。验证者无法区分真实的码（由高声誉权威支持）与噪声。这种不确定性正是执行机制：每一次请求都带有某位受人尊敬的权威可能在隐身观察的风险。维持这种压力的成本几乎为零（生成随机数），而对验证者而言不诚实的潜在代价却是灾难性的。即使在没有实际见证者的情况下，诚实行为也受到激励。

作为见证者的权威。为某条主张背书的权威可以捆绑见证服务："我为你的主张背书，并在验证期间充当隐身见证者。"这是一种自然的商业模式——权威本就掌握上下文。然而，这两种角色在逻辑上相互独立。

5.5 昂贵激进主义原则

三条成本渠道同时叠加（图 6）：

渠道 1：迭代成本。每一次拒绝都迫使一次消耗时间和资源的新迭代。线性增长。

渠道 2：文档膨胀。每一次迭代都会将验证者的完整响应加入主张文档。增长是线性的，但带有一个基础偏移量：初始载荷（主张内容、权威背书、密码学签名）本已具有不可忽略的大小 B_0。k 次迭代后的总大小：$S(k) = B_0 + k \cdot \Delta$，其中 Δ 是平均响应大小。

渠道 3：验证者风险溢价。风险是主观的。验证者评估请求方所声明的策略是否与验证者自身的商

业、社会或政治利益相冲突。溢价可能随着与验证者“理想”的感知距离而指数级攀升：$P(d) = \alpha \cdot e^{\beta d}$，其中 d 是验证者的主观距离度量。超过某个个人的禁区边界，验证者可能完全拒绝。

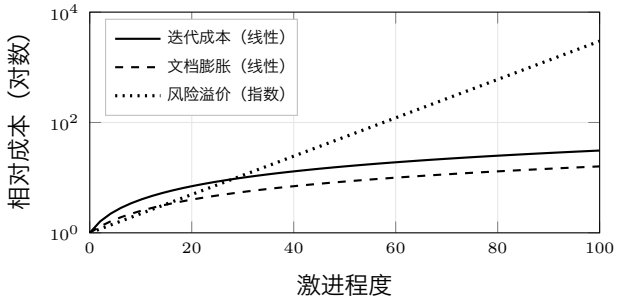

Figure 6: 三条渠道上的成本升级。在高激进程度处，风险溢价占主导。

至关重要的是，这不是审查。没有任何主张被禁止。系统为风险定价（表 1）。

Table 1: 各类主张的成本比较。

类型	迭代	文档	费用	总计
可信	$k_{\min}$	B_0	$P_{\min}$	低
有争议	k_{med}	$B_0 + k\Delta$	$\alpha e^{\beta d}$	中等
激进	$k_{\max}$	$\gg B_0$	$\gg P_{\min}$	极高
欺诈	$\to \infty$	—	—	无法发布

6 声誉与风险评估

6.1 声誉累积模型

声誉呈现三种属性：**缓慢累积**（通过一致行为逐步增长）、**快速摧毁**（单一欺诈可灾难性地降低分数），以及**个体化评估**（每个消费方都可以应用其自己的聚合函数）。

6.2 可信权威

可信权威审查主张，如果满意，则共同签署——以其自身声誉作担保（图 7）。这种不对称是刻意设计的：获得声誉是缓慢的（每次诚实背书增量 $+\delta$），而失去它则是灾难性的（每次虚假背书 $-\Delta \gg \delta$；示意性地，例如 $+2\%$ 对 -70%）。最优策略始终是拒绝可疑主张。δ 和 Δ 的具体数值是系统参数。

6.3 多维声誉

声誉不是一个标量，而是一个跨多个维度的向量：财务可靠性、个人诚信、专业能力、公民参与，以及其他（图 8）。不同的评估提供者会根据情境以不同方式投影这一向量——放贷者优先考虑财务

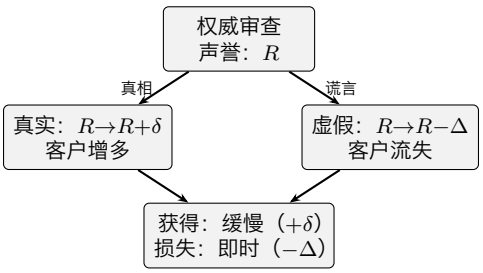

Figure 7: 可信权威——非对称激励。

可靠性，雇主看重专业能力，邻居关注公民行为。不存在单一“正确”的声誉分数；只有各种视角。

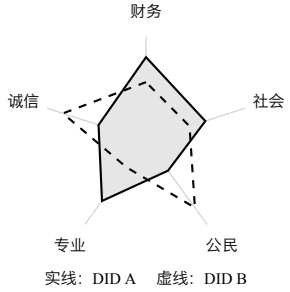

Figure 8: 多维声誉向量。不同的 DID 在各维度上呈现不同的画像。

6.4 民主化的风险评估

RSN 使系统性风险评估民主化——这一能力目前集中于金融机构，但其适用范围远超银行业。工业集团评估供应商可靠性、保险公司为行为风险定价、并购中的尽职调查、制造业中的设备寿命估算——只要交易对手风险需要评估，RSN 就提供一种去中心化、市场驱动的替代方案。一个解读服务市场将原始的多维声誉数据转化为可付诸行动的摘要，使任何参与者都能以边际成本进行交易对手评估。

7 共识与纠纷解决

7.1 去中心化正义模型

RSN 将正义重新定义为一个双边协商问题。永久且可验证的声誉损害之威胁，比受害方无力负担的法律诉讼更能起到威慑作用。

7.2 涌现式共识

每个参与者都维持一个个人满意度阈值。网络的总体共识作为数千次双边协商的统计平均值而涌现（图 9）。远高于共识（野蛮）的回应发布起来昂贵。接近共识（相称）的回应则廉价。共识不是一条规则——它是一个价格信号。

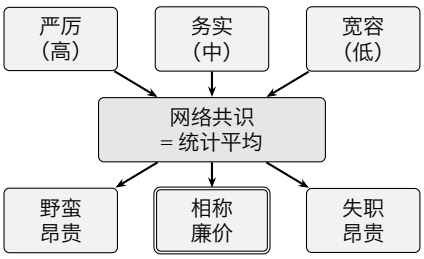

Figure 9: 从个体满意度阈值中涌现的共识。

7.3 惩罚校准

每个 DID 持有者都公开声明其对特定不当行为类别的回应。过于严苛或过于宽纵的声明会招致负面声誉信号。相称的声明则毫无摩擦地被接受——一种自我校准的惩罚系统。

共识声明本身也受制于伪善检测：在声明上承诺某个共识立场，却行为不一致，构成一种比底层偏离更严重的声誉过失，因为它展示了蓄意的欺骗。

7.4 委托

一个 DID 内的所有活动——只有一个例外——都可以委托给另一个 DID。**主体角色——即主张所针对其行为的那个 DID——不可委托；它不可转让地绑定于主张所指的身份持有者。**其他活跃角色——发起者、权威、见证者、验证者——都可以转移给一个指定的受托 DID，无论是用于验证、主张提交、共识参与还是纠纷解决。委托可随时撤销。这使得专业化成为可能（例如专业验证服务），同时持有者保留最终的控制权和责任。委托适用于整个系统，并且对可扩展性至关重要。

7.5 从个体道德到涌现式社会契约

每个 DID 声明的策略代表着个体道德的形式化：持有者认为什么是可接受的、他们如何回应特定行为、他们对交易对手有何要求。这些策略并非由系统设计者强加——它们由每个参与者选择，并以机器可读的形式表达。

这种形式化促成一种三阶段涌现。第一，个体道德被编码为**策略**——一套 DID 承诺遵守的声明规则、阈值和回应函数。第二，网络中所有策略的总和产生**涌现式伦理**——统计上可观察的规范，没有任何单一参与者设计过它们，但它们从数千个个体道德立场的互动中产生 [9]。第三，这些涌现式伦理，作为通过双边价格信号执行的共享行为规范来表达，构成一份**可度量的社会契约**——不是一个无人签署的理论构想 [10]，而是一套由实际行为支撑、可经验观察的规则。

这一过程呼应了道德基础理论 [11] 的发现：人类道德是直觉的、多元的、且随文化而变的。RSN 不要求道德共识作为输入；它产生行为共识作为输出。由此产生的社会契约并非静态——它随着参与者的加入、离开以及根据网络反馈调整其策略而演化。与经典的社会契约理论不同，这份契约是可撤销的、可观察的，并且无需一个主权者即可执行。

8 经济模型

前面各节描述了一种工具——RSN 的验证机制、成本结构和涌现式共识。本节描述一种将此工具应用于财政和治理转型的方法论。工具是通用的；下述方法论是一种可能的应用。

8.1 激励结构

作为资本的声誉为诚实行为创造了直接激励。在正常运作中，参与者通过日常交易和履行的承诺自然地建立声誉。主要支出用于负面主张——记录他人所造成的伤害。这种不对称激励有益且可靠的行为：诚实不需额外成本，而有害则会留下他人愿意付费保存的记录痕迹。伪善——声明规则却不遵守——是代价最高的失败模式，因为它展示了蓄意的欺骗。

8.2 并行财政系统

三个组件促成竞争压力：（1）**简单税**——单一的比例税率，无任何例外，初始略低于现有的实际税率；（2）**电子支出登记（ESR）**——将捷克 EET 模式反转，使公民监督国家支出；（3）**公民定向税款分配**——在第一年从百分之几起步，十年之后达到从低两位数到完全迁移至公民定向系统之间的任何水平，取决于采用速率。实际节奏取决于国家服务的自由市场替代品出现的速度，以及国家配合这一转型的意愿；时间的函数不必是线性的，也没有理由为采用最终可能达到的水平设定上限。

9 迁移路径

迁移经历三个阶段（图 10）：**阶段 1：叠加**（RSN 作为信息层，国家是唯一提供者）、**阶段 2：竞争**（RSN 提供功能性替代方案，双系统并用），以及**阶段 3：转型**（RSN 表现优于国家等价物，自愿迁移）。这一过渡在每一个阶段都是可逆的。

主要的压力机制是财政性的：当有条件的纳税人达到“经济上举足轻重的少数 X”——一个大到足以使针对它的镇压造成非平凡经济损害、并使公民不服从成为可信威胁的群体——时，国家便

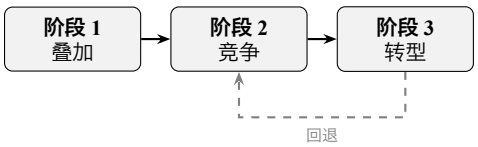

Figure 10: 三阶段迁移——在每一个阶段都可逆。

进入谈判。作为示例，我们采用 $X \approx 15\%$ 的 GDP，但实际门槛取决于具体经济体。

10 安全性分析

我们考虑五类对手：个体不良行为者、有组织团体、企业对手、国家级对手，以及协议级对手。

女巫抵抗 [12]。建立声誉需要持续、可验证的互动。一次成功的女巫攻击的成本随伪造身份数量线性增长，随目标声誉水平二次方增长。并行操作多个 DID 是可能的，但设计上代价高昂——每个身份都必须通过真实活动独立累积声誉。在自由社会中，这一成本遏制随意的滥用；在独裁政权下，并行 DID 反而成为一种生存工具，使得分隔化的抵抗、地下协调和更安全的黑市周旋成为可能。

合谋防御。封闭团体之间相互背书的模式可通过社交图谱分析被检测。声誉聚合函数会对来自紧密聚集来源的主张打折扣。

国家级对手。洋葱路由、中心化基础设施的缺失，以及验证者角色在参与者群体中的分布，确保了压制需要与任何收益都不成比例的措施。

隐私与问责。匿名化——介于完全匿名与身份公开之间的中间地带——允许 DID 在不暴露持有者现实世界身份的情况下累积有意义的声誉。

11 隐私考量

RSN 的隐私模型建立在匿名化之上。持有者控制身份披露：最小（纯匿名）、选择性（关联特定凭证）或完全（关联法律身份）。已发布的主张是永久的——系统支持情境化的更正，但不支持抹除。持有者可以放弃一个已被泄露的 DID 并重新开始，代价是丧失累积的声誉。

12 相关工作

W3C DID Core 规范 [2] 和 Ceramic 网络 [8] 提供了身份基础。EigenTrust [13] 演示了无需中央权威的分布式声誉聚合。SBT [3] 引入了不可转让的社会代币。RSN 的不同之处在于，它强调将经济成本作为质量过滤器，并具有为激进主义定价的机制。

DAO [4] 和二次方投票 [5] 演示了去中心化治理的可行性。RSN 从双边价格协商而非投票中导出共识。

该提案借鉴了无政府资本主义理论 [14, 15] 中关于私营服务提供的观点，但在两个关键方面有所背离：它为怨恨和复仇冲动而设计，而非假定理性；并且它提出渐进转型而非革命。涌现式社会契约的概念在哈耶克的自发秩序理论 [16] 中有其先声。

无政府阿哥拉主义。 RSN 与阿哥拉主义反经济学 [17] 有显著的共同点——尤其是在强调建立并行制度和自愿交换方面。然而，阿哥拉主义倾向于假定理性的自利是一种充分的协调机制，并且往往缺乏一条从现有国家结构出发的具体迁移路径。RSN 明确纳入非理性的行为倾向，并为渐进转型提供一种财政压力机制。

精英体制。 RSN 也许代表了对精英体制 [18] 如何作为一种系统属性而非一句口号涌现的首次功能性描述。传统的精英体制之所以失败，是因为它们需要一个中央权威来定义和强制执行“功绩”。在 RSN 中，功绩从双边评估中涌现：那些创造价值的人累积声誉；没有委员会来裁定谁有功绩。

作为特权的财产。 RSN 从根本上背离了无政府资本主义和奥地利学派将财产权视为自然且绝对的处理方式。在 RSN 框架中，财产是一种特权——一种社会认可，在极端情况下，当参与者从根本上违反共享规范时（背叛、在生存危机中拒绝保卫社区、系统性剥削），社区可以将其撤回。这不是国家征收，而是双边关系网络对社会认可的撤回。

13 讨论与局限

冷启动。 RSN 的价值取决于网络效应；早期采用者在参与达到某个门槛之前面临有限的价值。然而，即便在早期阶段，DID 网络也可作为一个有用的补充信息来源。早期的声誉评估和权威评估预计将随着日益成熟而逐步发展。

反吸血与访问控制。目标 DID 可对来自低声誉 DID 的查询施加分级费用和访问限制。这不是二元的，而是一个连续的尺度：查询方 DID 的声誉越少，它接收到的信息就越少、等待越久、支付越多。这一机制激励真实参与，而非被动消费。

访问的不平等。发布主张的成本可能会将来自经济弱势参与者的正当申诉过滤掉。缓解措施：每个参与者同时充当潜在的验证者（或委托这一角色）并赚取验证费。在足够长的时间跨度内，一个非激进的参与者应当接近财务中性——验证收入大致抵消发布成本。这是一种统计上的预期，而非保证。

声誉的不平等。早期采用者累积的优势可能为后来者制造障碍。然而，声誉评估由第三方提供者

执行，他们可以选择通过其聚合算法来缓解先发优势。以良好声誉率先入场是一种正当的比较经济优势——而这正是设计使然。

开放问题。最优成本函数、聚合标准、协议治理，以及与 AI 生成的合成声誉数据的交互，仍是未来工作的领域。

本文并不声称 RSN 会在所有情况下产生最优结果。它声称的是，RSN 代表一种可行的替代方案——技术上可实现、经济上可自我维持，并且在社会上与人类行为倾向的本来面目相容。

14 结论

本文提出了声誉社交网络，一种使匿名化、可验证的声誉交换成为可能的去中心化协议。昂贵激进主义原则确保欺诈性主张在不经审查的情况下因成本过高而被排除。所提出的迁移路径使从现有机构进行渐进、可逆的转型成为可能。这一设计将人性的本来面目纳入其中——包括琐碎、怨恨以及对复仇快感的渴望——而非要求意识形态上的转变。其结果不是乌托邦，而是一个正义可及、声誉靠赚得、不当行为的成本由造成它的一方承担的系统。

References

[1] Czech Republic, “Act no. 549/1991 coll., on court fees, as amended,” 1991, fee schedule for court proceedings; cost rules per Act No. 99/1963 Coll. (Code of Civil Procedure) and Decree No. 177/1996 Coll. (Attorney Tariff).

[2] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, and C. Allen, “Decentralized identifiers (DIDs) v1.0,” W3C Recommendation, Jul. 2022.

[3] E. G. Weyl, P. Ohlhaver, and V. Buterin, “Decentralized society: Finding Web3’s soul,” *SSRN Electronic Journal*, May 2022.

[4] V. Buterin, “DAOs, DACs, DAs and more: An incomplete terminology guide,” Ethereum Blog, May 2014.

[5] V. Buterin, Z. Hitzig, and E. G. Weyl, “A flexible design for funding public goods,” *Management Science*, vol. 65, no. 11, pp. 5171–5187, 2019.

[6] D. Karger, E. Lehman, T. Leighton, R. Panigrahy, M. Levine, and D. Lewin, “Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web,” in *Proc. 29th ACM STOC*, 1997, pp. 654–663.

[7] D. Buchner, “ION — a scalable DID method based on Bitcoin,” Microsoft, 2021.

[8] Ceramic Network, “Ceramic protocol specification,” GitHub, 2023. [Online]. Available: https://github.com/ceramicnetwork

[9] É. Durkheim, *The Division of Labor in Society*. Free Press, 1893, english translation 1984.

[10] J. Rawls, *A Theory of Justice*. Harvard University Press, 1971.

[11] J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage Books, 2012.

[12] J. R. Douceur, “The Sybil attack,” in *Proc. 1st International Workshop on Peer-to-Peer Systems (IPTPS)*, 2002, pp. 251–260.

[13] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, “The EigenTrust algorithm for reputation management in P2P networks,” in *Proc. 12th International Conference on World Wide Web*, 2003, pp. 640–651.

[14] M. N. Rothbard, *For a New Liberty: The Libertarian Manifesto*. Macmillan, 1973.

[15] H.-H. Hoppe, *Democracy: The God That Failed*. Transaction Publishers, 2001.

[16] F. A. Hayek, *Law, Legislation and Liberty*. University of Chicago Press, 1973.

[17] S. E. I. Konkin, *An Agorist Primer*. KoPubCo, 2006.

[18] M. Young, *The Rise of the Meritocracy*. Thames and Hudson, 1958.