Yellow Paper: Nebulas Rank

Nebulas Research

June 2018
Version:1.0

Contents

1 Introduction

Nowadays, more and more scenarios benefits from *decentralization*, which is the core of blockchain systems. For example, Bitcoin, the origin of blockchain, has proven it's significance to digital assets, while Ethereum has proven how important is decentralization to DApps. And there are more and more blockchain projects explore how they can leverage decentralization.

Obviously, the backbone of decentralization in blockchain is the openness and features of anonymity.

Yet, openness and anonymity obstruct the emergence of value measurements [1]. There are two aspects. First, it is difficult to infer if some accounts belong to the same user, which means it is difficult to build a mechanism like HTTP Cookie [2], or to use traditional data analysis technologies to understand user characteristics. Second, the openness of blockchain makes it vulnerable to manipulation, especially for value measurements. Attackers can easily get all details about the value measurements, and figure out the weakness of the whole system. This largely differs from traditional value measurements which are close or independent.

We believe that the effective value measurement is the foundations of blockchain's prosperity. Both the lack of and ineffectiveness of value measurement may confine blockchains to limited use cases.

First of all, we need a methodology to quantify the value of data, applications and accounts on blockchains. The root cause is cooperations on blockchain keeps scaling up, and the requirements of efficiency keeps growing. Without value measurement, such collaboration may be negatively affected.

Second, blockchains are still at the very early stage, and the value of data and assets on the blockchains is still underground and waiting to be found. Effective value measurements will uncover the value and empower more applications and enable more application scenarios, for example, loans, credit, data search, personalized recommendation and cross-chain interaction.

Third, incentives, which is based on value means, is necessary to healthy blockchain ecosystems. Without effective value measurements, incentives may lead a blockchain system to corruption and eventual collapse.

As a conclusion, an effective value measurement for blockchain needs to be

- **Truthful.** The rank needs to measure some characteristic of a blockchain sys-

tem, and thus can be trusted in some way;

- **Fair.** This means the rank need to be manipulation-resistant, and it is the core of the rank algorithm;
- **Diverse.** There will be different ranking requirements from different applications on blockchain, thus a good rank algorithm should cover different scenarios.

We believe Nebulas Rank shall be an effective value measurement for blockchains.

For truthfulness, we define Nebulas Rank to be quantification of an account's contribution to the blockchain system after considering many different metrics.

We believe that cryptocurrencies should have the attributes of money, and three functions of money: medium of exchange, store of value, and unit of account. Blockchains themselves are economic systems and the classical monetary theory still has the instruction value. Furthermore, we believe the value of cryptocurrencies comes from the liquidity. Specifically, each transaction between users increases the liquidity of cryptocurrencies, and endows the value of cryptocurrency eventually. Thus, the on-chain transactions are effective and natural data sources for effective value measurement.

To evaluation the effectiveness of Nebulas Rank, we calculate the sum of all accounts' Nebulas Rank on Ethereum, and compare it with the market capital given by `coinmarketcap.com`. Our evaluation shows strong correlation between them, about 0.84. That means Nebulas can measure accounts' contribution at the micro-level, while it can measure the value of blockchain systems at the macro-level.

For justness, we involve a special function to resist manipulation, and our analysis demonstrates its performance to be manipulation-resistant.

Based on the theory of Nebulas Rank, we divide Nebulas Rank into Core Nebulas Rank and Extended Nebulas Rank for different applications and scenarios.

Core Nebulas Rank defines the algorithm to calculate an account's contribution to the whole blockchain system in a certain period of time. And such calculate involves two factors: the median stake of an account in a certain period, and the in-and-out degree of the account in a certain period.

Extended Nebulas Rank is for different applications and scenarios, and it is based on Core Nebulas Rank. For example, we show how to rank smart contracts based on Core Nebulas Rank; we also show how to extend Core Nebulas Rank to a multi-dimensional vector.

Besides theory and methodology of Nebulas Rank, we also present our consideration about how to implement Nebulas Rank, including whether to put ranking scores

on-chain, how to update the algorithm of Nebulas Rank, and our future work on Nebulas Rank.

> Special Hint: The content is this yellow paper may be different from the description in our whitepaper (version 1.02 released on April 2018) [3]. This is because we keep thinking and verifying the algorithm in our whitepaper. And now we are more confident and capable to make it more rigorous. We use a different format (like this paragraph) to emphasize the relevant updates presented in this yellow paper.

2 Background

In this chapter, we introduce the background of blockchain and associated technology. Due to the absence of value measurement, we discuss the implementation of typical ranking algorithms in the area of blockchain and their drawbacks.

2.1 The Development Status of Blockchain

Satoshi Nakamoto published Bitcoin whitepaper [4] in October of 2008. As the earliest application of blockchain, Bitcoin is the most striking example of the concept of a *decentralized cryptocurrency system*. The production of Bitcoin is depend on massive computations executing a special algorithm instead of any organization, which guarantee the consistency in the distributed ledger system.

With specific scripting language, Bitcoin can be used for third-party payments, efficient micro-payments, and so on. Then, a wave of experiments originating from Bitcoin emerged which include features more complex than the basic currency property. For example, Namecoin [5] represented a distributed Domain Name System and others like the Open Assets [6] based on colored coins, both are copy of intelligent assets which follow the traceability of Bitcoin.

Unfortunately, the scripting language of Bitcoin has many design flaws, such as lacking of instructions and failing in Turing-complete, limiting its usefulness.

With the development of blockchain technology, more successors have merged and tried to extend the functions related to different applications. The most significant one is Ethereum [7], providing Turing-complete smart contracts, which opens new possibilities of applications.

Smart contracts are the contracts enforced by technical method in blockchain system. The Ethereum smart contract runs on the Ethereum Virtual Machine (EVM),

which isn't in the control of any entity, and EVM ensures the consistency of output as well as smart contract itself via consensus algorithm.

People can develop distributed applications (DApp) with complex functions based on the Ethereum smart contract. These DApps provide the solutions for various fields other than basic transactions, such as voting, crowdfunding, lending, property rights and so on. However, even if Ethereum extends the possibility of blockchain application, there is no killer apps in Ethereum platform because of the lack of value measurement.

For a system that supports smart contract, there are two kinds of account, externally owned accounts (EOA) and smart contract accounts, and both lack reasonable value measurement. In the meantime, invaluable information is usually concealed in the invocation process of smart contract. The information has more dimensions compared with traditional transaction data, and cannot be evaluated by classical value measurements.

In early 2015, Chris Skinner came up with the idea of *value web* [8], noting that a value ecosystem should include value exchanges, value stores and value management systems. Chris points out that there are clear difference between cryptocurrency platform and traditional society in value measurement, which poses challenge to evaluating the value of data and information in the cryptocurrency platform.

2.2 Node Ranking Algorithms Based on Graph

The new generation of blockchain projects such as Ethereum build a complex ecosystem, more than a cryptocurrency trading platform. However, there is no reasonable method to evaluate the value of entity on chain. For example, we have no idea about which one has the bigger contribution to the blockchain system and how to measure these contributions.

Here, we introduce PageRank algorithm [9], a typical reputation measurement on the Internet at first. As early Google's core algorithm, PageRank is proposed to solve the ranking problem in web link analysis. With the development of research on PageRank, it has been widely used in many fields, such as to rank the importance of academic papers, web crawlers, keywords extraction, user reputation ranking in social networks, and so on.

Some research focuses on using PageRank on blockchains. Fleder, Kester, Pillai et al. use PageRank to discover the Bitcoin account address and analyze its activities [10]. However, their main method is just manual analytical work with the help of PageRank.

As the classical ranking algorithm formed in web 2.0, PageRank suffers limitations in online reputation evaluation.

More research improving on the PageRank has emerged, and one of the most famous is LeaderRank [11]. LeaderRank improves the transition probability by introducing ground node and weighted bidirectional links instead of using the same transition probability in PageRank, which makes the nodes have different transition probability in and out. But there are limits: LeaderRank counts the reputation ranking iteratively with the consideration of relation between nodes only, while lacking evaluation of user activities.

Note that these kinds of PageRank algorithms are not resistant to Sybil attacks [12] which is the strategy where an adversary subverts the reputation system in symmetric network by creating a large number of pseudonymous identities.

The most relevant work with Nebulas Rank is NEM [13]. Different from Bitcoin's Proof-of-Work and Ethereum's Proof-of-Stake consensus strategy, NEM adopts Proof-of-Importance consensus protocol and NCDawareRank [14] as the ranking algorithm. The NCDawareRank exploits the clustering effect of network topology with clustering algorithm based on SCAN algorithm [15] [16] [17]. Although community structure does exist in transaction graph and should be helpful to handle with spam nodes, it does not guarantee that all nodes on blockchain controlled by one entity in the real world are mapped into one cluster, which leads to large room for manipulation.

2.3 Manipulation Resistance

The ability of resisting manipulation, a.k.a. truthfulness, is the most significant and challenging goal of Nebulas Rank.

Hopcroft et al. find that PageRank fails at evaluating user reputation under manipulation [18]. Zhang et al. point out that, the adversary can diminish the degree of non-sybil users reputation effectively even if the evaluation index of node reputation is built [19].

This is because PageRank algorithms work based on the network topology, while the adversary could get the same or higher reputation score by creating an image network [20] [12].

With blockchain systems, some manipulation methods are as follows:

1. Loop transfer. The attacker transfers along a loop topology, which allows the same money flow over same edges repeatedly. By this means, the attacker hopes to raise the weight of related edges;

2. Transfer to random addresses, so that the out-degree of sybil node is increased, and the propagation of fund is increased as well;

3. Form an independent network component with addresses controlled by the attacker, so the attacker can pretend to be a central node;

4. Interact with authoritative exchange service addresses frequently, i.e. transfer the same money in and out an authoritative exchange service address repeatedly, so that the attacker can acquire better structural position in the network.

We should take these into consideration to keep the fairness of Core Nebulas Rank during the design stage.

3 Economic Model

Cryptocurrencies are endowed with economic significance, either as a kind of trading medium or intelligent asset. Therefore, a reasonable economic model can help us to establish a value measurement standard on the blockchain, which is also the objective of Core Nebulas Rank. This chapter first introduces the mathematical representation of cryptocurrency, and then analyzes the cryptocurrency with a simple but well-recognized monetary model. During the analysis, we introduce the Core Nebulas Rank as an important argument.

3.1 Representation of Cryptocurrency

The biggest difference between cryptocurrency and traditional economy is that all the transactions on the cryptocurrency can be traceable. This provides the data sources for us to analyze the impact of each transaction on the economic system.

In general, a cryptocurrency system can be defined as a pair $(\mathcal{L}, \mathcal{U})$, where $\mathcal{L}$ is the ledger system, and $\mathcal{U}$ is the set of cryptocurrency users. Further, the ledger system can be described as a triple as below:

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

where $\mathcal{A}$ is the set of accounts, $\mathcal{D}$ is the set of initial balances of each account, and $\mathcal{T}$ is the set of transactions. Each transaction can be recorded as a tetrad as below:

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

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

where $a \to d$ represents the balance d corresponding to the account a (d is a positive real number, in other words, we do not takes the accounts with zero balance into consideration). s, t, w and τ represents the source account, target account, amount and time of a transaction respectively.

An account is controlled by a related user, who can propose a transaction with the account, which can be denoted as:

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

On one hand, a user can control multiple accounts, represented as:

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

On the other hand, an account can only be controlled by a single user, shown as:

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

Note that the model described above is a reasonable simplification of any cryptocurrency system. In the model, we do not distinguish the on-chain data from off-chain data, and do not introduce either transaction price or invocations of smart contracts and so on. In addition, the accounts of exchanges are type-specific. Generally speaking, the transactions in an exchange can be divided as two categories: normal transactions that will be recorded on the chain, and intra-exchange transactions that will not be recorded in a centralized database of the exchange. This leads to an outcome where we will lose the intra-exchange transactions if we only obtain the data from the chain. However, if the intra-exchange transactions can be obtained with the cooperation of the exchange, we can further map an exchange account into multiple accounts, so as to use the model above.

3.2 Model of Cryptocurrency

Although the cryptocurrency differs largely from the traditinal commodity currency and fiat money, the classical monetary theory still has the practical leading meaning nowadays. As the current money of a new economic entity [21], cryptocurrency is born with the attributes of the money and has three functions of money: medium of exchange, store of value, and unit of account.

Hereby, we establish a simple and classic monetary model to help us understand the physical significance of Nebulas Rank.

First of all, we try to give the indicator to measure the *velocity factor* in the cryptocurrency ecosystem.

Another concept needed to be differentiated from the *velocity factor* in the economics is *liquidity*. *Liquidity* is used to describe the difficulty level of exchanging the assets for the medium of exchange. As money itself is a medium of exchange in the economics, money is the assets with the best *liquidity*.

> In the Nebulas Technical White Paper [3], we used the word *liquidity*. However, there is no rigid definition of *liquidity*, whose meaning is very broad even in the economics. For example, the entries to explain the *liquidity* includes three totally different aspects in *The New Palgrave: A Dictionary of Economics*. R. S. Kroszner pointed out that there were 2795 independent papers mentioning *liquidity* during the past 6 months, each of which raised a typically different statement though [22]. The *liquidity* in this yellow paper is referred to as the **velocity of money**, meaning the turnover times of a monetary unit within a certain period of time.

We use the velocity of money to represent the turnover rate of cryptocurrency [23], namely the turnover of a monetary unit within a certain period of time (one day in this paper), which is represented with V. According to the classical quantity theory of money, the equation is expressed as below:

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

where M, V, P and Y represent the total monetary amount of the economical system, the velocity of money, the price level (measured by the money of unit economical output, thus the money price is $\frac{1}{P}$), and real economical output (real GDP) respectively.

The equation illustrates that the product of monetary amount and velocity of money equals the product of price of goods and their output.

As for the monetary amount M, Nebulas is similar to Ethereum in that the monetary amount maintains steady growth (the additional issuance percentage of Nebulas money is set as 4% at present), which is different from Bitcoin as the total monetary amount of latter will be stable at 21 billion at last. The velocity of money V can be described as the ratio of the circulated monetary amount and the monetary supply. As a result, the equation 7 can be further expressed as:

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

where Δm is the additional monetary supply.

In terms of price level P, it is acceptable that the value of price is determined by the relationship between the monetary supply and demand, both by the classical theory of money and New Keynesian Models. In the long term, the total price level will be adjusted to make the monetary supply and demand equal.

However, the total price level does not always construct a balance between monetary supply and demand in the short term. In a healthy economical system, the growth rate of price level is usually smaller than that of velocity of money. By increasing the monetary supply (reduce the interest rate in other words), both the price level P and goods/service demands Y will increase in the meantime. On the other side, the increase speed of price level should be controlled, to prohibit the users from holding the cryptocurrency for a long time, thus reducing the velocity. The reason for the users to hold the cryptocurrency is that they expect the price of cryptocurrency will raise.

With regard to the real economical output Y, it is usually represented as real GDP by economists, namely *a monetary measure of the market value of all final goods and services produced in a period of time*. We believe that the value of cryptocurrency is based on its velocity, namely each transaction makes contributions to the total economic aggregate to a certain extent. In other words, once a transaction takes place, it both increases the velocity of cryptocurrency and people's approval and belief of cryptocurrency to some degree. As a result, we think that Y in the equation 8 is consisted of each transaction. Given that the subjects of a economical system are accounts, we can also explain Y as the transactions issued by each account as below:

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

where $\mathcal{C}(a)$ represents the contributions made by account a to the economical output, namely Core Nebulas Rank.

The development of cryptocurrency relies on the development of the community. Therefore, we consider that quantifying the contribution made by each account is the basis of designing the reasonable incentive mechanism. Based on this, the economical system can create either explicit incentives (e.g., Proof of Devotion in Nebulas technical white paper) or implicit incentives (e.g., the sorted search results provided by search engines). The directive and primitive incentives in the cryptocurrency is the additional issuance of money, which is different from that in the traditional monetary theory.

4 Core Nebulas Rank

Core Nebulas Rank is used to measure the contributions of a user to the whole economy **in a certain period of time**. It is quite complicated to calculate the contribution precisely, so we provide an approximation algorithm for it. In this approximation algorithm, we consider two critical factors, the coinage and the account position information in the transaction network. And we will proof the effectiveness of the approximation algorithm in the evaluation section below.

We use the transaction history on the mainnet in a certain period as the data source of Core Nebulas Rank. All the transactions in a period of time $[t_0 - T,\ t_0]$, can be specified as a set:

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

Based on $\Theta(t_0)$, we can define a weighted directed graph, the node is the address of the account, the edge from node s to node d represents one transaction, the weight of the edge is w, the time of the edge is τ.

For account $a \in \mathcal{A}$, the calculation of Core Nebulas Rank $\mathcal{C}(a)$ is based on $\Theta(t_0)$, which can be represented as:

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

$\beta(a)$ is the median stake of account a in a certain period; $\gamma(a)$ is the in-and-out degree of account a in a certain period.

> Different from the way we calculate the Core Nebulas Rank in Nebulas white paper [3], we made some updates blow:
> 1. We don't use the top K highest transaction amount as the weight when building the transaction graph;
> 2. We don't rely on the weight of nodes in LeaderRank to get the importance of the node.
> First, we remove the transaction loops before we calculate the in-and-out degree β, so it can resist a loop attack. At the same time we still consider the strength of the edge. For some cases of homogeneous topology graph, PageRank and some other symmetric function (such as LeaderRank) has been proved to not be able to resist sybil attack [20]. In this yellow paper, we didn't use the Topology-like ranking strategies. We propose a asymmetric calculation function 21 which is effective for reducing the rewards by faking the low stake nodes in § 4.3.

Below, we are going to discuss three issues in equation 11: Median Account Stake $\beta(a)$, In-and-Out Degree $\gamma(a)$, and selection of function Ω and Ψ.

4.1 Median Account Stake $\beta(a)$

In time period of $[t_0 - T, t_0]$, there are n blocks in the blockchain system, marked as:

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

B_i is the parent block of B_{i+1}. For account $a \in \mathcal{A}$, the balance of the account at end of each block is

$$d^a_0, d^a_1, \ldots, d^a_n$$

We can get a new list by sorting the items ascending

$$d^a_{(0)}, d^a_{(1)}, \ldots, d^a_{(n)}$$

where $d^a_{(i)} < d^a_{(i+1)}, 0 \leq i \leq n-1$, thus, the $\beta(a)$ can be expressed as:

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

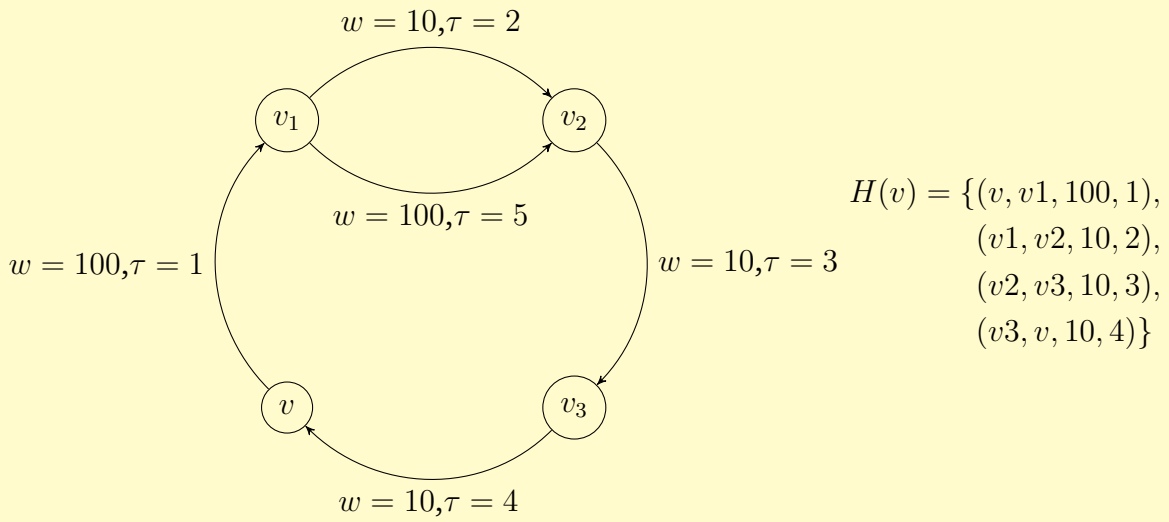

Figure 1: Forwarding loop in a transaction

The median account stake represents the coinage in a certain way, that means the account need to hold the stake for more than half of the time period.

4.2 In-and-Out Degree $\gamma(a)$

Consider the adversary would increase the in-and-out degree by using loop attack, so, we will need to remove the forwarding loop before we calculate the In-and-Out degree for the transaction graph. Forwarding loop is a loop of transaction in a sequence of time. It starts and ends on same node v, which is a set of edges in the transaction graph. A forwarding loop can be marked as $H(v)$, which is

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

where $\forall 1 \leq i \leq n : \tau_i \leq \tau_{i+1}$. As shown in Fig. 1, there is a forwarding loop, and note that transaction $(v_1, v_2, 100, 5)$ is not included in the forwarding loop.

After figuring out the forwarding loop, we need to remove the loop before use it. Assuming that there are n forwarding loops in the system, and the forwarding loops are listed by the sequence of occurrence as below:

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

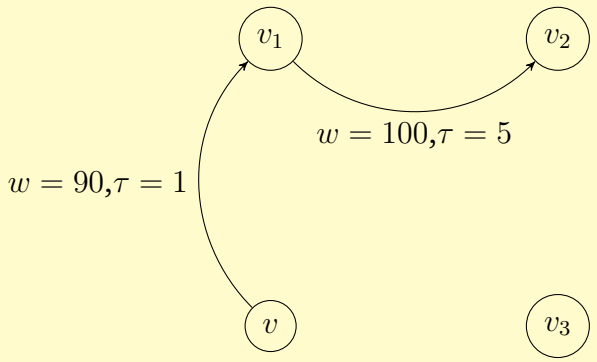

Figure 2: The transaction graph after removing forwarding loop in Fig. 1

The minimal amount of the transaction in $H^i(v_i)$ is $(s^i_m, t^i_m, w^i_m, \tau^i_m)$, and

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

Then, for each transaction in $H^i(v_i)$, we need ot minus the minimal transaction amount w^i_m accordingly and remove this transaction if the latest transaction amount is 0, which is

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

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

Fig. 2 shows the non-loop transaction graph after removing the forwarding loop in Fig. 1.

Set the transfer-in amount of node v as $p(v)$, then

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

Similarly, transfer-out amount of node v is

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

In this case, for node v, its in-and-out degree $\gamma(v)$ is

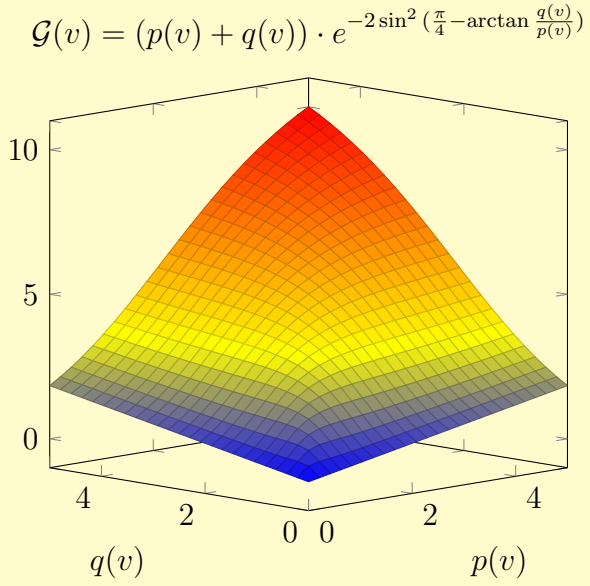

Figure 3: The curve of the In-and-Out degree function

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

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

where θ, μ, λ are the parameters to be determined.

And Fig. 3 shows the curve of the function 14.

4.3 Wilbur Function

It would be extremely complicated to calculate Core Nebulas Rank if we consider different usage case and its properities. However, we can provide a general function for Nebulas Rank.

We define the Core Nebulas Rank calculation function as $f(x)$, namely *Wilbur Function*[1],where x is the factor of Core Nebulas Rank, it can be account stake, coinage or the in-and-out degree. $f(x)$ satisfies flowing two properties:

[1]The name *Sybil Attack* derives from 1970's TV mini-series Sybil, in which an young woman is diagnosed as suffering from muliple personalities and receives treatment from psychiatrist named Dr. Cornelia Wilbur.

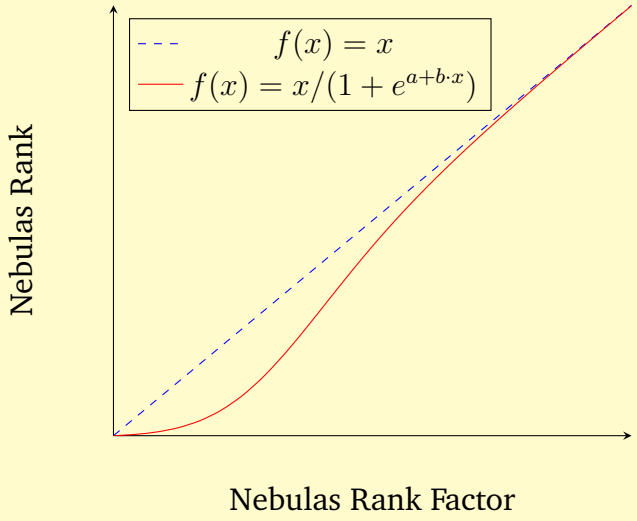

Figure 4: The curve of the Nebulas Rank function

Property 1. *For any two variables a and b, which are both larger than 0, the sum of the two functions is smaller than the function of sum of two variables.*

$$f(a+b) > f(a) + f(b) \quad a > 0, b > 0 \tag{18}$$

Property 2. *For any two variables are infinity, the sum of the two functions is approximately equal to the function of the sum of the two variables.*

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

These properties described above ensure, under given transaction behaviors, the benefits of splitting stakes into smaller accounts is smaller than keep them in single account. At same time, when the stake is larger enough, the cost of splitting the stakes into small accounts can be ignored.

There is more than one function that satisfies the two properties above. Here, we provide a succinct function, the curve of the function is shown in Fig. 4.

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

Detailed proofs for the function are given in Appendix A

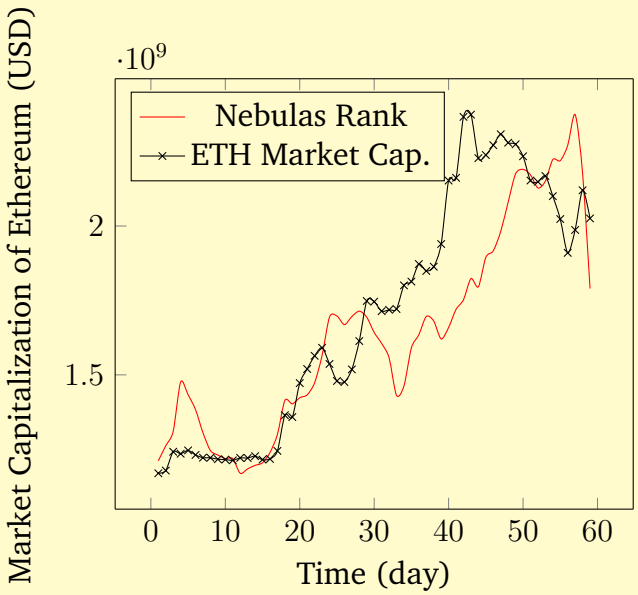

Figure 5: The market capitalization and Core Nebulas Rank of Ethereum

In summary, equation 11 can be expressed further as below:

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

where a, b, c, d are parameters to be determined.

In order to verify the effectiveness of the function, we calculate the Core Nebulas Rank for all the accounts in Ethereum in certain period of time. We collected all the transaction records from May 1st 2017 to June 30th 2017 (block height: from 3629091 to 3955158), besides, we also collected average daily ETH token price (in USD) and transaction volumes [24].

Fig. 5 shows the trending of ETH market capitalization and Core Nebulas Rank of the Ethereum, where the black solid line indicates the market capitalization (in USD) of Ethereum, while the red solid line represents the summation of all accounts' Core Nebulas Rank based on function 21.

We can see that the Core Nebulas Rank reflects the market capitalization changes of Ethereum precisely. The correlation coefficient is 0.84427, p (p-value) is $4.48 \times 10^{-17} < 0.001$. That means, the function 11 shows the success in depicting the contributions of users to the economic system on chain, which demonstrates the validity of Core Nebulas Rank.

5 Manipulation-resistance of Core Nebulas Rank

This chapter is the analysis on how Core Nebulas Rank resists manipulation, i.e. the fairness of Nebulas Rank.

Manipulation is the fact that an attacker can take specific actions to obtain most benefit. The action space of attackers is to launch asset transfers, by making use of the assets and accounts controlled by them and their cooperators. Among the transfers, the amount of asset doesn't exceed the asset owned by the attacker; the source of transfer is either the accounts owned by the attacker and its cooperators, or some institutes' accounts who serve as exchanges. Usually, the benefit obtainable is determined by the accounts whose private keys are known by the attacker. A simple case is that the attacker's benefit is the sum of all of these accounts' ranking scores. Of course, it could be noticed that the private keys of institutes' accounts mentioned before are not controlled by the attacker.

The analysis of this section is based on the action space and attackers' benefit in simple case defined above. First, we discuss the upper-bound for a single account's ranking score enhancement. Then we analyze the upper-bound for multiple accounts. Last, collusion is included and we discuss the situation of more than one attacker.

5.1 Ranking Score Enhancement for One Account

In order to raise the ranking score for one account, according to formula 21, the ranking score of account is positively correlated with the amount of assets and in-and-out degree. The amount of assets in the account, i.e. β, has an upper bound, i.e. it is no more than the asset of the total assets owned by the attacker, denoted by β_0. And in-and-out degree γ represents the volume of transfers, which means the attacker needs to increase the transfers amount of one controlled account as much as possible.

The increasing of transfer amount includes two parts: increasing in-degree and increasing out-degree. Increasing in-and-out degree needs two accounts as participants, one of which is the target account whose ranking score is aimed to raise, the other account could either be a controlled account or an uncontrolled account. If it is an uncontrolled account, increasing degree means transacting with other people, this situation is discussed in § 5.3. The other case is that the attacker sends assets to strangers unconditionally, which is too costly that it won't be discussed in this section. Therefore typically, it could be defined that, the actions of attackers mainly focus on increasing the transfers among the accounts controlled by themselves. Since the as-

sets controlled by attackers are limited and the time period for ranking is also limited, it holds that the degree of an account has an upper-bound which is decided by the amount of assets held by the attacker.

As analyzed above, we consider the case of transferring with accounts of the same owner. Based on the computation method 21 as defined in § 4.3, the attacker's benefit will lower down if it split the asset transfers into multiple ones. Thus the attacker will try to make its transaction amount to be as high as possible, i.e. it tries to transfer all assets it owns into the account and then transfer it out all. Due to the cycle-removal algorithm, the attacker's asset cannot be transferred in again during this period. And the in-and-out degree is $\gamma = 2\beta_0$. The ranking score is $\mathcal{C} = \frac{2\beta_0^2}{(1+e^{a+b\cdot\beta_0})(1+e^{c+2d\cdot\beta_0})}$.

Additionally, we consider a more advanced way to manipulate. Consider the case that the attacker manages to acquire the asset again somewhere else by transacting off-line. Then it could transfer the asset into the account again and the upper-bound of in-and-out degree is the asset amount times the number of off-line transactions. Since the ranking time period is limited, the upper-bound of the number of off-line transactions is a constant integer, i.e. γ is bounded by $2T \cdot \beta_0$, where T is a constant integer indicating the length of ranking time period. Therefore the upper-bound of score is $\mathcal{C} = \frac{2T\cdot\beta_0^2}{(1+e^{a+b\cdot\beta_0})(1+e^{c+c\cdot d\cdot\beta_0})}$.

5.2 Ranking Score Enhancement for Multiple Accounts (Sybil Attack)

Sybil Attack refers to that the attacker obtains falsely high ranking score by creating large number of pseudo-identities to tamper the reputation system of P2P network [25].

An entity on a peer-to-peer network is a piece of software which has access to local resources. An entity advertises itself on the peer-to-peer network by presenting an identity. More than one identity can correspond to a single entity. In other words, the mapping of identities to entities is many to one. Entities in peer-to-peer networks use multiple identities for purposes of redundancy, resource sharing, reliability and integrity. In peer-to-peer networks, the identity is used as an abstraction so that a remote entity can be aware of identities without necessarily knowing the correspondence of identities to local entities. By default, each distinct identity is usually assumed to correspond to a distinct local entity. In reality, many identities may correspond to the same local entity. An adversary may present multiple identities to a peer-to-peer network in order to appear and function as multiple distinct nodes. The

adversary may thus be able to acquire a disproportionate level of control over the network, such as by affecting voting outcomes [26].

Here we assume the attacker's payoff is the sum of all accounts controlled by the attacker. Considering the strategy to enhance ranking score for one account, which is analyzed at last subsection, the attacker could apply the same strategy to multiple accounts: starting from any one account, the attacker transfer part of its asset into the next account, finally forming a linked asset flow. In this case, since Core Nebulas Rank requires that no more than valid amount of asset stays in the account for no more than half of the period, by no means the attacker could make β for more than one account to be the total amount of assets owned by it. Thus the attacker should adopt another strategy where its assets are evenly distributed into all its accounts. Suppose the length of link is N, i.e. there are N controlled accounts, and for every account, $\beta = \frac{\beta_0}{N}$. The in-and-out degree analysis is same with § 5.1, the upper-bound of γ is $K \cdot \beta$, where $K = 2 \cdot N$ is a constant integer. Therefore the upper-bound of the sum of all accounts owned by the attacker is:

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

5.3 Coalition Manipulation

The result of coalition manipulation is no different with the case that one attacker owns the original total asset of two attackers. So here we can analyze the case of coalition manipulation by analyzing the consequence of a single attacker's assets increasing.

6 Implementation of Core Nebulas Rank

The complete implementation of Core Nebulas Rank is out of scope of this proposal, so we will just discuss the key points of the implementation here.

6.1 On chain or not?

As explained in previous chapters, the Core Nebulas Rank actually shows each account's contribution to the overall economic aggregation. Normally, each node can calculate the contribution for any specific account, however, do we really need to put NR on chain periodically?

In our opinion, it's unnecessary or unsuitable to put NR on chain, because:

- NR data size will be huge, while it's certainly unsuitable to put it on chain. Even for IPFS, Genaro etc [27] [28], it's inappropriate to store every account's NR periodically either, even data storage is actually their focus.
- It will affect the performance of block generation. The computing complexity of Core Nebulas Rank is high, so it will significantly affect the block generation and verification performance, and eventually, TPS is affected.

Overally, we suggest that each node can calculate the Core Nebulas Rank individually.

However, if each node does the calculation individually, how can we make sure that the Core Nebulas Rank is reliable and trustful. For instance, the node may viciously modify the NR calculation result, and then gives out incentive based on the NR calculation result. For important applications, we should verify the NR calculation result, to assure the fairness of the calculation result; on the other hand, for those applications that are not so important, it depends on the applications themselves that how they use the NR result and whether they want to verify the NR result.

The other important situation we should also consider is: the node may refuse to calculate the NR with the concern of power consumption. Thinking about that, a trustful Core Nebulas Rank service can be introduced, so, repeated calculation can be avoided. We can either offer the service for free, or charge by number of times. Complete implementation and service detail are out of scope in this paper.

6.2 Core Nebulas Rank Upgrade

As we all know, Core Nebulas Rank is associated with the economy of a whole encrypted digital currency. As the economy changes, the algorithm of Core Nebulas Rank calculation will also need to be changed, especially its parameters. It's very important to figure how we can update the algorithm rapidly. Our solution is: upgrade the Nebulas Rank calculation algorithm through Nebulas Force.

More specifically, we will upgrade the block data structure, the new structure will include the Core Nebulas Rank algorithm and parameters (based on LLVM IR). Nebulas Virtual Machine (NVM) will be the execution engine of the algorithm: it fetches the algorithm code and parameters from the block, then execute the code, and eventually obtains the Core Nebulas Rank within the node.

Whenever the algorithm or the parameters need to be updated, we will work together with the community, making sure the new algorithm and parameters are in-

cluded in the new blocks, so the update will be timely and smooth, such that potential forks can be avoided down the road.

7 Extended Nebulas Rank

Core Nebulas Rank is used to evaluate an individual account's contribution to the economy aggregate, it's very important for Proof of Devotion (PoD) and Developer Incentive Plan (DIP), and the Core Nebulas Rank actually matches with their use cases. However, as we also noticed, there are some other use cases that they may need a different evaluation. Consequently, we also designed Extended Nebulas Rank. Extended Nebulas Rank is based on the Core Nebulas Rank, to guarantee continuous incentive to the whole Nebulas economy even under different use cases.

7.1 Smart Contract Oriented Extended Nebulas Rank

In the whole economy, the rank of smart contract plays an important role. On the one hand, it helps user find high quality DApps, on the other hand, it will also motivate the developers who build high quality Dapps, so the economy can grow healthily and stably.

The rank of smart contract depends on two truths: the call from users account address to smart contract, and the calls between different smart contracts. The call from user account address to smart contracts reflects the truth that the users account address is actually distributing its contribution to the aggregate economy of all smart contracts, since each smart contract has its own initial NR. The calls between smart contracts can also be treated as a directed acyclic graph. Therefore, we use Page Rank algorithm to calculate NR for each smart contract.

7.2 Multi-dimension Extended Nebulas Rank

We also found out that some applications need multi-dimensional data in order to compute the correlation between different kinds of data on chain. For example, in a blockchain based advertisement system, we need to get the correlation between the advertisement and the user from different dimensions. Under this situation, Extended Nebulas Rank is multi-dimensional, we can represent it as a vector, where the Core Nebulas Rank is one of the dimensions.

Extended Nebulas Rank is multi-dimensional, except the Core Nebulas Rank, the other dimensions all depend on concrete applications. How to implement those di-

mensions also depends on the application itself. Nevertheless, the calculation algorithms can always reference the calculations of the Core Nebulas Rank algorithm.

Starting from a real use case, we design the Extended Nebulas Rank for smart contracts, and have described an implementation method of Extended Nebulas Rank. We also illustrated the corresponding evaluation mechanism of this algorithm, and proposed the multi-dimensional Extended Nebulas Rank, which shows the possibility for our evaluation mechanism to be used in other use cases.

8 Future Work

The goal of Nebulas Rank is to provide a necessary value measurement for blockchain, from the perspective of giving an evaluation on the contribution from user account addresses to aggregate economy. There will be more work along the way. Here we briefly summarized the future work we will be working on:

- Cross-chain Nebulas Rank. We can foresee there will be high demand on cross chain data transfer in the near future. To name a few examples, cross-chain data interaction and digital asset transfer will certainly need value measurement on different chains. For example, when developers transfer their Dapps from one chain to the other chain, a method for how to calculate the Nebulas Rank of the Dapps will also need an unique value measurement among different chains.
- More contribution indictors based on economy aggregate. Nebulas Rank is based on the contribution to the economy aggregate. However, the development of the blockchain industry needs its community. Therefore, in terms of the economy aggregate, we cannot ignore the contribution of the community. So, how do we evaluate an individual's or organization's contribution in the community, and how this is reflected in Nebulas Rank, certainly has tremendous implications.

References

[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/`.

Appendix A Proof

A.1 Proof of Property 1

Proof. For any $x_1 > 0$, $x_2 > 0$, we have

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

In formula 21, we have $b < 0$, so $0 < e^{b\cdot x_1} < 1$, $0 < e^{b\cdot x_2} < 1$, moreover,

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

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

is actually:

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

□

A.2 Proof of Property 2

Proof. For any $x_1 > 0$, $x_2 > 0$, we have

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

Here we use function $g(x_1, x_2)$ represents the left part, $h(x_1, x_2)$ represents the right part:

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

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

So (23) for x_1 and x_2, their limits can be represented as:

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

we have

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

Calculate limit for $\frac{x}{1+\frac{1}{e^{a+b\cdot x}}}$, according to L'Hospital's rule,

$$\begin{aligned}\lim_{x\to\infty}\frac{x}{1+\frac{1}{e^{a+b\cdot x}}}&=\lim_{x\to\infty}\frac{1}{(e^{-a-b\cdot x})'}\\&=\lim_{x\to\infty}\frac{1}{-b\cdot e^{-a-b\cdot x}}\\&=0\end{aligned}$$

According to A.1, we have $g(x_1,x_2)>0$, so according to sandwich theorem:

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

Similarly, we can get:

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

So,

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

□

Appendix B Change Log

- 1.0 Release.