Extracting Tangle Properties in Continuous Time via Large-Scale Simulations

Bartosz Kuśmierz* Philip Staupe†

IOTA Foundation

January 21, 2018

Abstract

In this paper we analyze fundamental properties of the Tangle – a directed acyclic graph adapted for decentralized information storage. This technology has been adapted in the IOTA protocol – a scalable, permissionless distributed ledger designed for Internet of Things and Web 3.0. We use computer simulations to analyze cumulative weight evolution and tip count stability using a continuous-time model. As a byproduct of our analysis of average tip count we derive analytical formulas for the average number of tips in the discrete-time model. The paper also introduces and analyses the influence of a non-constant number of directly-approved transactions and scaling-invariance properties of the Tangle.

*Author's contact information: `bartosz.kusmierz@iota.org`

†Author's contact information: `your IOTA email?`

Contents

1 Introduction

1.1 Overview

This is the second paper in which we present preliminary results obtained with a computer simulation of the Tangle - directed acyclic graph (DAG) adapted for decentralized information storage. The Tangle was adapted as cornerstone technology of the IOTA network. IOTA is a scalable, permissionless distributed ledger designed for the Internet of Things industry. DAG based architecture of cryptocurrencies allows for bypassing drawbacks of standard blockchain architecture, like scalability (growing problem of micropayments) and division of users into two groups (issuers and approvers/miners of transactions).

The first paper [1] was devoted to the simulations of the discrete time model. Undeniable advantage of the discrete model is its simplicity. Time discrete model is both easy to implement and convenient for analytical calculations. Having said this, bulk of obtained results agrees with the predictions of the IOTA Whitepaper [2] in which the Tangle was introduced, what confirms usefulness of the discrete model. The paper [1] confirmed the existence of two distinct phases of cumulative weight growth and behavior of number of tips as a function of time. However, the simulation results of the discrete model exhibited clear discrepancies to the predictions of the Whitepaper regarding the average number of tips. An hypothesis was made, that disparities were caused by differences between discrete and continuous models examined in the simulation and the Whitepaper respectively. In this paper we explore simulations of the continuous model. Obtained results allow us to confirm this hypothesis. We discuss disparities of two models in the section **3**. As a continuation of our analysis we derive analytical formula for average number of the tips in the discrete model (see appendix **A**).

In this paper we verify the predictions of the Whitepaper with a continuous version of the simulations. Furthermore, we explore novel concept of non-constant number of directly approved transactions k. We perform this generalization by substituting the fixed quantity $k \in \mathbb{N}$ for some random variable $K : \Omega \to \mathbb{N}$. Although random variable K has to be integer-valued, its expected value does not. This opens possibility of interpreting certain forms of random variable K as fractional k (number of directly approved transactions). Subsections **3.3.1** and **3.3.2** provide support of this claim with comparison of average number of tips obtained with simulations and formulas predicted by the Whitepaper and analysis of the influence of standard deviation of K.

The paper is organized as follows: in the next section we provide an introduction to the Tangle, we discuss basic concepts needed to understand this DAG and briefly discuss tip selection mechanisms. In sections **2** and **3** we present analysis of cumulative weights and number of tips respectively. Section **4** discusses scaling transformations of the Tangle and provides Tangle visualizations. In the last section **5** we conclude our findings.

1.2 Methodology

The computer simulation is based on formalism and algorithms proposed in the Whitepaper [2] (although we introduce certain novel concepts). In particular we use the same terminology. In order to issue a transaction user must directly approve k other transactions where k is an integer and does not necessarily equal 2. If transaction A directly approves B it is denoted $A \rightsquigarrow B$. In terms of DAG transactions are vertices and relation $A \rightsquigarrow B$ means that there is an edge between A and B. One says A indirectly approves Z if there is a sequence of length at least three of transactions satisfying: $A \rightsquigarrow B \rightsquigarrow ... \rightsquigarrow Z$. In general, the weight of the transaction is proportional to the amount of work that the issuing node invested into it, however in our simulations all of the transactions have the same, constant weight (equal to 1). An important notion introduced in the Whitepeper is the cumulative weight of some transaction x which is denoted $\mathcal{H}_x$. Cumulative weight is the sum of the own weight of the transaction and own weights of all transactions that directly or indirectly approve this transaction.

When transactions are issued by a large number of roughly independent entities, the process of incoming transactions can be modeled by a Poisson point process. This claim is justified by both analytical considerations [2, 3] and experimental observations [4]. We denote the rate of this process by λ. For the sake of simplicity, we assume that this rate remains constant in time. Then elapsed time between two consecutive transactions is given by the exponential distribution $\mathbf{Exp}(\lambda)$. In the simulations we assume all of the computational work that is needed to issue a transaction is done within $h = 1$ units of time (as we demonstrate in the section **4.1** properties of the Tangle does not depend on both λ and h, but rather on theirs product $\lambda \cdot h$). Until this time passes other transactions do not detect incoming transaction, in particular transaction approved by incoming transactions can be recognized as a tip (unconfirmed transaction). Two tips selection mechanism are considered:

• URTS (uniform random tip selection) algorithm - tips are chosen from the list of available tips randomly (uniform distribution).

• MCMC (Markov Chain Monte Carlo) algorithm - random walk of particles towards the tips. Particles are released in the tangle, the first k particles at distinct tips determine which transactions are approve.

In the case of MCMC algorithm adjusted for selection of k tips, we use a random walk of $3k+4$ particles. Such number of particles was chosen arbitrarily, when $k = 2$ formula agrees with 10 particles proposed in the Whitepaper. Moreover, used formula provides enough particles to account for some of them reaching the same tip, while not keeping theirs number unnecessarily high (as a function of k). Too many particles would make random walk too computationally demanding. When required number of tips tips is reached random walk of others stops. In the case all of $3k+4$ particles end up in less than k distinct tips, one completes list of transactions to approve with transactions approved by already selected tips. Starting positions of particles are chosen randomly (uniform distribution) from transactions issued between 100λ and 200λ transactions ago (since $h = 1$, this corresponds to time 100λ - 200λ before). During simulations we noticed that any starting position placed further than 10λ - 20λ provide the same growth of number of tips and we conclude that such starting position does not influence directly this parameter. However, to make sure that enough randomness is provided we decided to increase it by one order of magnitude).

Moreover one accepts chosen transaction as particle starting position only if its cumulative weight is at least 5 (number 5 had been chosen arbitrarily). In the absence of this additional requirement (not discussed in the Whitepaper), one can observe oscillations in number of tips. This behavior occurs only for large values of α and is most probably an artifact of large number of transactions being tip, deep in the tangle. This phenomena will be discussed elsewhere [5].

When a particle is located at the transaction x (and x is not a tip) then the transition probability P_{xy} of particle moving from transaction x to y (y approves x directly, $y \rightsquigarrow x$) is analogous to the

Whitepaper[2]:

$$P_{xy} = \exp\big(-\alpha(\mathcal{H}_x - \mathcal{H}_y)\big)\left(\sum_{z:z\rightsquigarrow x} \exp\left(-\alpha(\mathcal{H}_x - \mathcal{H}_z)\right)\right)^{-1}. \tag{1}$$

Where one sums over all transactions z that directly approve x.

This form of transition probability function is analogical to the Gibbs distribution, where $\mathcal{H}_x$ plays role of energy of each state (transaction) and α is an inverse temperature of the system. Intuitively one can think that large α correspond to the "cold" system, then particles move along relativelly thin paths inside the tangle of transactions with highest weights. Small α coresponds to the "hot" stystem, in this case motion of particles is chaotic and very hard to predict. What is worth stressing, MCMC with $\alpha = 0$ is not exactly equivalent to the URST. Any random walk on the Tangle is dependent on the topology of the DAG and URST does not depend on the network structure whatsoever. As our simulations shows differences can be noticed in average number of tips, where result is higher for MCMC with $\alpha = 0$ (see fig. 4 and tab. 1).

Most of the results presented here had been gathered for $\lambda = 50$. However, we also examined different values of this parameter: $\lambda = 5, 50, 500$ and 5000. Qualitatively results agree for all of these values. Simulations were performed for Tangles with 10^6 up to 10^7 transactions. Such number of transactions is significantly larger than time needed for the system to achieve stability (equilibrium) for $\lambda = 50$.

2 Cumulative Weight

Results obtained with the simulations of the continuous time model of the Tangle (similarly as in the case of the discrete time model) confirm predictions of the Whitepaper regarding growth of cumulative weight of transactions. For both tip selection mechanisms: URTS and MCMC with small values of α (such as $0, 0.001$) we observe two phases of growth: "adaptation period" (exponential growth) and linear growth (see figs. 1 and 2). Behavior analogous to what is observed in figs. 1 and 2 is typical for almost all other transactions in this regime. The only exceptions are permanent tips. Permanent tips are transactions that are either approved by a negligible number of transactions or not approved at all. Example of cumulative weight growth (or lack of growth) of permanent tips can be observed in fig. 3. This figure represents typical cumulative weight growth when tip selection is guided by a relatively high $\alpha = 0.1$. In this case tip selection no longer bears similar characteristics to tip selection URTS. Transactions are either "lucky", get approved very quickly and fall into linear trend of growth, or are "unlucky" and stay permanent tips. Although the probability of becoming permanent tips is nonzero even for smaller α, we observe that the ratio of lucky to unlucky transactions is rather small for $\alpha \leq 0.001$. On the contrary, for $\alpha = 0.1$, out of 5 examined transactions only 1 went to linear growth phase. This characteristic, obviously, disqualifies MCMC with large α as a default tip selection mechanism in a real-life implementation of the Tangle. More in depth analysis of this feature will be presented in future paper [5].

[1] For a part (a) of the figure 1 $f(x)$ had been fitted on an interval $[29; 60]$ and $g(x)$ on an interval $[16,\ 28]$, $a_1 = 48.729 \pm 0.026$, $b_1 = -1249.3 \pm 1.2$; $c_1 = 0.3131 \pm 0.0017$, $d_1 = -3.787 \pm 0.045$. In a part (b) $f(x)$ had been fitted on an interval $[31,\ 60]$, $g(x)$ on an interval $[10,\ 30]$. Values of parameters: $a_2 = 49.549 \pm 0.030$, $b_2 = -1194.9 \pm 1.4$; $c_2 = 0.18287 \pm 0.00040$, $d_2 = 0.265 \pm 0.011$. One can notice that slopes of both linear functions are close to the value of $\lambda = 50$.

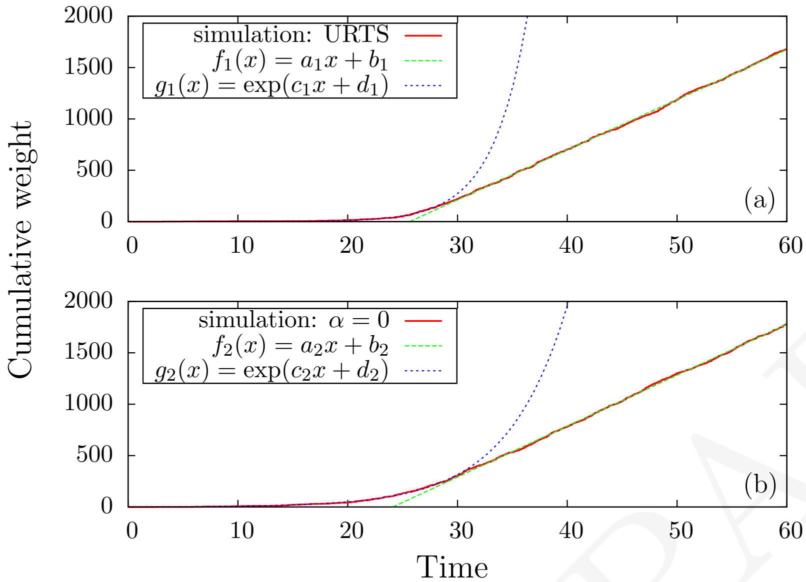

Figure 1: Cumulative weights of the 200th transactions issued during the simulations of the Tangle for: (a) - tip selection URTS; (b) - tip selection MCMC with $\alpha = 0.001$. Flow rate of new transactions: $\lambda = 50$, each transaction approves exactly $k = 2$ other transactions. Green and blue curves are best fitted linear and exponent functions respectively. Linear trend of cumulative weight continued for the length of the simulation.[1]

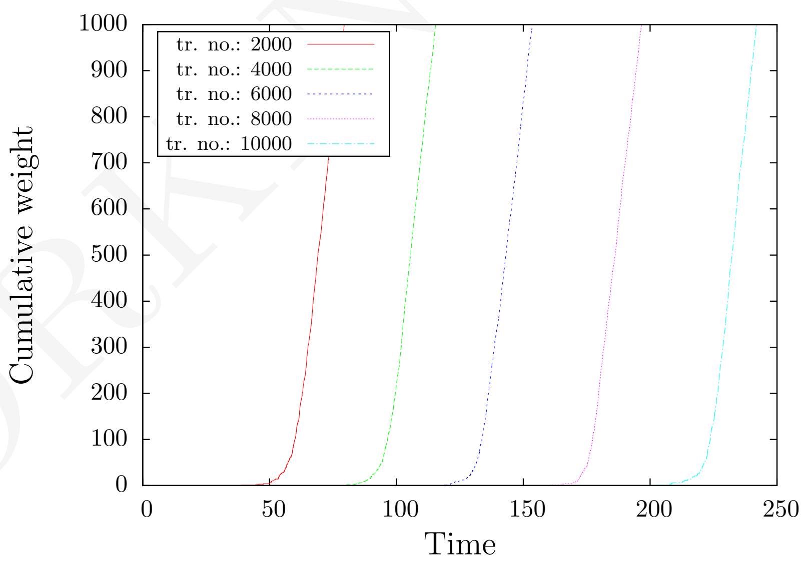

Figure 2: Cumulative weight of 2000th, 4000th, 6000th, 8000th and 10000th transaction issued during the simulation of the Tangle. Flow rate of new transactions: $\lambda = 50$; tip selection algorithm: MCMC random walk of 10 particles towards the tips for $\alpha = 0.001$; each transaction approves exactly $k = 2$ other transactions.

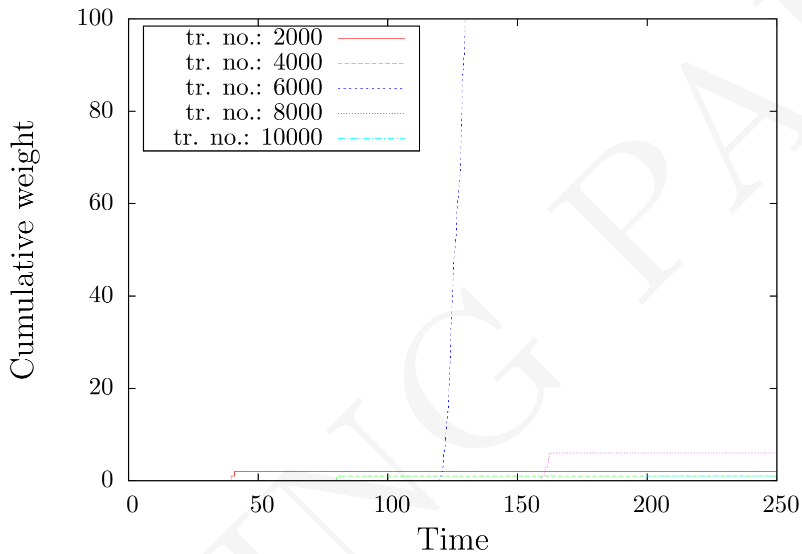

Figure 3: Cumulative weight of 2000th, 4000th, 6000th, 8000th and 10000th transaction issued during the simulation of the Tangle. Flow rate of new transactions: $\lambda = 50$; tip selection algorithm: MCMC random walk of 10 particles towards the tips for $\alpha = 0.1$; each transaction approves exactly $k = 2$ other transactions. From 5 examined curves only 1 experienced linear growth of cumulative weight. The four remaining curves correspond to the permanent tips: 2000th ans 8000th transaction were referenced by 2 and 6 transactions respectively. Other permanent tips were not referenced at all and acquire a cumulative weight of 1.

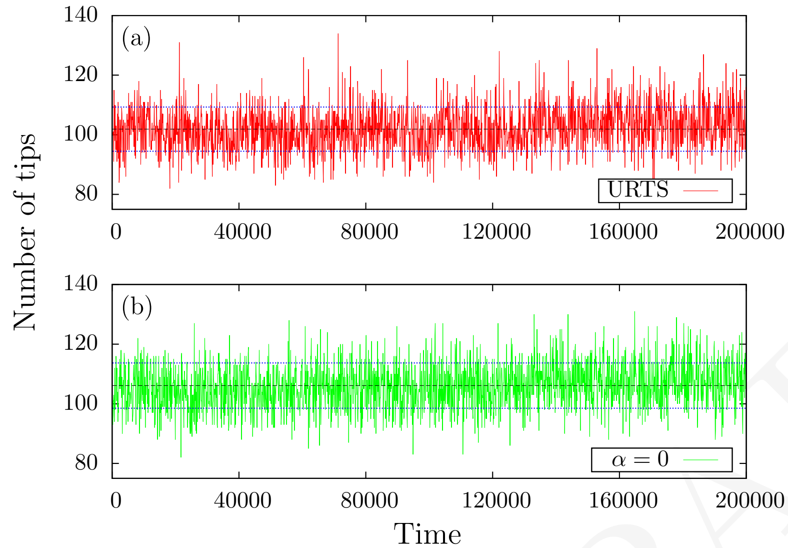

Figure 4: Values of $L(t)$ - number of the tips (unconfirmed transactions) as a function of time for: (a) URTS; (b) tip selection MCMC $\alpha = 0$. Flow rate of new transactions: $\lambda = 50$; each transaction approves exactly $k = 2$ other transactions; simulation involves 10^7 transactions. $L(t)$ varies greatly around average value (black long dashed curve) and most of the points are no more further from average than standard deviation (blue short dashed lines).

3 Tip Count

In this section we examine properties of quantity $L(t)$ - number of tips in the Tangle. We examine behavior of $L(t)$ as a function of time, its average value denoted L_0 and generate empirical probability density functions for $L(t)$. We also compare number of tips in continuous and discrete model. Reader may also look into the appendix, section **A**, where we derive an analytical formula for average number of tips as a function of approved transactions k.

3.1 Evolution in Time and Probability Density Functions

We analyze the number of tips $L(t)$ as a function of the time for different tip selection methods. Examined simulations involved Tangles 10^7 transactions in size. In the case of URTS, $L(t)$ remains stable (see fig. 4, part a). Although the number of tips varies greatly, it fluctuates around an average L_0 (see the Whitepaper [2] section **3 Stability of the system, and cutsets**). This fact can be illustrated with empirical probability density function of number of tips (i.e. normalized histogram of $L(t)$) - fig. 6.

In the case of MCMC, when α is small, $L(t)$ is expected to exhibit similar behavior as for URTS. Indeed, figs. 4 and 5 (part b) show that $L(t)$ seem to be stable, at least on examined time interval. In the last paper [1] we pointed out that when MCMC random walk starts from transactions located within certain bounded distance from the top of the Tangle, then $L(t)$ diverges. This is the situation in performed simulations (bound is 200λ). We also stress that the recent paper regarding equilibria in the Tangle [6] discusses an even stronger conjecture (**Conjecture 1.1.**). Thus unlike in the case of tips selection URTS it does not make sense to plot empirical density functions for $L(t)$ at an

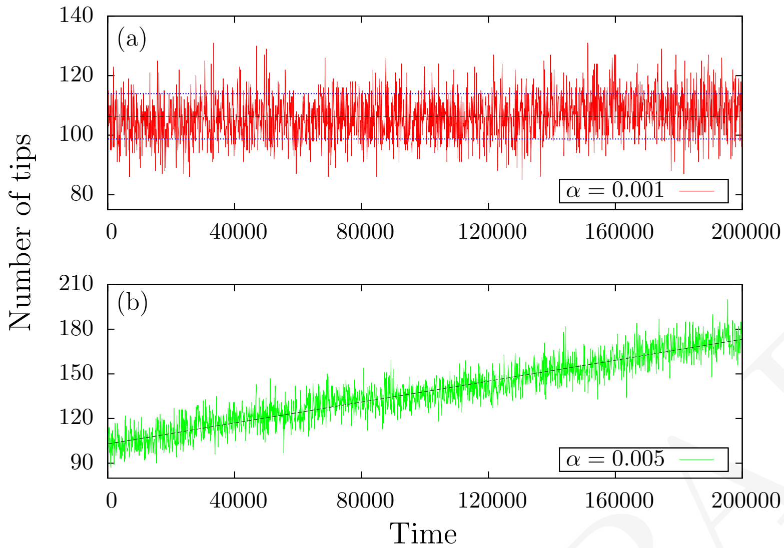

Figure 5: Values of $L(t)$ - number of the tips (unconfirmed transactions) as a function of time for: (a) - tip selection MCMC $\alpha = 0.001$; (b) - tip selection MCMC $\alpha = 0.005$. Flow rate of new transactions: $\lambda = 50$; each transaction approves exactly $k = 2$ other transactions; simulation involves 10^7 transactions. In part (a) $L(t)$ varies greatly around average value (black long dashed curve) and most of the points are not further from average than standard deviation (blue short dashed lines). Part (b) reveals linear trend indicated by best fitted linear function (black long dashed line).

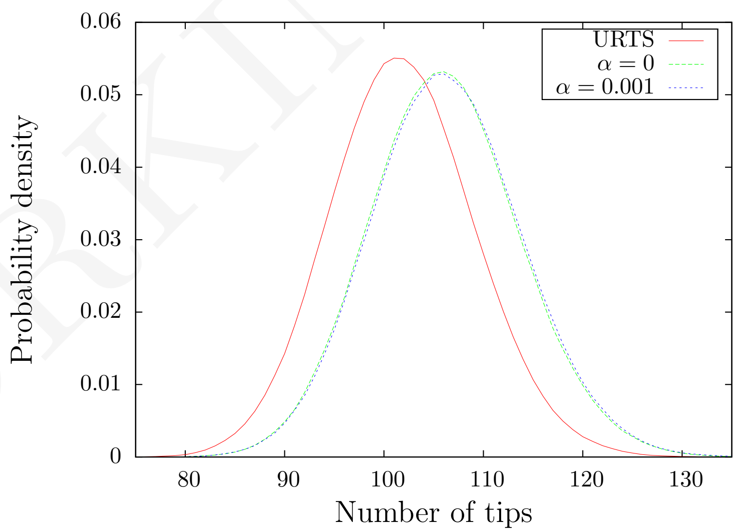

Figure 6: Empirical probability density function of number of tips $L(t)$ (normalized histogram). Tip selection algorithm: URTS (red curve), MCMC $\alpha = 0$ (green curve), $\alpha = 0.001$ (blue curve). Flow rate of new transactions: $\lambda = 50$, each transaction approves exactly $k = 2$ other transactions; simulations involve 10^7 transactions.

arbitrary time value. However, we can still plot empirical density functions for $L(t)$ in examined time interval. Such distributions are plotted in fig. 6.

When α is large, paths of particles would concentrate near fixed highways, determined by the highest cumulative weight. In this case MCMC tip selection mechanism no longer resembles tip selection URTS and $L(t)$ undergoes different evolution. As simulation suggests in this regime number of tips grows linearly and the slope can be significant. An example of such behavior is given in the figs. 5, part b for $\alpha = 0.005$. Although we do not present more plots in this paper, we also examined higher values of α and linear growth is even more significant. For example when $\alpha = 0.01$ best fitted linear curve to the number of tips reveal slope $(2.79237 \pm 0.009) \cdot 10^{-4}$. Although slope may seem to be small, number of tips after 10^7 transactions would be roughly 2900. Almost 3000 tips is substantially greater than value of tips for $\alpha = 0.005$ on analogous time period, which is below 200 (see fig. 5, part b).

We want to communicate that the slope of tips number is dependent on λ, especially when α is large. As we already pointed out when α is big system is “cold” and particles move along relativelly thin path inside the Tangle. In the limit of infinite α thickness of this path (understood as number of tips per given unit of time) is not dependent on λ, whereas number of all isued transactions is proportional to λ. Since only transactions inside thin path will be approved, slope of tips number depends on λ.

3.2 Dependency on Flow Rate of new Transactions λ

Our simulation shows that average tip count is a linear function of λ. For URTS method results agree with predictions of the Whitepaper $L_0 = 2\lambda$. For MCMC with $\alpha = 0$ we observe that slope is slightly higher (by 4-4.5%), see tab. 1. In the case of large values of λ time needed for the Tangle to stabilize can be more significant than for $\lambda = 50$. We noticed that for $\lambda = 5000$ Tangle stabilizes after approximately 250 000 transactions. Having this in mind we sample data starting from 500000th transaction to 10^6 every 1000 transaction.

We want to communicate that empirical probability density function of number of tips of the Tangle seems to be scalable, where its average value grows linearly with λ while its standard deviation grows as a square root. However further investigation is still needed [5].

λ	URTS	MCMC $\alpha = 0$
5	2.067 ± 0.021	2.128 ± 0.022
10	2.025 ± 0.014	2.115 ± 0.015
50	$2.0144 \pm 6.3 \cdot 10^{-3}$	$2.0946 \pm 6 \cdot 10^{-3}$
100	$2.0091 \pm 4.6 \cdot 10^{-3}$	$2.0914 \pm 4.6 \cdot 10^{-3}$
500	$2.0054 \pm 1.9 \cdot 10^{-3}$	$2.084 \pm 2.2 \cdot 10^{-3}$
1000	$2.0026 \pm 1.4 \cdot 10^{-3}$	$2.0863 \pm 1.2 \cdot 10^{-3}$
5000	XX missing	XX missing

Table 1: Average number of tips divided by λ for two tip selection methods: URTS and MCMC $\alpha = 0$ (second and third column respectively). The first column is flow rate of transactions. Number of directly-approved transactions $k = 2$. Uncertainties are standard deviation of average of 500 i.i.d. random variables.

3.3 Dependency on Number of Directly-Approved Transactions k

3.3.1 Average Value of k

Previously obtained results of the discrete time simulations [1] revealed some differences from the analytical predictions (see section **3 Stability of the system, and cutsets** of the [2]). In this paper we examine simulations of the continuous model and results largely agree with predictions of the Whitepaper. Table 2 displays the average number of tips for a series of Tangles. Each Tangle involved 10^6 transactions, number of directly approved transaction k was constant on the course of each simulation. For different simulations k varied from 2 to 10. In order to obtain data, we recorded exact number of tips every 1000th transactions (what gives 1000 numbers in total). Then for those data we calculated average value and standard deviation. Uncertainties present in the tables are standard deviation divided by $\sqrt{1000}$ (standard deviation of average of 1000 i.i.d. random variables). Results had been gathered for five methods of tips selection: URTS and MCMC random walk with $\alpha = 0, 0.001, 0.005, 0.01$. It is no suprise that results obtained with method URTS give results closes to the predictions of Whitepaper ($L_0 = k/(k-1)$), as this method had been used in the derivation. One observes that obtained results rarely differ from predictions by an amount larger than uncertainty. One also notices that results obtained for both MCMC random walks are fairly close to the predictions as well.

k	$^k/_{(k-1)}$	URTS	MCMC $\alpha = 0$	MCMC $\alpha = 0.001$	MCMC $\alpha = 0.005$	MCMC $\alpha = 0.01$
2	2	$1.9965 \pm 5 \cdot 10^{-3}$	$2.0822 \pm 5.1 \cdot 10^{-3}$	$2.0752 \pm 5.2 \cdot 10^{-3}$	$2.0945 \pm 5.2 \cdot 10^{-3}$	$2.1828 \pm 5.2 \cdot 10^{-3}$
3	$1\frac{1}{2} = 1.5$	$1.5053 \pm 4.6 \cdot 10^{-3}$	$1.4779 \pm 4.4 \cdot 10^{-3}$	$1.4807 \pm 4.6 \cdot 10^{-3}$	$1.4833 \pm 4.4 \cdot 10^{-3}$	$1.4918 \pm 4.5 \cdot 10^{-3}$
4	$1\frac{1}{3} = 1.3333$	$1.3349 \pm 4.6 \cdot 10^{-3}$	$1.2911 \pm 4.4 \cdot 10^{-3}$	$1.2902 \pm 4.6 \cdot 10^{-3}$	$1.2971 \pm 4.5 \cdot 10^{-3}$	$1.2906 \pm 4.4 \cdot 10^{-3}$
5	$1\frac{1}{4} = 1.25$	$1.2596 \pm 4.6 \cdot 10^{-3}$	$1.2059 \pm 4.5 \cdot 10^{-3}$	$1.2021 \pm 4.6 \cdot 10^{-3}$	$1.2046 \pm 4.5 \cdot 10^{-3}$	$1.2088 \pm 4.4 \cdot 10^{-3}$
6	$1\frac{1}{5} = 1.2$	$1.2042 \pm 4.5 \cdot 10^{-3}$	$1.1555 \pm 4.4 \cdot 10^{-3}$	$1.1534 \pm 4.6 \cdot 10^{-3}$	$1.1533 \pm 4.5 \cdot 10^{-3}$	$1.1587 \pm 4.5 \cdot 10^{-3}$
7	$1\frac{1}{6} = 1.1666$	$1.1666 \pm 4.6 \cdot 10^{-3}$	$1.1133 \pm 4.6 \cdot 10^{-3}$	$1.1149 \pm 4.4 \cdot 10^{-3}$	$1.1209 \pm 4.5 \cdot 10^{-3}$	$1.1247 \pm 5 \cdot 10^{-3}$
8	$1\frac{1}{7} = 1.1429$	$1.1441 \pm 5 \cdot 10^{-3}$	$1.0863 \pm 4.4 \cdot 10^{-3}$	$1.1032 \pm 5 \cdot 10^{-3}$	$1.1009 \pm 4.5 \cdot 10^{-3}$	$1.1078 \pm 4.6 \cdot 10^{-3}$
9	$1\frac{1}{8} = 1.125$	$1.1257 \pm 4.4 \cdot 10^{-3}$	$1.0847 \pm 4.4 \cdot 10^{-3}$	$1.0903 \pm 4.5 \cdot 10^{-3}$	$1.0851 \pm 4.6 \cdot 10^{-3}$	$1.0839 \pm 4.6 \cdot 10^{-3}$
10	$1\frac{1}{9} = 1.1111$	$1.1091 \pm 4.6 \cdot 10^{-3}$	$1.0702 \pm 5 \cdot 10^{-3}$	$1.0753 \pm 5 \cdot 10^{-3}$	$1.0728 \pm 4.6 \cdot 10^{-3}$	$1.0672 \pm 4.5 \cdot 10^{-3}$

Table 2: L_0/λ (average number of tips divided by flow rate of transactions λ) during simulations of the Tangle for different values of k. Consecutive columns are: number of directly approved transactions k; number $^k/_{(k-1)}$ which is analytical prediction of tips number for tips selection method URTS three last columns are results obtained with a simulation for tip selection methods: URTS, MCMC random walk with parameter $\alpha = 0, 0.001, 0.005, 0.01$ respectivelly. Flow rate of new transactions $\lambda = 50$. Each simualtion involved 10^6 transactions, data for average had been gathered every 1000th transaction. Uncertainties are standard deviation of average of 1000 i.i.d. random variables.

The Tangles analyzed in table 2 have constant k for all of the transactions. However we can explore simulations where number of directly approved transactions vary from transaction to transaction. We model this situation by associating certain random variable with number of directly approved transactions: $K : \Omega \to \mathbb{N}$. Where Ω is the set of all transactions, the value of K is always greater or equal 2 (with the exception of the genesis transaction). Although such random variable have to be integer-valued its expected value can be non-integer. In table 3 we examined average number of tips for number of directly approved transactions modeled with a special form of K -

two point distribution. Namely $\mathbb{P}(K = k) = \mathbb{P}(K = k + 1) = 1/2$. Such random variable have expectational value $\mathbb{E}K = k + {}^1\!/_2$. Obtained data shows that average number of tips seems to be following the formula for average number of tips derived in the Whitepaper: $k/(k - 1)$ (one notices that the derivation from the Whitepaper can be extended to non-integer values of k).

$\mathbb{E}K$	${}^{\mathbb{E}K}/_{(\mathbb{E}K-1)}$	URTS	MCMC $\alpha = 0$	MCMC $\alpha = 0.001$	MCMC $\alpha = 0.005$	MCMC $\alpha = 0.01$
2.5	$1\frac{2}{3} = 1.6667$	$1.6713 \pm 4.6 \cdot 10^{-3}$	$1.72 \pm 5 \cdot 10^{-3}$	$1.7129 \pm 4.6 \cdot 10^{-3}$	$1.7273 \pm 5 \cdot 10^{-3}$	$1.8292 \pm 5 \cdot 10^{-3}$
3.5	$1\frac{2}{5} = 1.4$	$1.4084 \pm 4.5 \cdot 10^{-3}$	$1.3664 \pm 4.5 \cdot 10^{-3}$	$1.3703 \pm 4.4 \cdot 10^{-3}$	$1.3650 \pm 4.3 \cdot 10^{-3}$	$1.3936 \pm 4.5 \cdot 10^{-3}$
4.5	$1\frac{2}{7} = 1.2857$	$1.2865 \pm 4.4 \cdot 10^{-3}$	$1.2477 \pm 4.6 \cdot 10^{-3}$	$1.2458 \pm 4.5 \cdot 10^{-3}$	$1.2418 \pm 4.4 \cdot 10^{-3}$	$1.2483 \pm 4.4 \cdot 10^{-3}$
5.5	$1\frac{2}{9} = 1.2222$	$1.2297 \pm 4.5 \cdot 10^{-3}$	$1.1774 \pm 4.5 \cdot 10^{-3}$	$1.1659 \pm 4.5 \cdot 10^{-3}$	$1.1771 \pm 4.4 \cdot 10^{-3}$	$1.1817 \pm 4.5 \cdot 10^{-3}$
6.5	$1\frac{2}{11} = 1.1818$	$1.1865 \pm 4.4 \cdot 10^{-3}$	$1.135 \pm 4.5 \cdot 10^{-3}$	$1.1319 \pm 4.5 \cdot 10^{-3}$	$1.1378 \pm 4.5 \cdot 10^{-3}$	$1.1348 \pm 4.6 \cdot 10^{-3}$
7.5	$1\frac{2}{13} = 1.1538$	$1.1548 \pm 4.5 \cdot 10^{-3}$	$1.1071 \pm 5 \cdot 10^{-3}$	$1.1065 \pm 4.4 \cdot 10^{-3}$	$1.1057 \pm 4.5 \cdot 10^{-3}$	$1.1073 \pm 5 \cdot 10^{-3}$
8.5	$1\frac{2}{15} = 1.1333$	$1.1414 \pm 4.6 \cdot 10^{-3}$	$1.0953 \pm 4.6 \cdot 10^{-3}$	$1.0947 \pm 4.6 \cdot 10^{-3}$	$1.0862 \pm 4.4 \cdot 10^{-3}$	$1.0912 \pm 4.6 \cdot 10^{-3}$
9.5	$1\frac{2}{17} = 1.1176$	$1.1156 \pm 4.5 \cdot 10^{-3}$	$1.0787 \pm 4.6 \cdot 10^{-3}$	$1.0955 \pm 4.5 \cdot 10^{-3}$	$1.0750 \pm 4.6 \cdot 10^{-3}$	$1.0852 \pm 5 \cdot 10^{-3}$

Table 3: L_0/λ (average number of tips divided by flow rate of transactions λ) during simulations of the Tangle for different expectational value of number of directly approved tips $\mathbb{E}K$. Random variable K is defined by $\mathbb{P}(K = k) = \mathbb{P}(K = k + 1) = 1/2$. Consecutive columns are: number of directly approved transactions k; number ${}^{\mathbb{E}K}/_{(\mathbb{E}K-1)}$ which is analytical prediction of tips number for tips selection method URTS; three last columns are results obtained with a simulation for tip selection methods: URTS and MCMC random walk with parameter $\alpha = 0$ and $\alpha = 0.001$ respectivelly. Flow rate of new transactions $\lambda = 50$. Each simualtion involved 10^6 transactions, data for average had been gathered every 1000th transaction. Uncertainties are standard deviation of average of 1000 i.i.d. random variables.

3.3.2 Standard Deviation of Random Variable K

Now we analyze influence of standard deviation of random variable K. Object of our interest is random variables $K_{k,\sigma}$ defined by two two point distribution

$$\mathbb{P}(K_{k,\sigma} = k - \sigma) = \mathbb{P}(K_{k,\sigma} = k + \sigma) = \frac{1}{2}. \tag{2}$$

Such random variable have expectational value k and standard deviation σ. In the table 4 one presents average number of the tips as a function of standard deviation σ. In the case of method URTS one notices very little influence of the standard deviation. As σ is getting bigger, L_0 reveals slight growing tendencies, however differences are within uncertainties. Moreover, empirical probability density functions of $L(t)$ presented in the figure 7 also is stable. Probability density functions for larger values of σ tend to be only very slightly skewed towards higher values but overall agreement between three curves is evident.

We want to stress that number of direct approved tips can influence internal structure of the Tangle (as one can observe in the table 2). Thus tip selection based on random walks on DAG should be expected to reveal bigger dependence on standard deviation of K. This is indeed the case as analysis of the table 4 reveals. One observes that for all of four MCMC tip selection mechanisms ($\alpha = 0, 0.001, 0.005, 0.01$) influence of standard deviation is visible only for large values of σ. Also one can observe that this influence is getting bigger with value of α. This observation is confirmed by study of empirical distribution functions - fig. 8.

We also examined analogous quantities for other random variables ($k = 4, 8$) and results were qualitatively the same: tip selection URTS seem to be unaffected by standard deviation of random variable K while random walk based tip selection reveals greater dependence and average number of tips grows with σ. However we want to stress that dependence is subtle when standard deviation is small.

σ	URTS	$\alpha = 0$	$\alpha = 0.001$	$\alpha = 0.005$	$\alpha = 0.01$
0	$1.1947 \pm 4.5 \cdot 10^{-3}$	$1.1556 \pm 4.6 \cdot 10^{-3}$	$1.1583 \pm 4.6 \cdot 10^{-3}$	$1.1533 \pm 4.5 \cdot 10^{-3}$	$1.1587 \pm 4.5 \cdot 10^{-3}$
1	$1.1995 \pm 4.7 \cdot 10^{-3}$	$1.1595 \pm 4.4 \cdot 10^{-3}$	$1.1600 \pm 5 \cdot 10^{-3}$	$1.1530 \pm 4.6 \cdot 10^{-3}$	$1.1602 \pm 4.4 \cdot 10^{-3}$
2	$1.1929 \pm 4.6 \cdot 10^{-3}$	$1.1818 \pm 4.6 \cdot 10^{-3}$	$1.1757 \pm 4.6 \cdot 10^{-3}$	$1.1769 \pm 4.4 \cdot 10^{-3}$	$1.1958 \pm 4.6 \cdot 10^{-3}$
3	$1.1986 \pm 5 \cdot 10^{-3}$	$1.2250 \pm 4.5 \cdot 10^{-3}$	$1.2182 \pm 4.6 \cdot 10^{-3}$	$1.2204 \pm 4.3 \cdot 10^{-3}$	$1.2541 \pm 4.6 \cdot 10^{-3}$
4	$1.2017 \pm 5 \cdot 10^{-3}$	$1.3158 \pm 4.6 \cdot 10^{-3}$	$1.3235 \pm 5 \cdot 10^{-3}$	$1.3582 \pm 5 \cdot 10^{-3}$	$1.7176 \pm 6 \cdot 10^{-3}$

Table 4: Average number of tips for simulations for different forms of random variable K (number of directly approved transaction). Random variable K is defined by $\mathbb{P}(K_{k,\sigma} = k - \sigma) = \mathbb{P}(K_{k,\sigma} = k + \sigma) = \frac{1}{2}$ for $k = 6$. First column is σ (standard deviation of random variable) and four next columns are results obtained with a simulation for tip selection methods: URTS, MCMC random walk with parameters $\alpha = 0, 0.001, 0.005, 0.01$ respectivelly. Flow rate of new transactions $\lambda = 50$. Each simualtion involved 10^6 transactions, data for average had been gathered every 1000th transaction. Uncertainties are standard deviation of average of 1000 i.i.d. random variables.

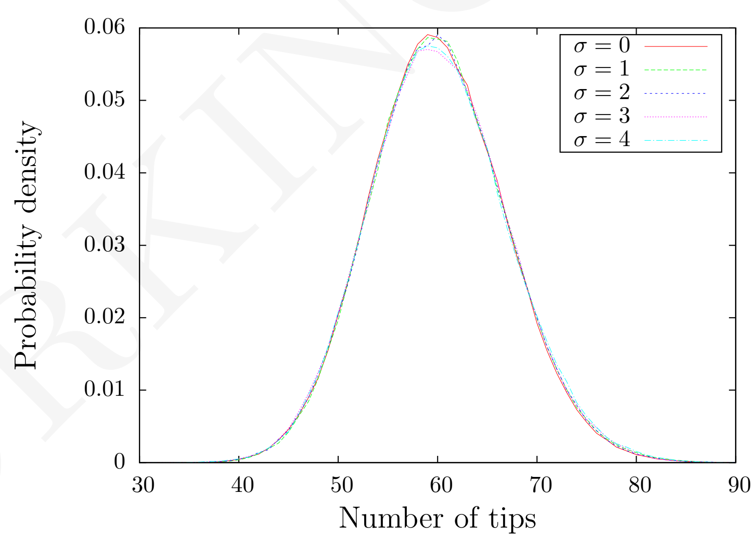

Figure 7: Empirical probability density function of number of tips $L(t)$ (normalized histogram). Flow rate of new transactions: $\lambda = 50$; tip selection algorithm: URTS; simulation involves 10^6 transactions. Every transaction approves number of transactions modeled with random variable $K_{k,\sigma}$, where $\mathbb{P}(K_{k,\sigma} = k - \sigma) = \mathbb{P}(K_{k,\sigma} = k + \sigma) = \frac{1}{2}$. We present data for $k = 6$ and $\sigma = 0, 1, 2, 3, 4$. One observes that all of the curves have very similar shape, both position of maximal value and overall spreadeness are virtually identical.

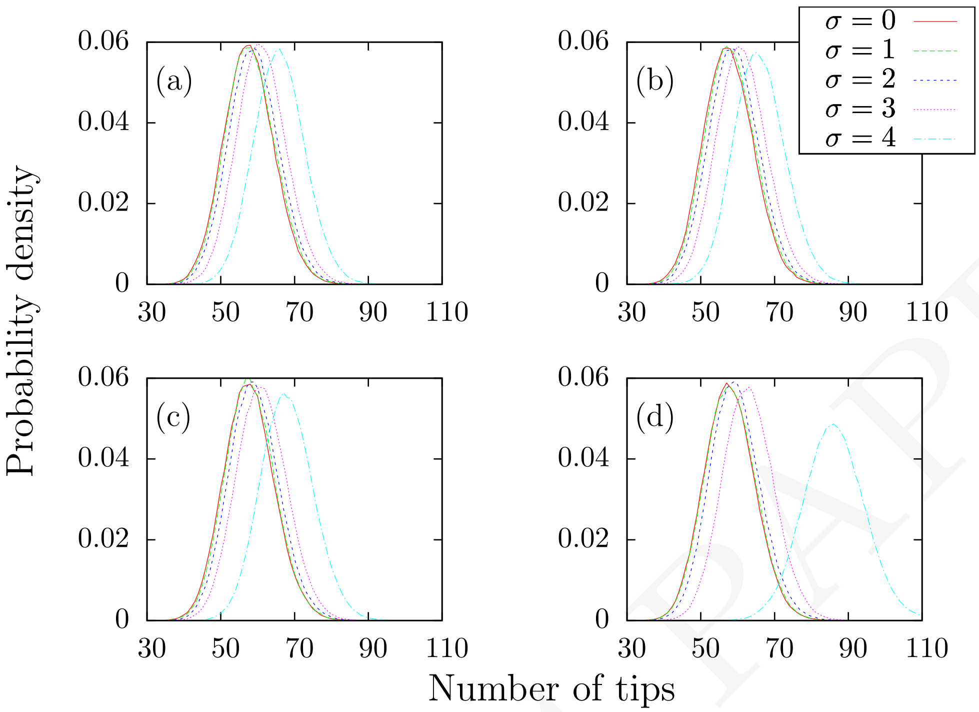

Figure 8: Empirical probability density function of number of tips $L(t)$ (normalized histogram). Flow rate of new transactions: $\lambda = 50$; tip selection algorithm: MCMC random walk with parameter $\alpha = 0$ (part (a)), $\alpha = 0.001$ (part (b)), $\alpha = 0.005$ (part (c)), $\alpha = 0.01$ (part (d)); simulation involves 10^6 transactions. Every transaction approves number of transactions modeled with random variable $K_{k,\sigma}$, where $\mathbb{P}(K_{k,\sigma} = k - \sigma) = \mathbb{P}(K_{k,\sigma} = k + \sigma) = \frac{1}{2}$. We present data for $k = 6$ and $\sigma = 0, 1, 2, 3, 4$.

4 Scaling of the Tangle and Visualizations

4.1 Time-Invariant Scaling

The Tangle as a dynamic system is invariant under the following transformation: $\forall \mathbf{a} \in \mathbb{R}$

$$h \to \mathbf{a} \cdot h, \quad \lambda \to \lambda/\mathbf{a}. \tag{3}$$

This transformation is equivalent to change of units of time by a factor $\mathbf{a}$. Equivalently one could say that the Tangle does not depend on λ and h but rather on the product $\lambda \cdot h$. This fact justifies our decision of fixing $h = 1$ in the performed simulations (one could say that we measure λ in units of h – flow rate of transactions per one h).

One can illustrate this with a simple example: Say $\lambda = 1 \ \left[\frac{\text{tx}}{\text{sec.}}\right]$ and $h = 1$ [sec.], i.e. there is 1 transactions issued per second. This statement is equivalent to saying that there is 60 transactions issued per minute:

$$\lambda = 1 \left[\frac{\text{tx}}{\text{sec.}}\right] \Leftrightarrow \lambda = 60 \left[\frac{\text{tx}}{\text{min.}}\right].$$

This fact can be illustrated with plots of the Tangle graphs for values of parameters fulfilling transformation given in equation 3 (see fig. 9 in the next subsection).

4.2 Tangle Visualizations

In the following we include graphical illustrations of the Tangle. Figure 9 demonstrates the point of Section 4.1, whereas Figure 10 and 11 highlight different choices for λ and α respectively. Approved transactions are green, tips are red.

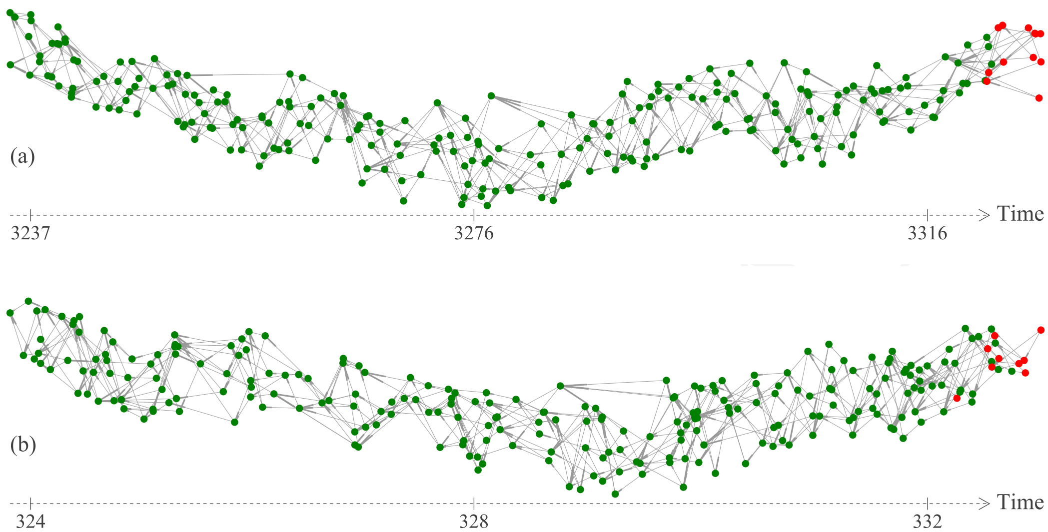

Figure 9: Graphical representation of the Tangle for (a) $\lambda = 3$, $h = 1$ over 100 units of time and (b) $\lambda = 30$, $h = 0.1$ over 10 units of time, both using MCMC with $\alpha = 0$.

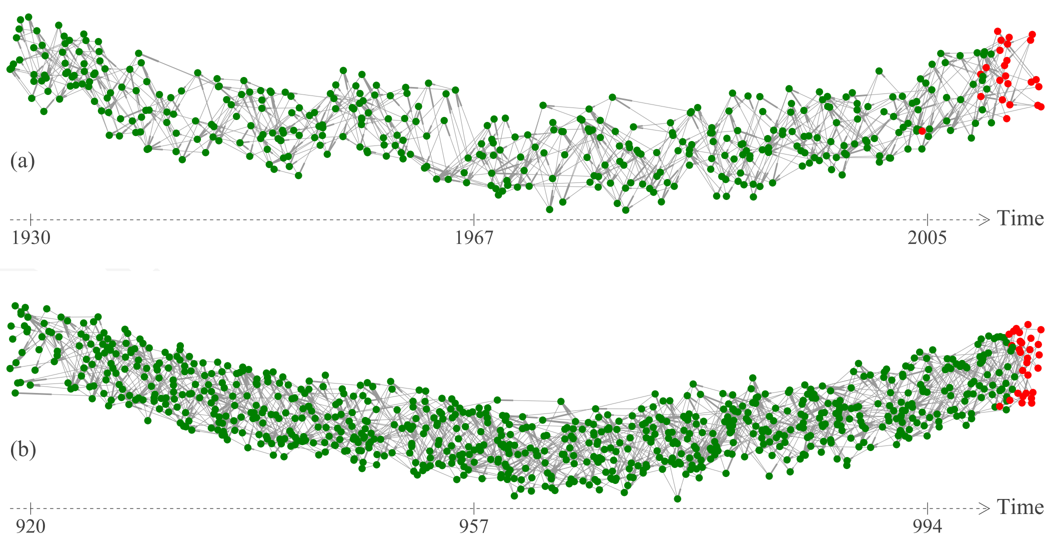

Figure 10: Graphical representation of the Tangle for (a) $\lambda = 5$, $h = 1$ and (b) $\lambda = 10$, $h = 1$ over 100 units of time, both using MCMC with $\alpha = 0$.

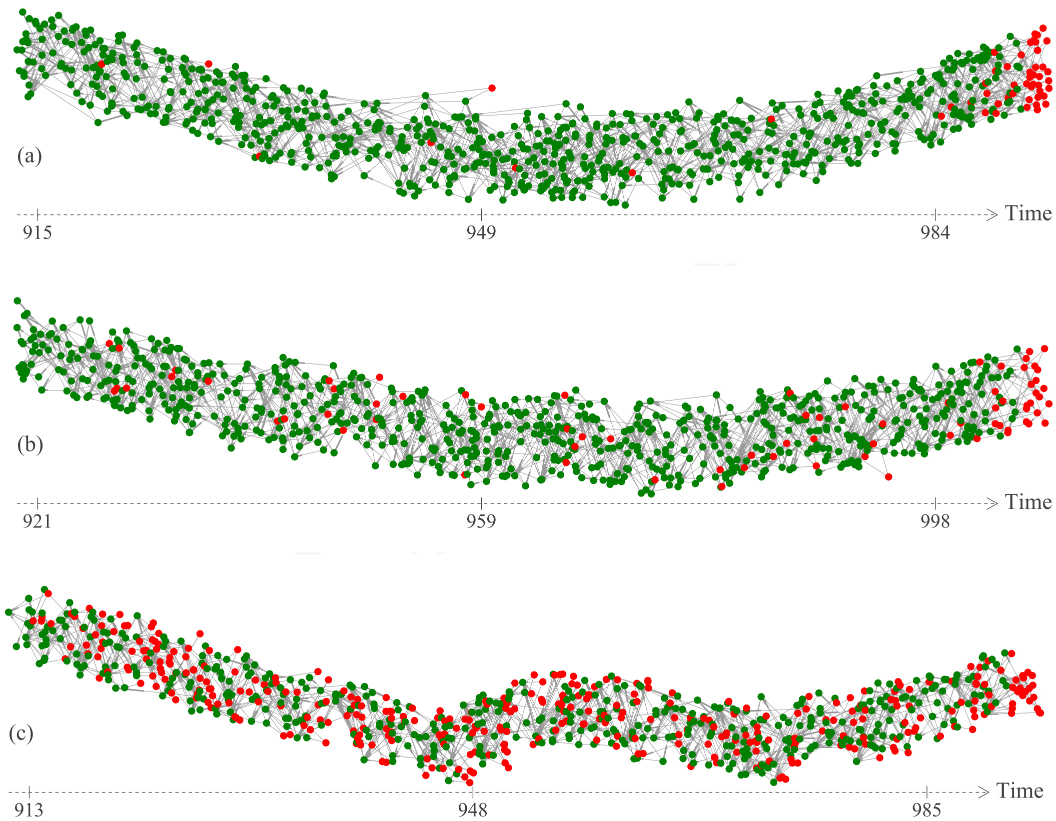

Figure 11: Graphical representation of the Tangle for $\lambda = 10$, $h = 1$ over 100 units of time, all using MCMC with (a) $\alpha = 0.05$, (b) $\alpha = 0.1$ and (c) $\alpha = 0.5$

5 Discussion and Conclusions

The results from the continuous-time model presented here largely confirm the analytical predictions of the original Whitepaper [2]. In particular, we demonstrated that the growth of the cumulative weight follows two stages: adoption phase (which appears to be exponential), followed by a linear phase (with slope λ). These findings were independent of the tip selection algorithm; be it URTS or MCMC with some small α. However, it is worth noting that for larger α we observed a non negligible probability of transactions being left behind (i.e. their cumulative weight growth stopped all together). This became particularly evident for larger choices such as $\alpha = 0.1$ (see fig. 3).

Furthermore, for URTS and differing choices of k, we found the total number of tips $L(t)$ as a function of time to be a mean-reverting process, always evolving around the analytical prediction

$$L_0 = \frac{k}{k-1} \cdot \lambda \cdot h \tag{4}$$

MCMC exhibited very similar mean-reverting behavior, but especially for larger α we additionally observed a linear upwards drift in the overall tip count over time. What is worth stressing even for $\alpha = 0$ average number of tips was greater than value predicted by the formula 4 (even though linear growth of number of tips was negligible at examined time intervals). This suggests that internal structure of the Tangle (important during random walks) is not uniform and paths leading to different tips are not equally accessible.

Lastly, we analyzed the average total tip count for when the number of selected tips followed a random variable K, and observed that the analytical prediction from (4) with $k = \mathbb{E}[K]$ generally matched the results produced by URTS very accurately, even when standard deviation of K is significant. On the other hand, analogical results obtained with MCMC started to deviate as we increased α and/or the standard deviation σ. Nevertheless for small values of α and σ results suggest that one can estimate number of tips of random variable K with just its expected value. This provides tools for discussion of non-integer number of directly approved transactions. We will explore this property and possible influence of K on other quantities in greater detail (including skewness of K) in the future paper [5]. What is more we plan on investigating other aspects of the Tangle. Topics of investigation will be the probability of transactions being left behind (becoming permanent tip), analyzing different particle starting positions (e.g. from milestones) and non-constant α (e.g. random or dependent on tangle meta data). We want to examine Tangle as a thermodynamical object and examine its properties with concepts knows from statistical physics, like: entropy, correlation functions, path branching of particles and others [5].

A Appendix: Average number of tips in discrete model

Observed differences in average number of tips between predictions of the Whitepaper [2] and the results of discrete time simulations [1] lead us towards further examination of this phenomena. For $k = 2$ analytical calculations estimated $L_0 = 2\lambda$, while obtained simulations place it around 1.2607 ± 0.003. Although both results are of the same magnitude, difference of 60% requires further investigation. In order to explain this discrepancy we estimated average number of tips in the discrete time model. Our derivation is based on tip selection algorithm URTS (uniform random tip selection), where new transactions randomly choose k tips from the list of available tips. As calculations below show, we were able to re-create results obtained in the previous paper within satisfying accuracy.

Fundamental unit of time in the discrete model is a *time step*. During each time step there is on average λ transactions issued, exact number is chosen from Poisson distribution $\mathbf{Pois}(\lambda)$. Each new transaction approves k older transactions. One assumes all of the calculations required for device to issue a transaction are done within one time step.

Assume the Tangle is at equilibrium and a new time step just begins. There is on average L_0 tips and λ new transactions. Each of them is going to refer k transactions, so there are $k\lambda$ tips to be chosen. To simplify, one ignores the fact that each new transaction have to chose k different tips. As such we examine situation that there are $k\lambda$ tips to be chosen independently. We are going to choose tips sequentially, one by one.

At the beginning, since no tip had been chosen yet, the first selected tip will not be referred by anyone. After this, average number of tips referred by some transaction equals $t_1 = 1$ (in this case it is exactly 1). What follows, when one chooses second tip, probability of selecting tip that is already referred by any transaction equals $1/L_0$. When the second tip is chosen, there are on average $t_2 = 2 - \frac{1}{L_0}$ tips referred by any transaction and probability of finding already referred transaction equals $2/L_0 - \frac{1}{L_0^2}$.

One can follow this logic for subsequent transactions. Then one obtains recursion formula for the expected number of tips referred by any transaction after choosing n tips - t_n:

$$t_n = t_{n-1} \cdot (1 - 1/L_0) + 1; \quad t_1 = 1.$$

This recursion is fairly easy to solve, after expanding few first few elements one finds that general solution satisfy $t_n = 1 + (1 - {}^1/_{L_0}) + (1 - {}^1/_{L_0})^2 + ... + (1 - {}^1/_{L_0})^n$. When one chooses $k\lambda$ transactions expected value of tips that had been approved at least once ("erased tips") equals

$$t_{k\lambda} = \frac{1 - (1 - {}^1/_{L_0})^{k\lambda+1}}{1 - (1 - {}^1/_{L_0})}. \tag{A.1}$$

where we used the fact that

$$\sum_{i=1}^{n} x^i = 1 + x + x^2 + \ldots + x^n = \frac{1 - x^{n+1}}{1 - x} \tag{A.2}$$

By stationarity, expected value of erased tips equals λ

$$\frac{1 - (1 - {}^1/_{L_0})^{k\lambda+1}}{1 - (1 - {}^1/_{L_0})} = 1 - (1 - {}^1/_{L_0})^{k\lambda+1} L_0 = \lambda.$$

When $\lambda \to \infty$, L_0 diverges as well. Then one can approximate expression in the brackets with value of e

$$1 - \frac{\lambda}{L_0} = \exp\left(-\frac{k\lambda + 1}{L_0}\right). \tag{A.3}$$

What is more, one can once more use limit of large λ and write $k\lambda + 1 \approx k\lambda$. Then one search for solutions of equation

$$1 - \frac{\lambda}{L_0} = \exp\left(-\frac{k\lambda}{L_0}\right). \tag{A.4}$$

Exact solutions are hard to obtain without further simplifications. For now we reach for approximate numerical solutions. For $k = 2$:

$$L_0 = 1.25628 \cdot \lambda.$$

This agrees with results of the discrete model simulation $L_0 = 1.2607 \cdot \lambda$ (where $\lambda = 50$, for different values of λ simulation give result within few % range). In the tab. 5 we present numerical solutions of the equation (A.4) for general k. Analysis of tab. 5 reveals that $\lambda/L_0 \to 1$ as $k \to \infty$. This makes us consider following substitution $1 - x = \lambda/L_0$ and try to solve equation for "small" x. One writes

$$x = \exp(-k)\exp(kx), \tag{A.5}$$

and for $\exp(kx) = 1 + kx + O(x^2)$, one leaves only linear term. Then

$$x = \exp(-k)(1 + kx),$$

$$x = \frac{\exp(-k)}{1 - k\exp(-k)} \tag{A.6}$$

and

$$\frac{L_0}{\lambda} = \frac{1}{1 - \frac{\exp(-k)}{1 - k\exp(-k)}} = \frac{1 - k\exp(-k)}{1 - k\exp(-k) - \exp(-k)}. \tag{A.7}$$

k	L_0/λ via Equation (A.4)	L_0/λ via Equation (A.7)
2	1.2562	1.2278
3	1.0633	1.0621
4	1.0202	1.0201
5	1.0070	1.0070
6	1.0025	1.0025

Table 5: Numerical solutions L_0/λ via equations (A.4) and (A.7) respectively.

Comparison of analytical formulas for average number of tips in discrete and continuous model as a function of k is presented in a fig. 12 (we want to stress that derivations of both formulas work for fractional values of k as well). One observes that average number of tips in the continuous model is bigger than in discrete for all values of k, also formula for discrete model reveal exponential decay contrary to following power law formula for continuous model.

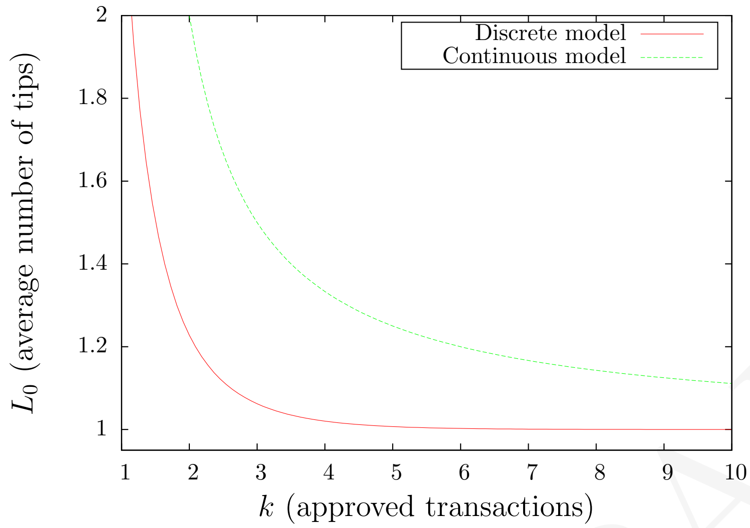

Figure 12: Plots of analytically derived formulas for average number of tips in "discrete model" (red curve) and "continuous model" (green curve, see [2]) as a function of directly approved transactions k.

References

[1] B. Kuśmierz, "The first glance at the simulation of the Tangle: discrete model", http://iota.org/simulation_tangle-preview.pdf (2017)

[2] S. Popov, "The Tangle", https://iota.org/IOTA_Whitepaper.pdf (2016)

[3] Sheldon M. Ross, "Introduction to Probability Models" 10th ed. (2012)

[4] P. Staupe, "IOTA Maths And Simulations", https://github.com/phil-zxx/papers/blob/master/IOTA_Maths_And_Simulations.pdf (2017)

[5] B. Kuśmierz, P. Staupe and S. Popov, "Simulation of the tangle", in preparation.

[6] S. Popov, O. Saa, P. Finardi, "Equilibria in the Tangle", https://arxiv.org/pdf/1712.05385.pdf (2017)