Quasi-Analytic Parasite Chain Absorption Probabilities in the Tangle

Philip Staupe

Abstract

We analyse with which probability the MCMC random walk gets absorbed into a parasite chain and underpin our arguments with simulation-based numerical results.

Last Updated: 31 Dec 2017

Table of Contents

1 Introduction

1.1 Disclaimer

The following results are **preliminary** and **should by no means be taken at face value**, as this is only the very first draft. All results should be understood as laying the groundwork for more comprehensive research on this topic.

1.2 Parasite Chains

The goal behind a *Parasite Chain* attack on the Tangle is to perform a double-spend. Specifically:

1. At time T_0 the attacker starts building a hidden chain of transactions – the *Parasite Chain* – parallel to the main tangle.
2. At time T_1 he places a specific transaction $\mathcal{T}_1$ (in the main tangle) whose funds he intends to spend again at some later point in time. In the meantime he continues building his chain in secret.
3. At time T_2 he adds a specific transaction $\mathcal{T}_2$ (to his chain) that stands in conflict to $\mathcal{T}_1$ (as it attempts to double-spend the funds used for $\mathcal{T}_1$). Concurrently he reveals his entire chain to the network.

The assumption is that between T_1 and T_2 some goods were "delivered" (whatever this means) paid for by transaction $\mathcal{T}_1$. And thus the attacker's intend is try to orphan $\mathcal{T}_1$ in order be able to spend the same funds a second time as well as keep the goods. By having built this parallel chain the hope is that the overall tangle consensus will branch off to his chain, consequently orphaning $\mathcal{T}_1$ and thus confirming the double-spend $\mathcal{T}_2$. Figure 1 demonstrates the Tangle structure of this endeavour.

1.3 Notations

Let us quickly re-cap some notation for modelling the Tangle in general as well as parasite chains.

λ	Tx issue-rate of the (honest) main tangle during one unit of time
h	Units of time it takes for one (honest) tx to go from hidden to revealed (coming from network latency & solving PoW)
μ	Tx issue-rate of the (malicious) parasite chain during one unit of time
T_0	Time at which the parasite chain begins
T_1	Time at which the first tx of the double-spend is issued
T_2	Time at which the second tx of the double-spend is issued

In addition to this, we recall the vertex transition probabilities from [PSF17] which are defined as

$$P_{xy}^{(f)} = \begin{cases} \dfrac{q}{|\mathcal{A}(x)|} & \text{if } y \in \mathcal{A}(x), \\ (1-q) \cdot \dfrac{f(\mathcal{H}_x^{(t-h)} - \mathcal{H}_y^{(t-h)})}{\sum_{z:\ x \in \mathcal{A}(z)} f(\mathcal{H}_x^{(t-h)} - \mathcal{H}_z^{(t-h)})} & \text{if } x \in \mathcal{A}(y), \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a monotone non-increasing function (such as $f(s) = \exp(-\alpha s)$ with $\alpha \geq 0$ being the inverse temperature), $\mathcal{A}(x)$ is the set of vertices which are approved/referenced by vertex x, and q is the probability of backtracking from vertex x to a previously-approved vertex in $\mathcal{A}(x)$, and $\mathcal{H}_x^{(t)}$ is the cumulative weight of vertex x at time t. Note that $|\mathcal{A}(x)| = 2$, in the case of IOTA.

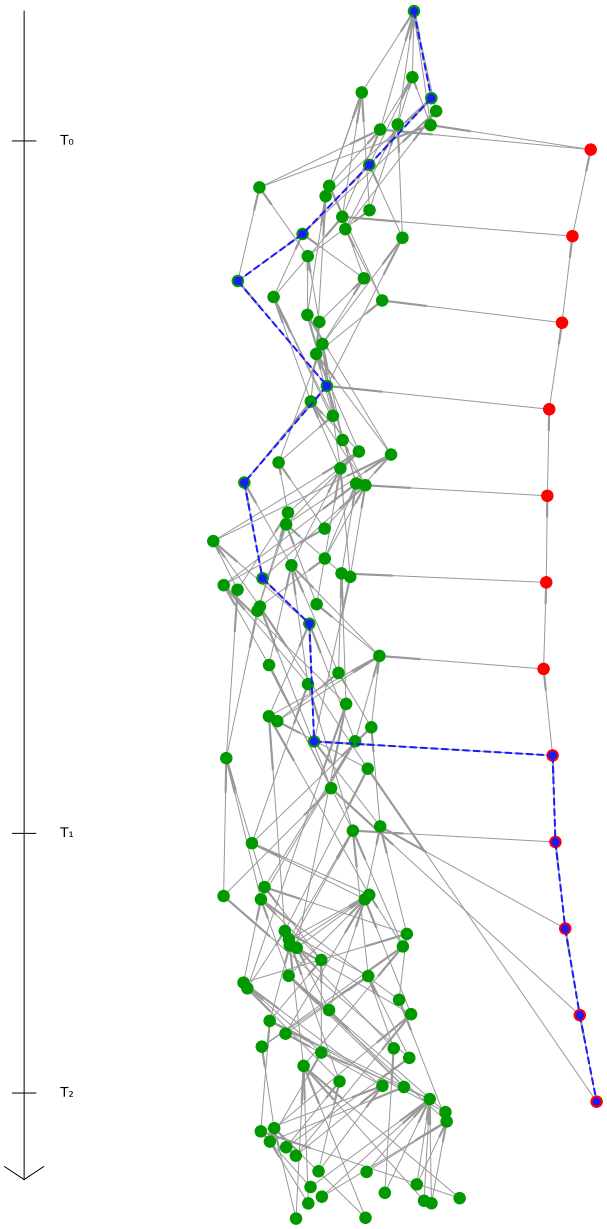

Figure 1: Tangle (in green) alongside a Parasite chain (in red) together with a sample random walk (in blue) which gets absorbed into the parasite chain.

2 Two-Way Markov-Chain Model

Trying to parameterise the Tangle as a full Markov-Chain is difficult, because the Tangle structure changes with every time step, thus making the Markov-Chain itself a random construct. We therefore attempt to break down the Tangle & parasite chain dynamics into their core elements.

We need a few observations for this to work:

1. The average time distance between two directly-connected vertices/transactions is $2h$.
   *In fact, the average time between a tx and its **first** approver is $2h$, so the average time distance to **any** of its approvers is slightly higher, but we shall ignore this inaccuracy.*
2. Every transaction approves **exactly** 2 transactions.
3. Every transaction is **on average** approved by 2 transactions.
4. For a time interval $[s, s+\delta]$ we expect $\lambda \cdot \delta$ transactions in the main tangle and $\mu \cdot \delta$ in the chain.
   So one can think of λ, μ as the breadth of the main tangle and parasite chain respectively whilst the time parameter serves as height.
5. After some adoption period, cumulative weights of transactions in the main tangle grow linearly with rate λ (before the parasite chain is revealed) and rate $\lambda+\mu$ (after the parasite chain is revealed). So for example, $\mathcal{H}_x^{(t)} \approx (\lambda+\mu) \cdot (t - t_x) + c$ for some constant $c \in \mathbb{R}$ and t_x the time at which transaction x got revealed to the network.

The idea of the *Two-Way Markov-Chain Model* is that we model the main Tangle and Parasite Chain as **two individual Markov-Chains**, only going **backwards** and **forward** in time. We additionally define specific transition probabilities between the two Markov-Chains. See Figure 2 for an example.

The main Tangle Markov-Chain jumps in time steps of $2h$ (by virtue of point 1). The Parasite Chain, on the other hand, jumps in time steps of μ^{-1} because issue-rate μ implies exponentially-distributed inter-arrival times with mean μ^{-1}. So when modelling the Tangle from T_0 to T_1 to T_2 we first calculate the number of Markov-Chain states between T_i and T_j in the main tangle and parasite chain respectively

$$\begin{aligned} n_{\text{tangle},ij} &= \lceil 1 + (T_j - T_i) \,/\, (2h) \rceil \\ n_{\text{parasite},ij} &= \lceil 1 + (T_j - T_i) \cdot \mu \rceil \end{aligned}$$

where $\lceil \varepsilon \rceil$ is the ceiling of $\varepsilon \in \mathbb{R}$, and then define the sets of Markov-Chain states

$$\begin{aligned} \mathcal{T} &= \{ i \in \mathbb{N}_0 \mid \quad 0 \le i < n_{\text{tangle},02} \} && \text{with} \quad |\mathcal{T}| = n_{\text{tangle},02} \\ \mathcal{P} &= \{ i \in \mathbb{N}_0 \mid n_{\text{tangle}} \le i < n_{\text{parasite},02} + n_{\text{tangle},02} \} && \text{with} \quad |\mathcal{P}| = n_{\text{parasite},02} \end{aligned}$$

The transition probability from (1.1) is then collapsed into two cases.

1. We have $x \in \mathcal{T}$, i.e. current state is in Main Tangle. Then

$$P_{xy}^{(f)} = \begin{cases} 1 & \text{if } x = n_{\text{tangle},02} - 1 = y \\ q & \text{if } x < n_{\text{tangle},02} - 1 \text{ and } x = y + 1, \text{ or } x = y = 0 \\ w_x^{(f)} & \text{if } x \le n_{\text{tangle},01} - 1 \text{ and } x = y - 1 \text{ and } x < n_{\text{tangle},02} - 1 \\ 1 - q & \text{if } x > n_{\text{tangle},01} - 1 \text{ and } x = y - 1 \text{ and } x < n_{\text{tangle},02} - 1 \\ 1 - q - w_x^{(f)} & \text{if } x \le n_{\text{tangle},01} - 2 \text{ and } y = \operatorname{argmin}\{ z \in \mathcal{P} \mid t_z \ge t_x \} \\ 1 - q - w_x^{(f)} - \sum_{z \in \mathcal{P}: z \ne y} v_z^{(f)} & \text{if } x = n_{\text{tangle},01} - 1, \text{ and } y \ge n_{\text{tangle},02} + n_{\text{parasite},01} - 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$w_x^{(f)} = (1-q)\cdot\Big(\\ \mathbf{1}_{\{x\le n_{\text{tangle},01}-2\}}\cdot\Big(\tfrac{\lambda}{\lambda+\mu}+\tfrac{\mu}{\lambda+\mu}\cdot\tfrac{2\cdot f(2h(\lambda+\mu))}{2\cdot f(2h(\lambda+\mu))+f((T_2-t_x)\lambda)}\Big)+\\ \mathbf{1}_{\{x=n_{\text{tangle},01}-1\}}\cdot\Big(\tfrac{2h\lambda}{2h\lambda+(T_2-T_1)\mu}+\tfrac{(T_2-T_1)\mu}{2h\lambda+(T_2-T_1)\mu}\cdot\tfrac{2\cdot f(2h(\lambda+\mu))}{2\cdot f(2h(\lambda+\mu))+\sum_{z\in\mathcal{P}:t_z\ge t_x}f((T_2-t_z)\lambda)}\Big)\Big)$$

$$v_z^{(f)} = (1-q)\cdot\frac{(T_2-T_1)\mu}{2h\lambda+(T_2-T_1)\mu}\cdot\frac{f\left((T_2-T_z)\lambda\right)}{2\cdot f(2h(\lambda+\mu))+\sum_{y\in\mathcal{P}:t_y\ge T_1}f\left((T_2-T_z)\lambda\right)}$$

and

$$t_z=\begin{cases}2hz & \text{if } z\in\mathcal{T}\\ (z-n_{\text{tangle},02})\cdot\mu^{-1} & \text{if } z\in\mathcal{P}\end{cases}$$

The formula for $w_x^{(f)}$ can be derived via a binomial Tree approach. Noting that each parasite chain vertex references/approves one vertex in the main tangle (and one in the parasite chain) the probability of any vertex in the main tangle being referenced is simply driven by the size of λ,μ (virtue of observation 4 above). That is,

$$\mathbb{P}(\text{A given tangle vertex is not referenced by any vertex from the parasite chain})=\tfrac{\lambda\delta}{\lambda\delta+\mu\delta}=\tfrac{\lambda}{\lambda+\mu}$$

Secondly, conditional on the current tangle vertex being referenced by some parasite vertex, the probability of continuing the random walk in the main tangle is driven by observations 3 & 5 and yields

$$\mathbb{P}(\text{Stay in Tangle}\mid\text{Current vertex } x \text{ is referenced by parasite chain})=\tfrac{2\cdot f(2h(\lambda+\mu))}{2\cdot f(2h(\lambda+\mu))+f((T_2-t_x)\lambda)}$$

2. We have $x\in\mathcal{P}$, i.e. current state is in Parasite Chain. Then

$$P_{xy}^{(f)}=\begin{cases}1 & \text{if } x=n_{\text{tangle},02}+n_{\text{parasite},02}-1=y\\ q & \text{if } x=n_{\text{tangle},02} \text{ and } y=0\\ q/2 & \text{if } n_{\text{tangle},02}<x<n_{\text{tangle},02}+n_{\text{parasite},02}-1 \text{ and } x=y+1\\ q/2 & \text{if } n_{\text{tangle},02}<x<n_{\text{tangle},02}+n_{\text{parasite},02}-1 \text{ and } y=\operatorname{argmin}\{z\in\mathcal{T}\mid t_z\ge t_x\}\\ 1-q & \text{if } x<n_{\text{tangle},02}+n_{\text{parasite},02}-1 \text{ and } x=y-1\\ 0 & \text{otherwise}\end{cases}$$

These probabilities are considerably easier to understand/derive as the ones from the previous case.

Figure 2 provides an illustrative example of the Two-Way Markov-Chain Model, together with sample probabilities $P_{xy}^{(f)}$ along the transition arrows.

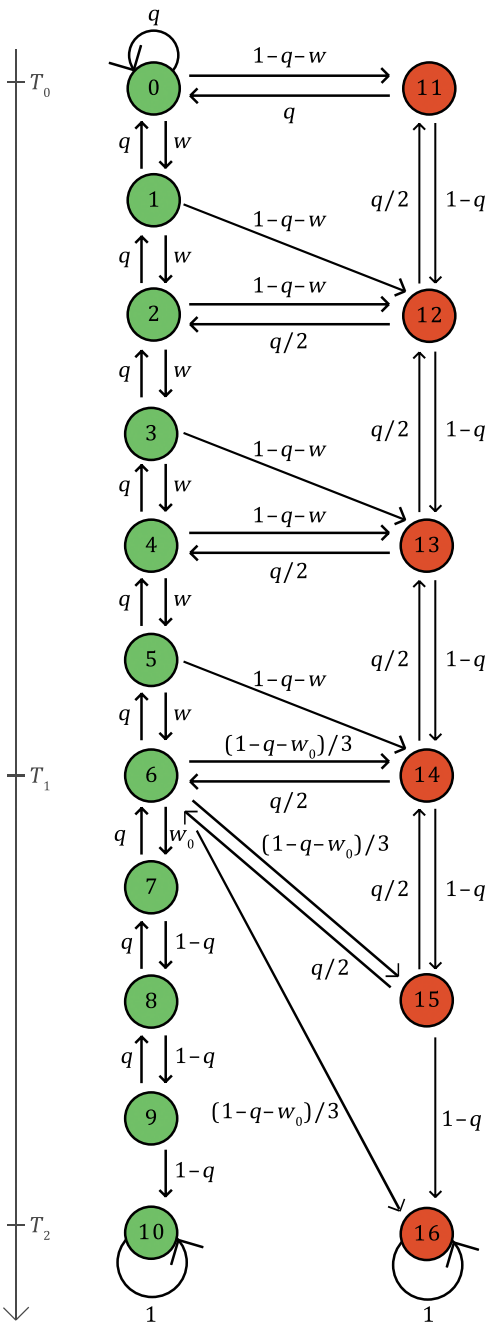

Figure 2: Illustration of the Two-Way Markov-Chain Model using parameters $h = 1$, $\mu = 0.25$, $T_1 - T_0 = 12$, $T_2 - T_1 = 8$

3 Absorption Probability

3.1 Derivation

The transition probabilities $P_{xy}^{(f)}$ from Section 2 can be used to build a full transition matrix $P \in \mathbb{R}^{m \times m}$ where $m = n_{\text{tangle},02} + n_{\text{parasite},02}$. Note that the final states in both the Tangle Markov-Chain and Parasite Markov-Chain are absorbing, so their rows in P only consist of 0s and one single 1. This means we can transform P to have the following shape:

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I_2 \end{array}\right) \qquad \text{with} \qquad I_k = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$$

where I_k is the identity matrix, $Q \in \mathbb{R}^{(m-2)\times(m-2)}$ contains the transition probabilities between non-absorbing states and $R \in \mathbb{R}^{(m-2)\times 2}$ contains the transition probabilities from non-absorbing states to the 2 absorbing ones.

Following standard procedure – cf. Section 11.2 from [Gri09] – we define

$$N = (I_{m-2} - Q)^{-1} \in \mathbb{R}^{(m-2)\times(m-2)} \qquad \text{and} \qquad B = N \cdot R \in \mathbb{R}^{(m-2)\times 2}$$

This gives

$$\begin{aligned} &\mathbb{P}(\text{Starting in the first tangle chain state \& Random walk is absorbed into last tangle state}) &&= b_{11} \\ &\mathbb{P}(\text{Starting in the first tangle chain state \& Random walk is absorbed into last parasite state}) &&= b_{12} \end{aligned}$$

with b_{ij} being the matrix entries of B. Clearly, $b_{11} + b_{12} = 1$.

Fundamental Result. Entry b_{12} is the probability that any MCMC particle walk (starting from genesis or another reasonable vertex) stops at a tip of the parasite chain.

In other words, b_{12} is the success probability of a parasite chain attack.

3.2 Field Tests

For initial simplicity we shall test the devised framework from Sections 2 and 3.1 by setting $f \equiv 1$, which is equivalent to

$$f(s) = \exp(-\alpha s) \quad \text{with} \quad \alpha = 0$$

Calculating the parasite chain success probability via the Two-Way Markov-Chain Model is virtually instantaneous. To estimate the "true" success probability we simulate the Tangle using a continuous-time model, generating 50 random tangles (with accompanying parasite chains) and then initiating 10 000 MCMC particle walks. The following plots provide a summary of these results.

3.2.1 Test 1

The derived formulae are tested for varying values of $\lambda = 2, 3, 4, 5, 6, 7, 8, 9, 10$ with fixed values $\mu = 0.1, h = 1, q = 0, T_1 - T_0 = 80, T_2 - T_1 = 10$.

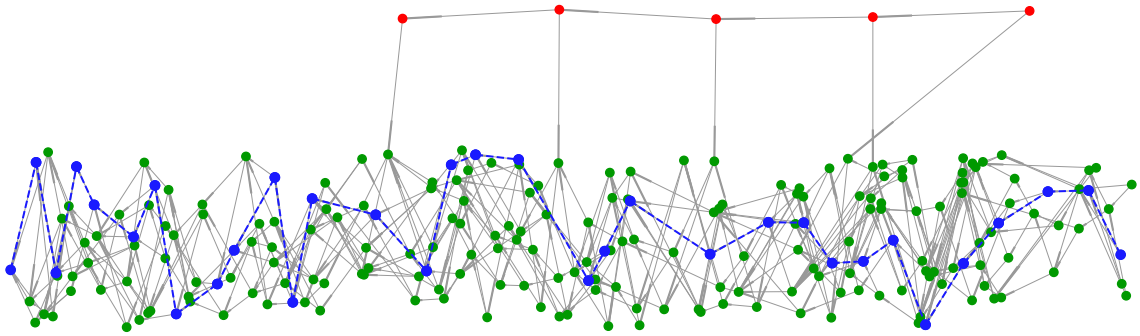

Figure 3: Sample Tangle (in green) for Test 1 alongside a Parasite chain (in red) together with a sample random walk (in blue)

The results are shown in Figure 4.

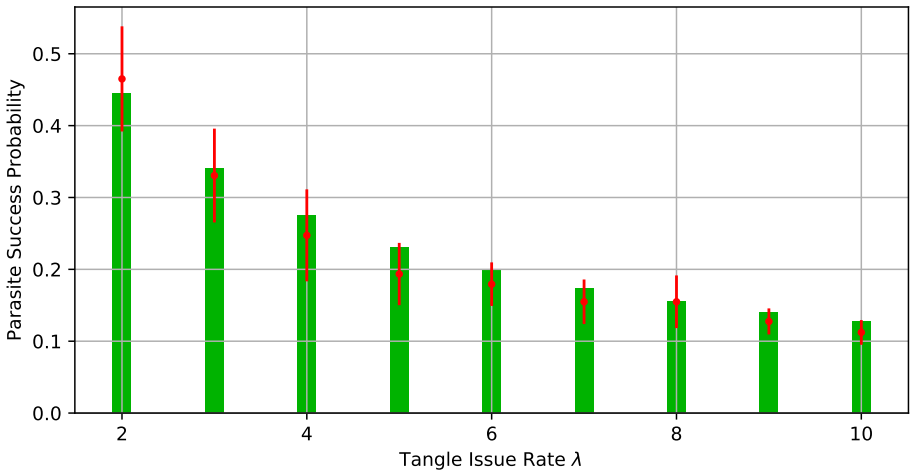

Figure 4: Results for Test 1 with a $[0.1, 0.9]$ confidence interval based on simulations (red lines) versus analytically computed numbers (green bars)

3.2.2 Test 2

The derived formulae are tested for varying values of $q = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45$ with fixed values $\mu = 0.25, h = 2, \lambda = 6, T_1 - T_0 = 100, T_2 - T_1 = 60$.

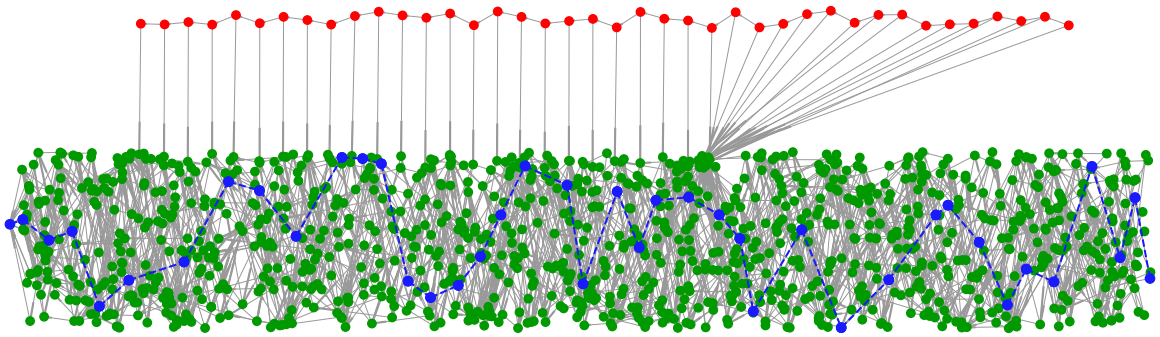

Figure 5: Sample Tangle (in green) for Test 2 alongside a Parasite chain (in red) together with a sample random walk (in blue)

The results are shown in Figure 6.

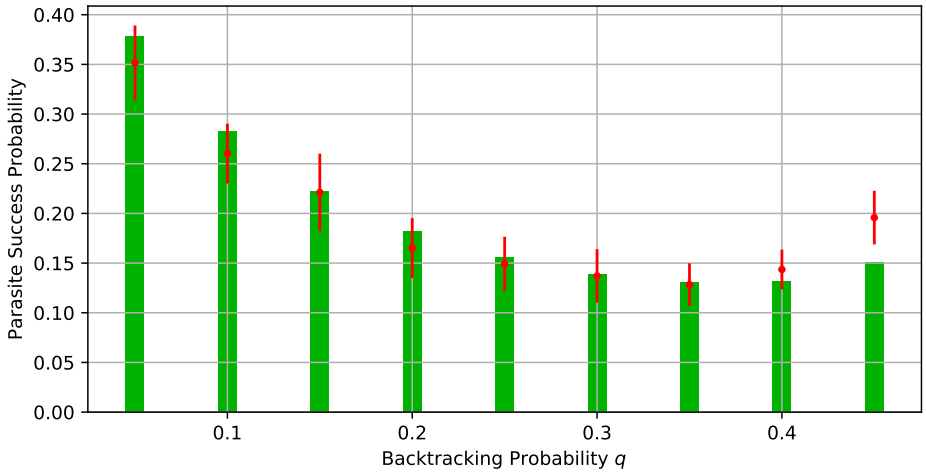

Figure 6: Results for Test 2 with a $[0.1, 0.9]$ confidence interval based on simulations (red lines) versus analytically computed numbers (green bars)

3.2.3 Test 3

The derived formulae are tested for varying values of $q = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45$ with fixed values $\mu = 0.125, h = 4, \lambda = 10, T_1 - T_0 = 304, T_2 - T_1 = 48$.

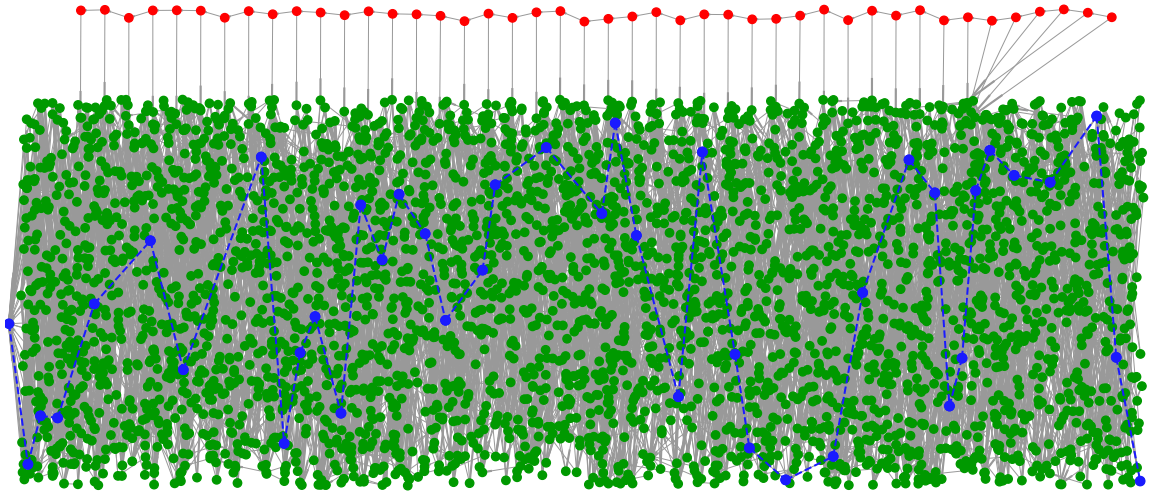

Figure 7: Sample Tangle (in green) for Test 3 alongside a Parasite chain (in red) together with a sample random walk (in blue)

The results are shown in Figure 8.

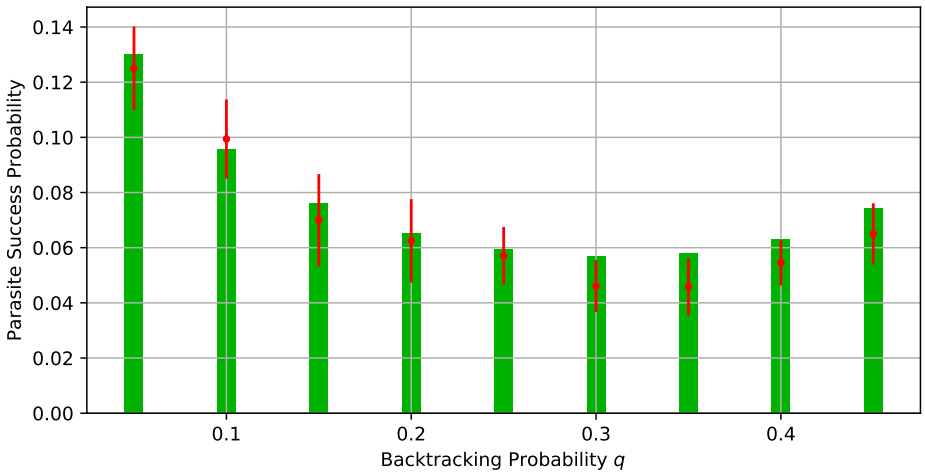

Figure 8: Results for Test 3 with a $[0.1, 0.9]$ confidence interval based on simulations (red lines) versus analytically computed numbers (green bars)

4 Conclusion And Extensions

The analysis from 3.2 is meant to verify that the probabilities produced by the Two-Way Markov-Chain Model are indeed viable. The results suggest that the quasi-analytic formulae are by no means perfect, but provide a very good first estimate. It is obvious that additional tweaking of transition probabilities is required to make the framework more accurate. In particular, the probabilities around T_1 (time when the first tx of the double-spend is placed) prove to be challenging as the single Markov state at T_1 can easily break the results.

As a next step, it would be desirable to analyse how the Tangle could defend an attacker by calibrating q and α as to minimise the success probability of a parasite chain attack. As Bart pointed out, we could also try using Perturbation Theory to derive analytical results (cf. Gateaux derivative) and additionally check whether Mean Field Approximation gives similar results.

As a final step, one could try extending this type of framework to assess the breadth of the *tip-selection exit distribution* (or entropy) dependent on α at any given point in time. Then by tweaking parameters α and q one can solve the optimisation problem of

$\nearrow$ Maximising breadth of tip-selection exit distribution,
$\searrow$ Minimising parasite chain success probabilities

If this framework proves successful, we could incorporate it into future IOTA nodes so that they can continuously select theoretically-perfect values for q and α, and then perform tip selection based on those calibrated parameters.

Bibliography

[Gri09] Charles Grinstead. *Grinstead and Snell's Introduction to Probability*. University Press of Florida, 2009.

[PSF17] Serguei Popov, Olivia Saa, and Paulo Finardi. *Equilibria in the Tangle*. `https://arxiv.org/abs/1712.05385`, December 2017.