BSc Thesis Applied Mathematics

Influence of bursty interaction patterns on tie strength in tie-decay temporal networks

Anna Dankers

Supervisor: dr. ir. M. de Graaf

January 24, 2020

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Preface

January 24, 2020

This paper was written to fulfill the graduation requirements of the bachelor Applied Mathematics at the University of Twente. The research was performed from September 2, 2019 up to January 24, 2020.

Influence of bursty interaction patterns on tie strength in tie-decay temporal networks

Anna Dankers

January 24, 2020

Abstract

This paper investigates the influence of bursty interactions patterns on the tie strength and the clustering coefficient in tie-decay temporal networks. To model bursty interaction patterns, we use a Markov process with two states and a high and low interaction probability. This Markov model is used to find expressions for the expected value and variance of tie strength for different degrees of burstiness. Moreover, we explore the behaviour of the clustering coefficient of networks with bursty interaction patterns. We find that the mean tie strength remains constant for different degrees of burstiness. However, the expected value of peaks in tie strength and the variance of tie strength increase as interaction patterns get more bursty, which results in a decreasing clustering coefficient.

Keywords: tie strength, temporal networks, bursty interaction patterns, weighted clustering coefficient

1 Introduction

In social networks, people interact with each other. The frequencies of interactions may differ per person and these interactions influence the structure of a network. How people relate to each other is an interesting property of a social network. For various purposes, for example the study of spread of information and disease, it is useful to analyze the properties of these networks. This can be done using variables such as *tie strength* and *clustering coefficient*. The clustering coefficient is a measure of the 'clustering' in a network: the extent to which neighbours of a node are connected to each other. In the binary case (either two nodes are connected, or they are not connected), the definition of clustering coefficient is straightforward. However, connections in real life are often not binary and the relation between two people has a certain (continuous) strength. This strength of connections is called *tie strength*. The tie strength of an edge is determined by the frequency of interactions between the nodes incident to that edge. Therefore, we should use another definition for clustering coefficient: one that includes the tie strength.

Besides the fact that ties are not binary, an important property of social networks is that they change over time. These types of networks are called temporal networks. Because of this, the clustering coefficient does also change over time. We separate the concepts *interactions* and *ties*: interactions occur in discrete time, while ties change continuously in time. To model this, we build on the work of Ahmad et al. [1] and Zuo and Porter [3]. The general model that is used is the tie-decay model of Ahmad et al [1]: at each time step, two entities can have an interaction or not; both possibilities affect the tie strength s_t of that pair of nodes at time t. If there is an interaction, the tie strength increases by 1. Otherwise, the tie strength is multiplied by a factor $e^{-\alpha}$, where $\alpha > 0$ is the decay factor. An illustration is given in Figure 1. In choosing a value for α, one can think of the half-life of a tie, $\eta_{1/2} = \frac{log2}{\alpha}$, where *log* denotes the natural logarithm. In this paper, we use $\alpha = 0.01$, unless otherwise stated.

Zuo and Porter have found an expression for the long-term expected value of tie strength, which we will revisit in Chapter 2. They used an interaction probability p for a pair of nodes. This

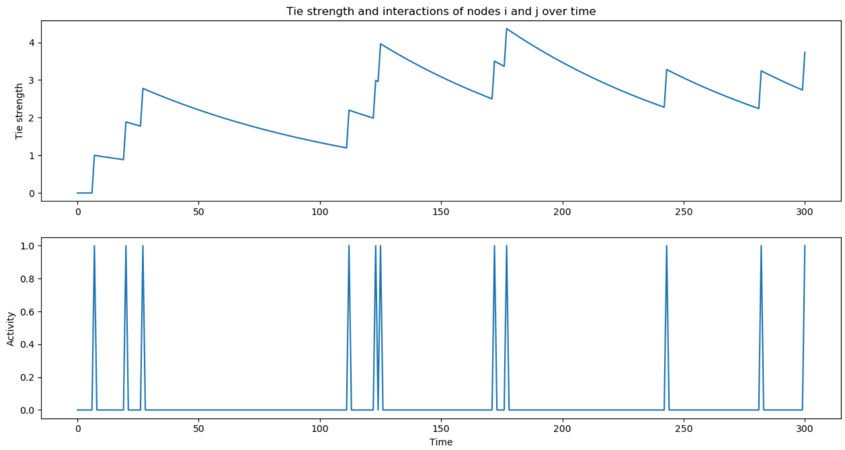

FIGURE 1: Illustration of tie strength of the edge between two nodes with interaction probability $r = 0.03$ and $\alpha = 0.01$. The peaks in the second graph represent interactions.

probability is constant over time and does not depend on the history of the edge. This creates a 'uniform' interaction pattern. However, in social networks interaction patterns often follow bursty patterns: sometimes people interact very frequently in a short time period, but it is also possible that the time between two interactions is long. We model these interaction patterns using a Markov process. We also investigate if the long-term expected value of tie strength differs from the value found by Zuo and Porter and calculate the variance of tie strength for different degrees of 'burstiness'. Finally, we present a definition for a weighted clustering coefficient and investigate the behaviour of this clustering coefficient when interaction patterns are bursty.

2 Behaviour of tie strength

2.1 Model of Zuo and Porter: one interaction probability

In the model of Zuo and Porter [3], at each time step there is an interaction between two entities with probability $0 \leq r \leq 1$. Similar to the model of Ahmad et al, the tie strength increases by 1 if there is an interaction. Otherwise, the tie strength is multiplied by a factor $e^{-\alpha}$.

Definition 2.1. If x_t is a Bernoulli random variable with parameter p that indicates whether there is an interaction ($x_t = 1$) or not ($x_t = 0$) at time t, we can write the tie strength s_t as

$$s_t = x_t + e^{-\alpha(1-x_t)} s_{t-1}.$$

Using $p = r$ as success probability for the Bernoulli random variable x_t, we obtain by the law of total expectation

$$E[s_t] = E\left[x_t + e^{-\alpha(1-x_t)} s_{t-1}\right] = E[x_t] + E\left[e^{-\alpha(1-x_t)} s_{t-1} | x_t = 0\right] P(x_t = 0) + E\left[e^{-\alpha(1-x_t)} s_{t-1} | x_t = 1\right] P(x_t = 1)$$

$$= r + (1-r)E\left[e^{-\alpha} s_{t-1}\right] + rE[s_{t-1}] = r(1 + E[s_{t-1}]) + (1-r)E[s_{t-1}]e^{-\alpha}. \quad (1)$$

It can be proved that we reach a stationary state as $t \to \infty$ [3]. In this state we have $E[s_t] = E[s_{t-1}]$, so the next theorem follows.

Theorem 2.1 (Zuo and Porter [3])**.** *In the long term, the tie strength of an edge with interaction probability r and decay factor α is given by*

$$E[s] = r(1 + E[s]) + (1-r)E[s]e^{-\alpha} = \frac{r}{(1-e^{-\alpha})(1-r)}.$$

As a check, for $r = 0.03$ and $\alpha = 0.01$, we have $E[s] \approx 3.11$. In a simulation of a network with 3 nodes where interactions between nodes are simulated with probability $r = 0.03$ at each time step, we obtain average values 3.19, 2.85 and 3.00 (see Figure 2). These values correspond to the expected value calculated using Theorem 2.1.

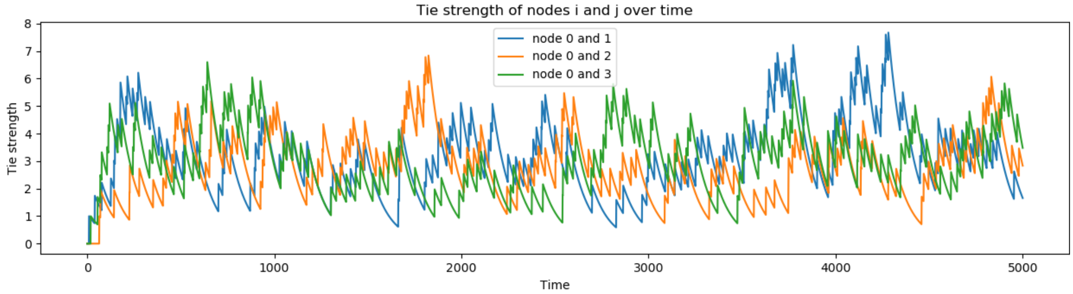

FIGURE 2: Simulation of tie strength of three edges, with $r = 0.03$ and $\alpha = 0.01$. The mean tie strength m_{ij} is $m_{01} \approx 3.19$, $m_{02} \approx 2.85$ and $m_{03} \approx 3.00$.

2.2 A Markov process

2.2.1 The model

In real life, most interaction patterns cannot be described by one interaction probability that is the same for each time. Interaction patterns alternate between *interaction periods* and *idle periods*, and most of the time these periods take more than one time step. To model this, we consider the random variable x_t again, but now we assume that it follows a Markov process. Let x_t be an indicator variable with sample space $S = \{0, 1\}$, where $x_t = 1$ if there is an interaction at time t. Otherwise, $x_t = 0$. Thus, the Markov process followed by x_t has two states: '0' (no interaction) and '1' (interaction).

Definition 2.2. An *interaction period* is an interval $[a, b]$ such that $x_t = 1$ for all $a \leq t \leq b$ and $x_{a-1} = x_{b+1} = 0$. An *idle period* is an interval $[a, b]$ such that $x_t = 0$ for all $a \leq t \leq b$ and $x_{a-1} = x_{b+1} = 1$.

We define

$$P(x_t = 1 | x_{t-1} = 1) = p$$
$$P(x_t = 0 | x_{t-1} = 1) = 1 - p$$
$$P(x_t = 1 | x_{t-1} = 0) = q$$
$$P(x_t = 0 | x_{t-1} = 0) = 1 - q$$

(see Figure 3).

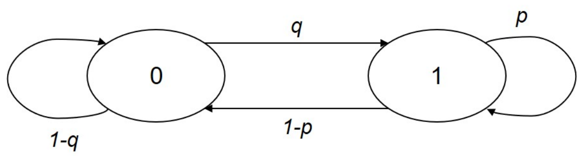

FIGURE 3: Transition diagram

In words, p is the probability that a pair of nodes that has an interaction now is still interacting the next time step, and q is the probability that a pair of nodes will interact during the next time step if there is no interaction now. We will call p the *active interaction probability* and q the *passive* *interaction probability.* The burstiness of interaction patterns is determined by the choice of p and q: the greater the difference between p and q, the greater the degree of burstiness. In this paper, we will have $p > q$. This is based on the fact that interactions in social networks occur in periods: the probability that two nodes 'keep interacting' is higher than the probability that nodes start interacting. Using these probabilities, we can define the transition probability matrix as follows:

$$T = \begin{bmatrix} 1-q & q \\ 1-p & p \end{bmatrix}$$

We can calculate steady state probabilities using the following equations (where t_{ij} is the ijth element of matrix T):

$$\pi_0 = \pi_0 t_{00} + \pi_1 t_{10}$$

$$\pi_1 = \pi_0 t_{01} + \pi_1 t_{11}$$

$$\pi_0 + \pi_1 = 1.$$

It can be shown that steady state probabilities π_0 and π_1 are given by

$$\pi_0 = \frac{1-p}{1-p+q}$$

$$\pi_1 = \frac{q}{1-p+q}.$$

Steady state probabilities can be interpreted as the fraction of time the system is in that state: π_1 is the fraction of time two nodes are interacting.

An example is given in Figure 4. Notice that the model of Zuo and Porter can also be modelled as a Markov process if we take $p = q = r$.

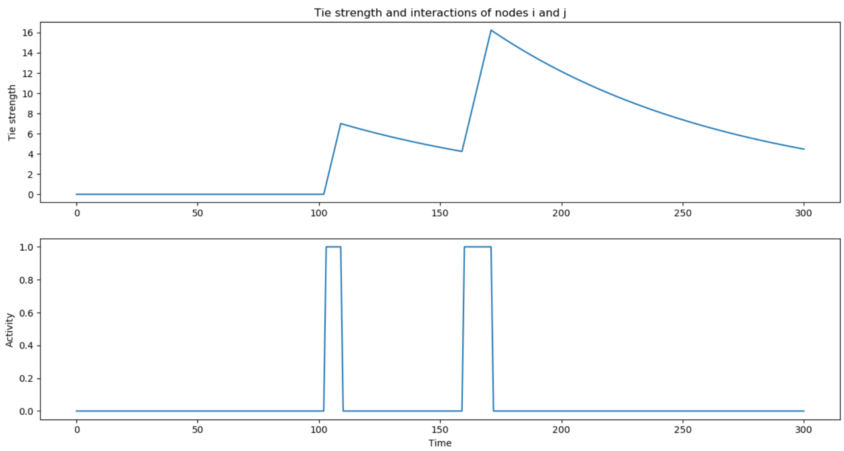

FIGURE 4: Simulation of tie strength of one edge, with $p = 0.81$ and $q = 0.006$ (so, $\pi_0 = 0.97$ and $\pi_1 = 0.03$). This simulation has two interaction periods ($[101, 108]$ and $[157, 169]$) and three idle periods.

2.2.2 Expected value

If we look at Figure 1 and Figure 4, it is clear that the tie strength in the second model grows much faster than the tie strength in the first model. However, idle periods in the second model last longer, which causes the tie strength to decay more. Therefore, it is natural to ask if the mean tie strength in the Markov model differs from the mean tie strength in the model of Zuo and Porter.

Theorem 2.2. *For equal stationary interaction probabilities (that is, $r = \pi_1$), the mean value of tie strength in a model where interactions follow a Markov process is equal to the mean value of tie strength in a model with constant interaction probability.*

Proof. For the expected value of tie strength in the Markov model, we have by the law of total expectation

$$E[s_t] = E[s_t|x_{t-1} = 0]P(x_{t-1} = 0) + E[s_t|x_{t-1} = 1]P(x_{t-1} = 1). \tag{2}$$

In stationary state, we have $P(x_{t-1} = 0) = \pi_0$ and $P(x_{t-1} = 1) = \pi_1$. Moreover,

$$E[s_t|x_{t-1} = 0] = E[x_t + e^{-\alpha(1-x_t)}s_{t-1}|x_{t-1} = 0] = E[x_t|x_{t-1} = 0]$$

$$+E[e^{-\alpha(1-x_t)}s_{t-1}|x_t = 0, x_{t-1} = 0]P(x_t = 0|x_{t-1} = 0)+E[e^{-\alpha(1-x_t)}s_{t-1}|x_t = 1, x_{t-1} = 0]P(x_t = 1|x_{t-1} = 0)$$

$$= q + (1-q)e^{-\alpha}E[s_{t-1}] + qE[s_{t-1}] = q(1 + E[s_{t-1}]) + (1-q)E[s_{t-1}]e^{-\alpha} \tag{3}$$

and

$$E[s_t|x_{t-1} = 1] = E[x_t + e^{-\alpha(1-x_t)}s_{t-1}|x_{t-1} = 1] = E[x_t|x_{t-1} = 1]$$

$$+E[e^{-\alpha(1-x_t)}s_{t-1}|x_t = 0, x_{t-1} = 1]P(x_t = 0|x_{t-1} = 1)+E[e^{-\alpha(1-x_t)}s_{t-1}|x_t = 1, x_{t-1} = 1]P(x_t = 1|x_{t-1} = 1)$$

$$= p + (1-p)e^{-\alpha}E[s_{t-1}] + pE[s_{t-1}] = p(1 + E[s_{t-1}]) + (1-p)E[s_{t-1}]e^{-\alpha}. \tag{4}$$

Substituting (3) and (4) in (2) yields

$$E[s_t] = (\pi_0 q + \pi_1 p)(1 + E[s_{t-1}]) + (\pi_0(1-q) + \pi_1(1-p)E[s_{t-1}]e^{-\alpha}) \tag{5}$$

Since

$$\pi_0 q + \pi_1 p = \frac{q(1-p)}{1-p+q} + \frac{pq}{1-p+q} = \frac{q}{1-p+q} = \pi_1$$

and

$$\pi_0(1-q) + \pi_1(1-p) = \frac{(1-q)(1-p) + q(1-p)}{1-p+q} = \frac{1-p}{1-p+q} = \pi_0,$$

(5) can be rewritten

$$E[s_t] = \pi_1(1 + E[s_{t-1}]) + \pi_0 E[s_{t-1}]e^{-\alpha}$$

and this yields the exact same result as (1) if we choose $r = \pi_1$. □

From this theorem, it is possible to deduce the expected value of tie strength in stationary state expressed in terms of π_0 and π_1.

Corollary 2.1. In stationary state, the expected value of tie strength for stationary interaction probabilities π_0 and π_1 is

$$E[s] = \frac{\pi_1}{\pi_0(1 - e^{-\alpha})}.$$

2.3 Interaction cycles

In every interaction pattern, interaction periods are alternated by idle periods. These periods can be viewed as the consecutive time steps the Markov chain is in a state. In this section, we no longer use the recursive expression for $E[s_t]$, but we use interaction and idle periods to calculate the expected value and variance of the strength. We introduce the following definition:

Definition 2.3. An *interaction cycle* is one interaction period and a subsequent idle period.

Furthermore, let a_k be the length of the kth interaction period in time steps and let b_k be the length of the kth idle period in time steps. With these definitions, we can state an expression for the tie strength after $n + 1$ interaction cycles u_n:

$$u_0 = a_0 e^{-\alpha b_0}$$

$$u_1 = (a_0 e^{-\alpha b_0} + a_1)e^{-\alpha b_1}$$

$$u_2 = ((a_0 e^{-\alpha b_0} + a_1)e^{-\alpha b_1} + a_2)e^{-\alpha b_2}$$

$$\vdots$$

$$u_n = (u_{n-1} + a_n)e^{-\alpha b_n}$$

The expression above can be explained by the fact that during an interaction period, the tie strength increases by one every time step. Thus, in interaction cycle k, the tie strength increases by a_k in total. After that, the tie strength is multiplied by $e^{-\alpha}$ for each time step in an idle period. So in total, the tie strength decreases by a factor $(e^{-\alpha})^{b_k}$.

We can also write

$$u_n = a_0 e^{-\alpha(b_0+b_1+\ldots+b_n)} + a_1 e^{-\alpha(b_1+b_2+\ldots+b_n)} + \ldots + a_{n-1}e^{-\alpha(b_{n-1}+b_n)} + a_n e^{-\alpha b_n} \tag{6}$$

or, in compact form:

$$u_n = \sum_{k=0}^{n} a_k e^{-\alpha \sum_{i=k}^{n} b_i}$$

Since an interaction cycle always ends with an idle period, u_n is the tie strength after the nth idle period (if we start counting after the first interaction period). This is illustrated in Figure 5. Analogously, we can state an expression for the tie strength after the nth interaction period v_n:

$$v_0 = a_0$$

$$v_1 = a_0 e^{-\alpha b_0} + a_1$$

$$v_2 = (a_0 e^{-\alpha b_0} + a_1)e^{-\alpha b_1} + a_2$$

$$\vdots$$

$$v_n = v_{n-1} + a_n$$

which can be written as

$$v_n = \sum_{k=0}^{n} a_k e^{-\alpha \sum_{i=k}^{n-1} b_i}.$$

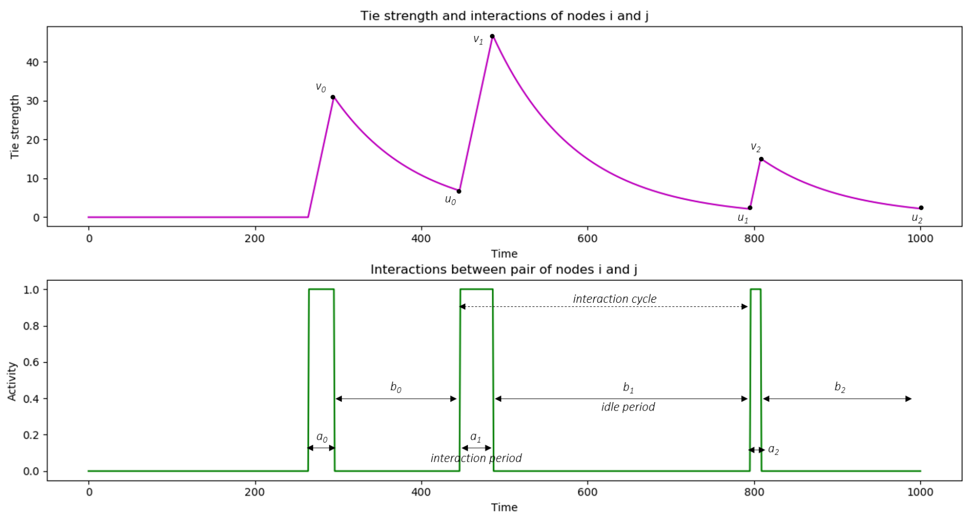

Figure 5: Illustration of interaction cycles, interaction periods and idle periods. The length of the kth interaction period is a_k, the length of the kth idle period is b_k. The tie strength after the kth interaction period is v_k, the tie strength after the kth interaction cycle is u_k.

2.3.1 Expected value

In this section, we do not assume a_k and b_k to be known. Instead, we treat them as random variables. They are defined as follows:

- a_k is the length of the kth interaction period in time units. a_k is geometrically distributed with success probability $1-p$, mean $\frac{1}{1-p}$ and probability mass function $P(a_k = x) = p^{x-1}(1-p)$ for all k.
- b_k is the length of the kth 'no-interaction period' in time units. b_k is geometrically distributed with success probability q, mean $\frac{1}{q}$ and probability mass function $P(b_k = y) = (1-q)^{y-1}q$ for all k.

Here, p and q are the active and passive probability respectively, as defined earlier (Figure 3).

Lemma 2.1. If b_k is the length of the kth idle period in time units, then

$$E[e^{-\alpha b_k}] = \frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}$$

for all k.

Proof. For discrete variables, the expected value of a function of a random variable is

$$E\left[f(X)\right] = \sum_{x \in D} f(x)P(X=x).$$

Therefore, we have

$$E[e^{-\alpha b_k}] = \sum_{i=1}^{\infty} e^{-\alpha i}(1-q)^{i-1}q = \frac{q}{1-q}\sum_{i=1}^{\infty}\left((1-q)e^{-\alpha}\right)^i.$$

Since

$$\left|(1-q)e^{-\alpha}\right| < 1,$$

the series converges and the sum is

$$E[e^{-\alpha b_k}] = \frac{q}{1-q}\cdot\left(\frac{1}{1-(1-q)e^{-\alpha}} - 1\right) = \frac{q}{1-q}\cdot\frac{(1-q)e^{-\alpha}}{1-(1-q)e^{-\alpha}} = \frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}$$

for all k. □

It can be shown that the values of u_k and v_k converge to numbers u and v. The next theorem gives expressions for the long-term expected value of tie strength after an interaction period and after an interaction cycle:

Theorem 2.3. *The expected value of tie strength after the nth interaction period is given by*

$$E[v_n] = \frac{(1-(1-q)e^{-\alpha})\left(1-\left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{n+1}\right)}{(1-p)(1-e^{-\alpha})}$$

and converges to

$$E[v] = \frac{1-(1-q)e^{-\alpha}}{(1-p)(1-e^{-\alpha})}$$

as $n \to \infty$. Furthermore, $E[u_n]$ also converges as $n \to \infty$ and

$$E[u] = \frac{qe^{-\alpha}}{(1-p)(1-e^{-\alpha})}.$$

Proof. We can write v_n as

$$v_n = a_0 e^{-\alpha(b_0+\ldots+b_{n-1})} + a_1 e^{-\alpha(b_1+\ldots+b_{n-1})} + \ldots + a_{n-1} e^{-\alpha(b_{n-1})} + a_n. \quad (7)$$

To calculate the expected value, notice that all a_k and b_k are independent, and hence

$$E\left[a_k e^{-\alpha(b_k+\ldots+b_n)}\right] = E\left[a_k\right] E\left[e^{-\alpha b_k}\right] \ldots E\left[e^{-\alpha b_n}\right].$$

Now, using Lemma 2.1 and the fact that $E[a_k] = \frac{1}{1-p} = E[a]$ for all k, we can rewrite (7) as follows:

$$E[v_n] = \sum_{k=0}^{n} E[a] \left(E\left[e^{-\alpha b}\right]\right)^k = \frac{1}{1-p} \sum_{k=0}^{n} \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right)^k.$$

This series can be rewritten and the sum is

$$E[v_n] = \frac{(1-(1-q)e^{-\alpha}) \left(1 - \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{n+1}\right)}{(1-p)(1-e^{-\alpha})}.$$

Since

$$\left| \frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right| < 1,$$

we obtain

$$E[v] = \frac{1-(1-q)e^{-\alpha}}{(1-p)(1-e^{-\alpha})}$$

in the long term limit (as $n \to \infty$). Moreover,

$$u_n = v_n e^{-\alpha b_n}$$

and hence

$$E[u_n] = E[v_n]E[e^{-\alpha b_n}] = \frac{(1-(1-q)e^{-\alpha}) \left(1 - \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{n+1}\right)}{(1-p)(1-e^{-\alpha})} \cdot \frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}$$

which can be rewritten

$$E[u_n] = \frac{qe^{-\alpha} \left(1 - \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{n+1}\right)}{(1-p)(1-e^{-\alpha})}.$$

We conclude that

$$E[u] = \frac{qe^{-\alpha}}{(1-p)(1-e^{-\alpha})}$$

as $n \to \infty$. □

Some numerical results are given in Table 1. Here, it can be seen that $E[u]$ is constant and that $E[v]$ grows larger as the interaction pattern gets more bursty, that is, if the difference between p and q gets larger. This can be explained by the fact that for a greater difference in p and q, interactions periods last longer, so the tie strength grows much faster, resulting in a growing value of $E[v]$. This can be deduced from the expected value of the length of an interaction period: $E[a_k] = (1-p)^{-1}$, so for greater p, $E[a_k]$ is also greater. However, idle periods last longer too: $E[b_k] = q^{-1}$, so if q gets smaller, $E[b_k]$ grows. Because of this, the tie strength also decreases more, which results in a constant value of $E[u]$.

π_1	p	q	$E[v]$	$E[u]$
0.03	0.03	0.0300	3.077	4.11
0.03	0.60	0.0124	3.077	5.58
0.03	0.80	0.0062	3.077	8.08
0.03	0.90	0.0031	3.077	13.08
0.03	0.95	0.0015	3.077	23.08
0.03	0.98	0.0006	3.077	53.08
0.03	0.99	0.0003	3.077	103.08

TABLE 1: $E[u]$ and $E[v]$ for different active and passive interaction probabilities p and q.

2.3.2 Variance

Now that we know that the expected value of v_n grows larger as the burstiness gets larger (that is, the difference between p and q gets larger), we investigate if the variance does also grow if interaction patterns get more bursty. First, we present a preliminary result.

Lemma 2.2. The covariance of two terms of v_n, provided $i < j$, is

$$cov\left(a_i e^{-\alpha(b_i+\ldots+b_{n-1})}, a_j e^{-\alpha(b_j+\ldots+b_{n-1})}\right) = \frac{b(i,j)}{(1-p)^2}$$

where

$$b(i,j) = \left(\frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}}\right)^{n-j}\left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{j-i} - \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{2n-i-j}.$$

Proof. To calculate the covariance, we use the following simplified expression:

$$cov(X,Y) = E[XY] - E[X]E[Y].$$

Let $x_k = a_k e^{-\alpha(b_k+\ldots+b_{n-1})}$. Since

$$x_i \cdot x_j = a_i e^{-\alpha(b_i+\ldots+b_{n-1})} \cdot a_j e^{-\alpha(b_j+\ldots+b_{n-1})} = a_i a_j e^{-\alpha(b_i+\ldots+b_{j-1}+2(b_j+\ldots+b_{n-1}))}$$

for $i < j$, we have that

$$cov(x_i, x_j) = E\left[a_i a_j e^{-\alpha(b_i+\ldots+b_{j-1}+2(b_j+\ldots+b_{n-1}))}\right] - E\left[a_i e^{-\alpha(b_i+\ldots+b_{n-1})}\right] E\left[a_j e^{-\alpha(b_j+\ldots+b_{n-1})}\right].$$

Since all factors are independent, we can rewrite this as follows:

$$E[a_i]E[a_j]E\left[e^{-\alpha b_i}\right]\cdot\ldots\cdot E\left[e^{-\alpha b_{j-1}}\right]E\left[e^{-2\alpha b_j}\right]\cdot\ldots\cdot E\left[e^{-2\alpha b_{n-1}}\right] -$$

$$E[a_i]E[a_j]E\left[e^{-\alpha b_i}\right]\cdot\ldots\cdot E\left[e^{-\alpha b_{j-1}}\right]\left(E\left[e^{-\alpha b_j}\right]\cdot\ldots\cdot E\left[e - \alpha b_{n-1}\right]\right)^2. \quad (8)$$

In a similar fashion to the proof of Lemma 2.1, it can be shown that

$$E[e^{-2\alpha b_k}] = \frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}}$$

for all k, and therefore (8) becomes

$$\left(\frac{1}{1-p}\right)^2\left(\left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{j-i}\left(\frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}}\right)^{n-j} - \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{j-i}\left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{2n-2j}\right)$$

which yields

$$\frac{1}{(1-p)^2}\left(\left(\frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}}\right)^{n-j}\left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{j-i} - \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}}\right)^{2n-i-j}\right).$$

□

This lemma is used for the following theorem.

Theorem 2.4. *The variance of v_n is given by*

$$var(v_n) = \sum_{k=0}^{n} \frac{1+p}{(1-p)^2} \left(\frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}} \right)^k - \frac{1}{(1-p)^2} \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right)^{2k} + cv_n$$

where

$$cv_n = 2 \sum_{0 \le i < j \le n} cov(a_i e^{-\alpha(b_i+\ldots+b_{n-1})}, a_j e^{-\alpha(b_j+\ldots+b_{n-1})}).$$

Proof. Since v_n can be written as a sum:

$$v_n = a_0 e^{-\alpha(b_0+\ldots+b_{n-1})} + a_1 e^{-\alpha(b_1+\ldots+b_{n-1})} + \ldots + a_{n-1} e^{-\alpha(b_{n-1})} + a_n,$$

we apply the following rule for the variance of a sum of dependent random variables:

$$var\left(\sum_{i=0}^{n} X_i \right) = \sum_{i=0}^{n} var(X_i) + 2 \sum_{0 \le i < j \le n} cov(X_i, X_j). \tag{9}$$

Let $x_k = a_k e^{-\alpha(b_k+\ldots+b_{n-1})}$. For the variance of a single term of v_n, we have

$$var(x_k) = E[x_k^2] - E[x_k]^2 = E\left[\left(a_k e^{-\alpha(b_k+\ldots+b_{n-1})} \right)^2 \right] - E\left[a_k e^{-\alpha(b_k+\ldots+b_{n-1})} \right]^2 =$$

$$E[a_k^2] E[e^{-2\alpha b_k}] \cdot \ldots \cdot E[e^{-2\alpha b_{n-1}}] - \left(E[a_k] E[e^{-\alpha}] \cdot \ldots \cdot E[e^{-\alpha b_{n-1}}] \right)^2.$$

For $E[a_k^2]$, we use the fact that a_k is geometrically distributed with probability $1-p$:

$$E[a_k^2] = var(a_k) + E[a_k]^2 = \frac{p}{(1-p)^2} + \frac{1}{(1-p)^2} = \frac{1+p}{(1-p)^2}.$$

Using that, we obtain

$$var(x_k) = \frac{1+p}{(1+p)^2} \left(\frac{qe^{-2\alpha}}{1/(1-q)e^{-2\alpha}} \right)^{n-k} - \left(\frac{1}{1-p} \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right)^{n-k} \right)^2. \tag{10}$$

Now, since $v_n = \sum_{k=0}^{n} x_k$, by (9) and (10) we have

$$var(v_n) = \sum_{k=0}^{n} \frac{1+p}{(1+p)^2} \left(\frac{qe^{-2\alpha}}{1/(1-q)e^{-2\alpha}} \right)^{n-k} - \frac{1}{(1-p)^2} \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right)^{2(n-k)} + 2 \sum_{0 \le i < j \le n} cov(x_i, x_j)$$

which can be rewritten as

$$var(v_n) = \sum_{k=0}^{n} \frac{1+p}{(1-p)^2} \left(\frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}} \right)^{k} - \frac{1}{(1-p)^2} \left(\frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right)^{2k} + 2 \sum_{0 \le i < j \le n} cov(x_i, x_j).$$

□

It can be shown that $var(v_n)$ converges as $n \to \infty$. We will not prove this, but the proof is based on the fact that

$$\left| \frac{qe^{-\alpha}}{1-(1-q)e^{-\alpha}} \right| < 1$$

and

$$\left| \frac{qe^{-2\alpha}}{1-(1-q)e^{-2\alpha}} \right| < 1$$

and therefore both terms go to zero as $n \to \infty$, so v_n converges to a stationary value. Some numerical results of this stationary value of the variance are presented in Table 2.

π_1	p	q	$var(u)$
0.03	0.03	0.0300	1.59
0.03	0.60	0.0124	9.83
0.03	0.80	0.0062	33.73
0.03	0.90	0.0031	119.04
0.03	0.95	0.0015	439.66
0.03	0.98	0.0006	2601.52
0.03	0.99	0.0003	10204.62

TABLE 2: $var(v)$ for different active and passive interaction probabilities p and q.

We conclude that the variance of v differs for varying degrees of burstiness: in a model where interactions follow a Markov process, the variance of tie strength at the end of an interaction period grows as the difference between p and q grows.

3 Clustering coefficient

In this section, we investigate the effect of bursty interaction patterns on the clustering coefficient. In case of unweighted networks, the clustering coefficient is defined as follows:

$$C = \frac{\delta_c}{\delta_o},$$

where δ_c is the number of closed triplets and δ_o is the number of open triplets. A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) ties. However, in this paper we are dealing with weighted ties, so there is no clear distinction between closed and open triplets. Therefore, we should use a definition of clustering coefficient that includes tie strength. We use the definition proposed by Kalna and Higham [2], but there are also other definitions of a weighted clustering coefficient. First, we present a method to calculate tie strength for a network in matrix form. This matrix is used to calculate the clustering coefficient.

Let $A(t)$ be the *interaction matrix* at time t. If there is an interaction between nodes i and j at time t, $a_{ij}(t) = 1$. Otherwise, $a_{ij}(t) = 0$. Since we only consider undirected networks, we have $a_{ij}(t) = a_{ji}(t)$ for all $i \neq j$, i.e. A(t) is symmetric.

Theorem 3.1. *If interactions between nodes i and j occur at times $\tau_0^{ij}, \ldots, \tau_n^{ij} < t$, the tie strength between nodes i and j at time t is given by entry s_{ij} of the matrix $S(t)$, for which we have*

$$S(t) = \sum_{k=1}^{t} e^{-\alpha(t-k)} A(k) + B(t), \tag{11}$$

where the matrix $B(t)$ is defined as follows:

$$b_{ij}(t) = \sum_{k=0}^{n} e^{(t-\tau_k^{ij})} \left(e^{\alpha(n-k)} - 1 \right).$$

Proof. For each interaction between pair of nodes i and j at time τ_k^{ij}, the tie strength is increased by $a_{ij}(\tau_k^{ij}) = 1$. If an interaction occurs at time $\tau_k^{ij} < t$, its contribution to the tie strength ($= a_{ij}(\tau_k^{ij})$) decreases every time step when there is no interaction. The number of time steps from time τ_k^{ij} to time t is $t - \tau_k^{ij}$, so the maximum factor by which the tie strength decreases is $e^{-\alpha(t-\tau_k^{ij})}$. Therefore, the first part of (11) can be interpreted as the 'minimum' tie strength at time t, i.e. the tie strength if at time t if there are no interactions between time τ_k^{ij} and time t. However, if there is an interaction, the tie strength that is already 'present' does not decrease.

Therefore, if interactions occur at times $\tau_0^{ij}, \ldots, \tau_n^{ij}$, the total contribution to the tie strength at time t of the interaction at time τ_k^{ij} is

$$a(\tau_k^{ij})e^{-\alpha(t-\tau_k^{ij}-(n-k))} = e^{-\alpha(t-\tau_k^{ij}-(n-k))}.$$

Now we solve the following equation:

$$e^{-\alpha(t-\tau_k^{ij}-(n-k))} = e^{-\alpha(t-\tau_k^{ij})} + x_k^{ij} \Rightarrow x_k^{ij} = e^{\alpha(t-\tau_k^{ij})}(e^{\alpha(n-k)} - 1).$$

So, the total tie strength of pair of nodes i and j at time t is the 'minimum' tie strength $\sum_{k=1}^{t} e^{-\alpha(t-k)} a_{ij}(k)$ increased by $\sum_{k=0}^{n} x_k^{ij} = b_{ij}(t)$. In matrix form:

$$S(t) = \sum_{k=1}^{t} e^{-\alpha(t-k)} A(k) + B(t).$$

□

We define the clustering coefficient of node k at time t to be

$$clust_k(t) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{s}_{ki}(t)\tilde{s}_{kj}(t)\tilde{s}_{ij}(t)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{s}_{ki}(t)\tilde{s}_{kj}(t)}$$

where N is the number of nodes and $\tilde{s}_{ij}(t)$ is the normalized tie strength of the edge between nodes i and j at time t. This normalization can be done in the following way:

$$\tilde{s}_{ij}(t) = \frac{s_{ij}(t)}{\beta},$$

where β is chosen large enough such that $\beta > s_{ij}(t)$ for all i,j,t. This normalization process is done to make sure that all values are between 0 and 1. The 'total' clustering coefficient of a network at time t is

$$clust(t) = \frac{\sum_{k=1}^{N} w_k(t) clust_k(t)}{\sum_{k}^{N} w_k(t)}$$

where $w_k(t)$ is the weighted degree of node k at time t:

$$w_k(t) = \sum_{i=1}^{N} s_{ik}(t).$$

Our choice of the definition of a weighted clustering coefficient is consistent with [2].

With these definitions, it is possible to make a theoretical analysis of the behaviour of the clustering coefficient for bursty interaction patterns. However, in this section we focus on simulations where interactions between nodes are simulated with various active and passive probabilities p and q. We run multiple simulations for a network with 4 nodes, where each pair of nodes has stationary interaction probability $\pi_1 = 0.03$. However, we vary the active and passive interaction probabilities p and q and check how these values influence the behaviour of the clustering coefficient.

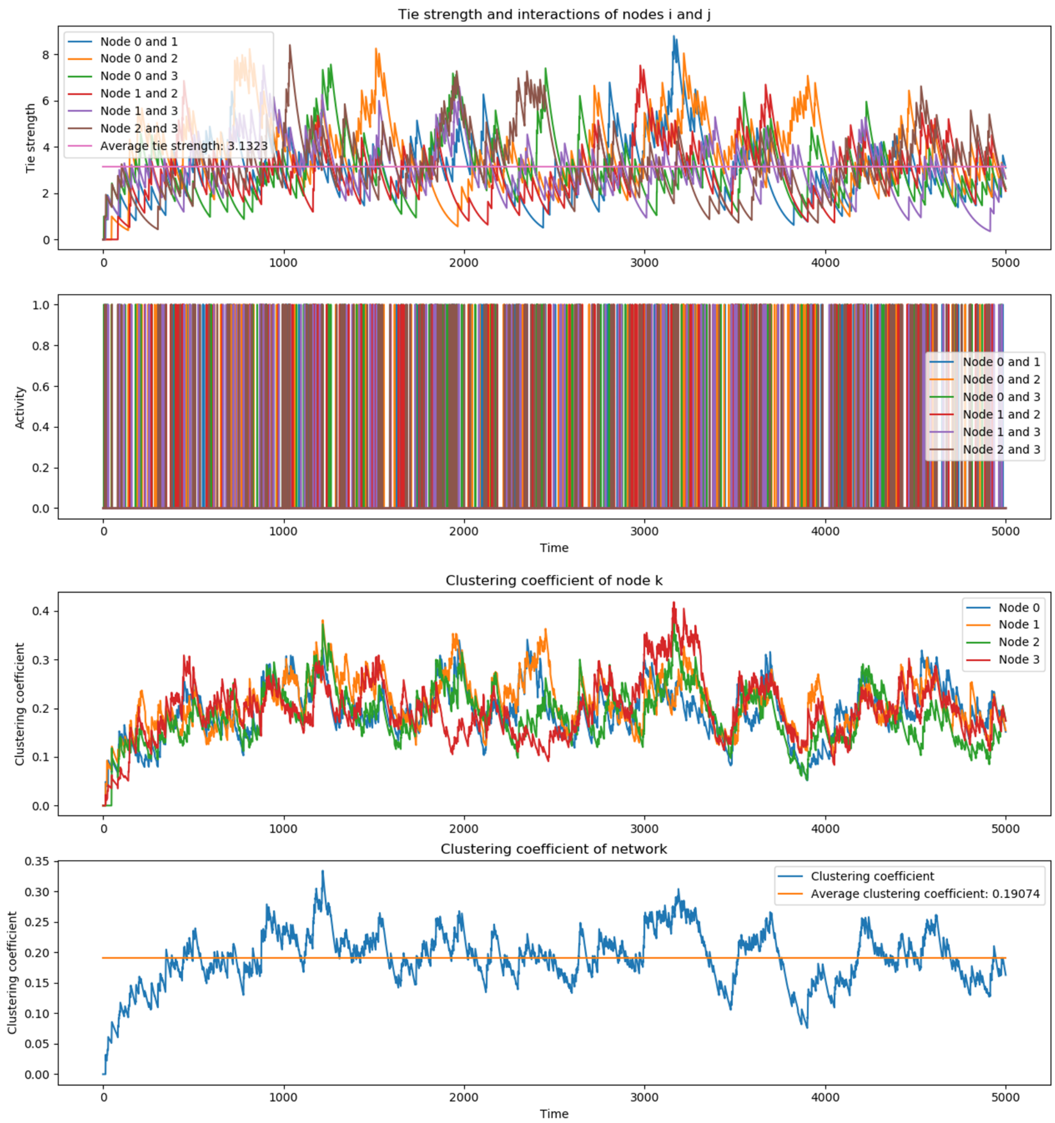

Figure 6: Tie strength, interactions, clustering coefficient per node and clustering coefficient of the network for a network with $N = 4$, $p = q = 0.03$, $\alpha = 0.01$ and $\beta = 150$.

First, we take $p = q = 0.03$, $\alpha = 0.01$ and $\beta = 150$ for all edges. This is a special case of the Markov model: we have $p = q$, so this is equivalent to the model of Zuo and Porter with $r = 0.03$. For convenience, all clustering coefficient values are multiplied by 15. The result can be seen in Figure 6. All individual clustering coefficients follow similar patterns, the mean network clustering coefficient is $c = 0.19074$ and the mean tie strength is $s = 3.1323$.

Now, for a much more bursty interaction pattern, we take $p = 0.95$ and $q = 0.0015$. All other parameter values are kept constant. The result can be seen in Figure 7. In this case, the burstiness of both individual and network clustering coefficients is much higher. Furthermore, the average network clustering coefficient is $c = 0.0405$, which is more than four times smaller than the clustering coefficient in Figure 6, while the average tie strength is slightly larger: $s = 3.6290$.

Table 3 shows some more results. The accompanying figures can be found in the appendix.

π_1	p	q	Average tie strength	Average clustering coefficient
0.03	0.03	0.03	3.1323	0.1907
0.03	0.6	0.012371134	3.1207	0.1579
0.03	0.8	0.006185567	3.5243	0.1144
0.03	0.9	0.003092784	3.6719	0.0942
0.03	0.95	0.001546392	3.629	0.0405
0.03	0.98	0.000618557	3.3391	0.0001

TABLE 3: Average clustering coefficient for different degrees of burstiness.

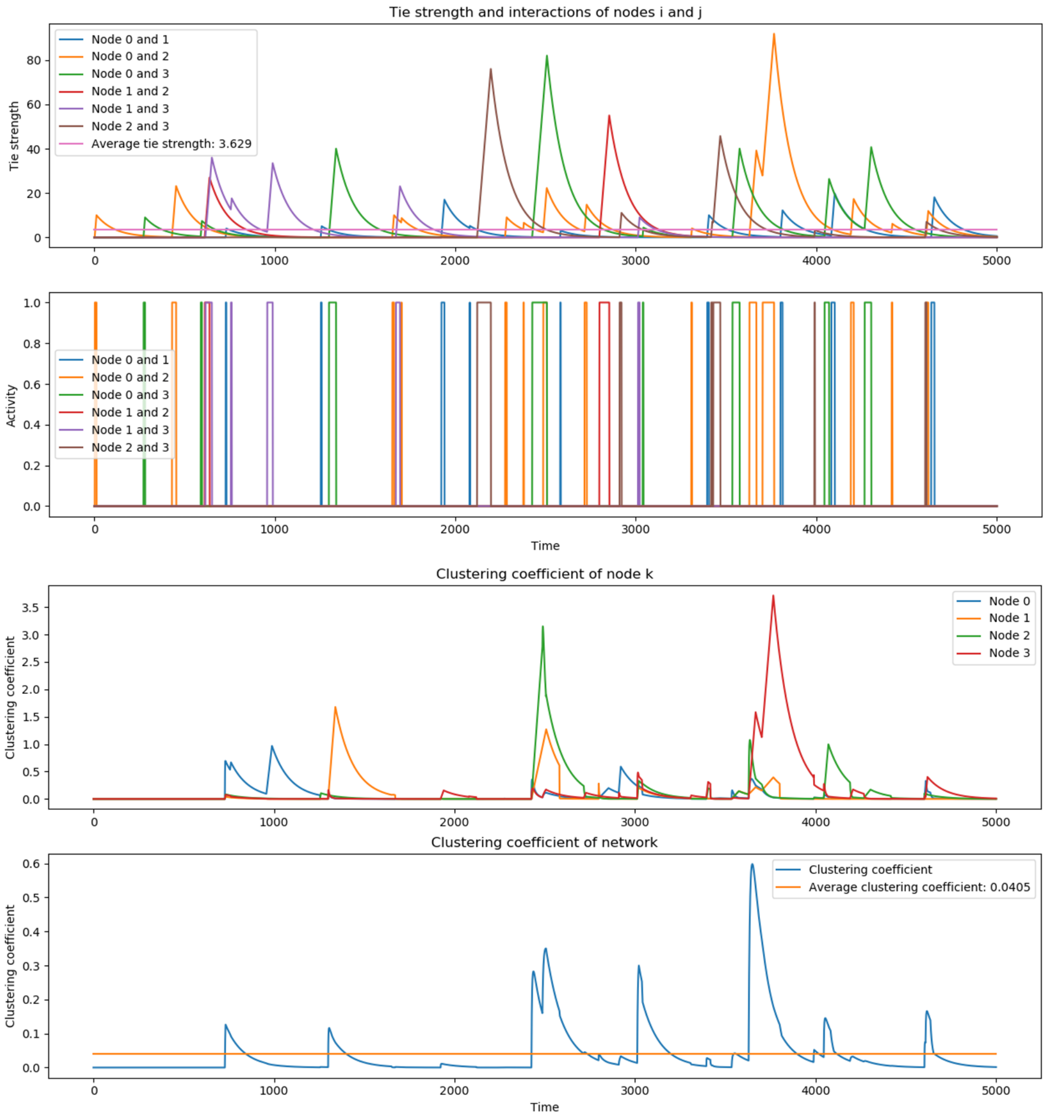

FIGURE 7: Tie strength, interactions, clustering coefficient per node and clustering coefficient of the network for a network with $N = 4$, $p = 0.95$, $q = 0.0015$, $\alpha = 0.01$ and $\beta = 150$.

4 Conclusions

In this paper, we developed a model that incorporates bursty interaction patterns. One of the main questions of this research was how the tie strength behaves on the long term and if the expected value of tie strength in a Markov model with $p > q$ differs from the expected value of tie strength in a model with one interaction probability. We found that for equal stationary interaction probability, the long-term tie strength does not depend on the burstiness of interaction patterns. However, we also proved that the expected value of tie strength at the end of an interaction period does differ: when interaction patterns are bursty, the tie strength grows much larger than if there is one interaction probability. This leads to the fact that the variance of the tie strength at the end of an interaction period is higher for higher degrees of burstiness.

We also investigated the behaviour of a weighted clustering coefficient for interaction patterns with different degrees of burstiness. Here we found that the mean clustering coefficient decreases as interaction patterns get more bursty, while the mean tie strength does not change. This could be explained by the fact that the clustering coefficient counts the number of 'triplets': for a high clustering coefficient, it is necessary to have sets of three ties with a high tie strength. For bursty interaction patterns, the variability of tie strength is higher and therefore the probability that one of the three ties of a 'triplet' has a low tie strength is higher than for non-bursty interaction patterns, where values of tie strength are less variable. One tie with a tie strength close to zero is sufficient to make the value of its triplet close to zero, and since the probability that a tie strength is close to zero is higher for bursty interaction patterns, there will be more triplets with a value close to zero. This results in a smaller clustering coefficient when interaction patterns are bursty.

It is clear that the way one models tie strength influences the clustering coefficient, even if the average tie strength remains constant. In this paper, we only did simulations of interaction patterns to calculate the clustering coefficient. However, in earlier sections we provided all definitions to make a theoretical analysis of the effect of bursty interaction patterns on the clustering coefficient. These definitions can be used in further research into the influence of bursty interaction patterns on tie strength and the clustering coefficient.

References

[1] Walid Ahmad, Mason A Porter, and Mariano Beguerisse-Díaz. Tie-decay temporal networks in continuous time and eigenvector-based centralities. *arXiv preprint arXiv:1805.00193*, 2018.

[2] Gabriela Kalna and Desmond Higham. A clustering coefficient for weighted networks, with application to gene expression data. *AI Commun.*, 20:263–271, 01 2007.

[3] Xinzhe Zuo and Mason A. Porter. Models of continuous-time networks with tie decay, diffusion, and convection. *CoRR*, abs/1906.09394, 2019.

Appendix A: clustering coefficient for different degrees of burstiness

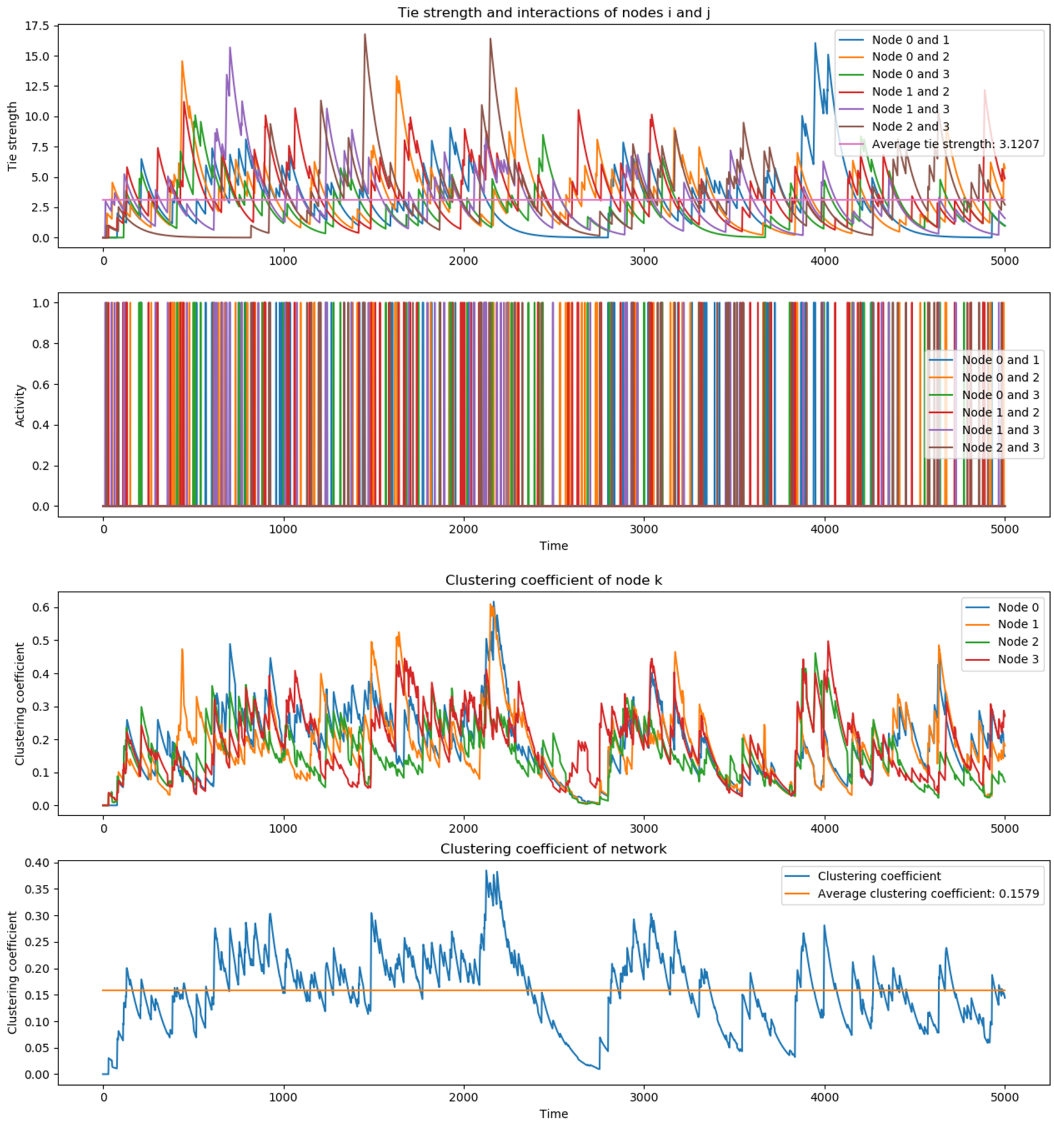

FIGURE 8: Tie strength, interactions, clustering coefficient per node and clustering coefficient of the network for a network with $N = 4$, $p = 0.6$, $q = 0.0124$, $\alpha = 0.01$ and $\beta = 150$.

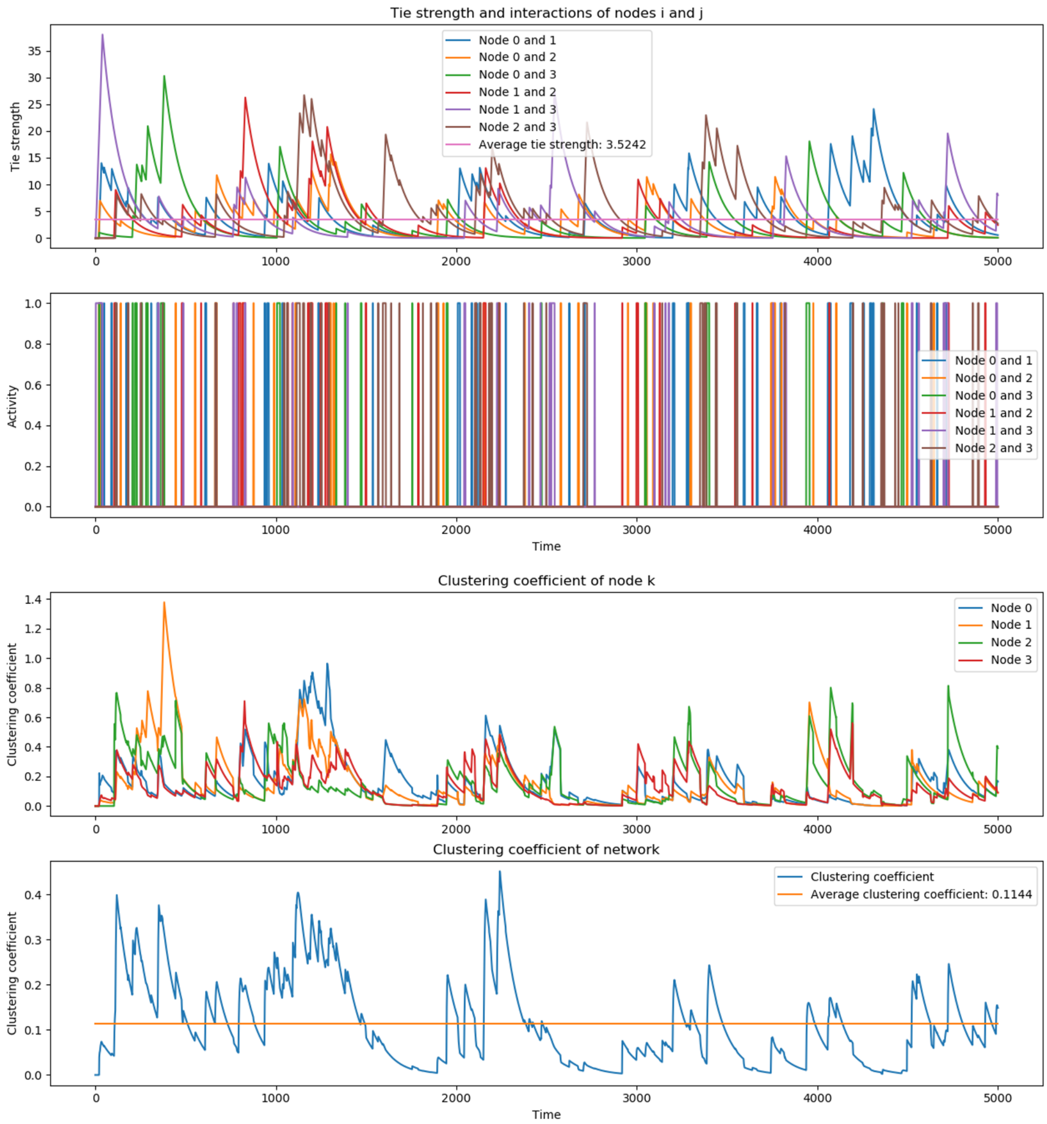

FIGURE 9: Tie strength, interactions, clustering coefficient per node and clustering coefficient of the network for a network with $N = 4$, $p = 0.8$, $q = 0.0062$, $\alpha = 0.01$ and $\beta = 150$.

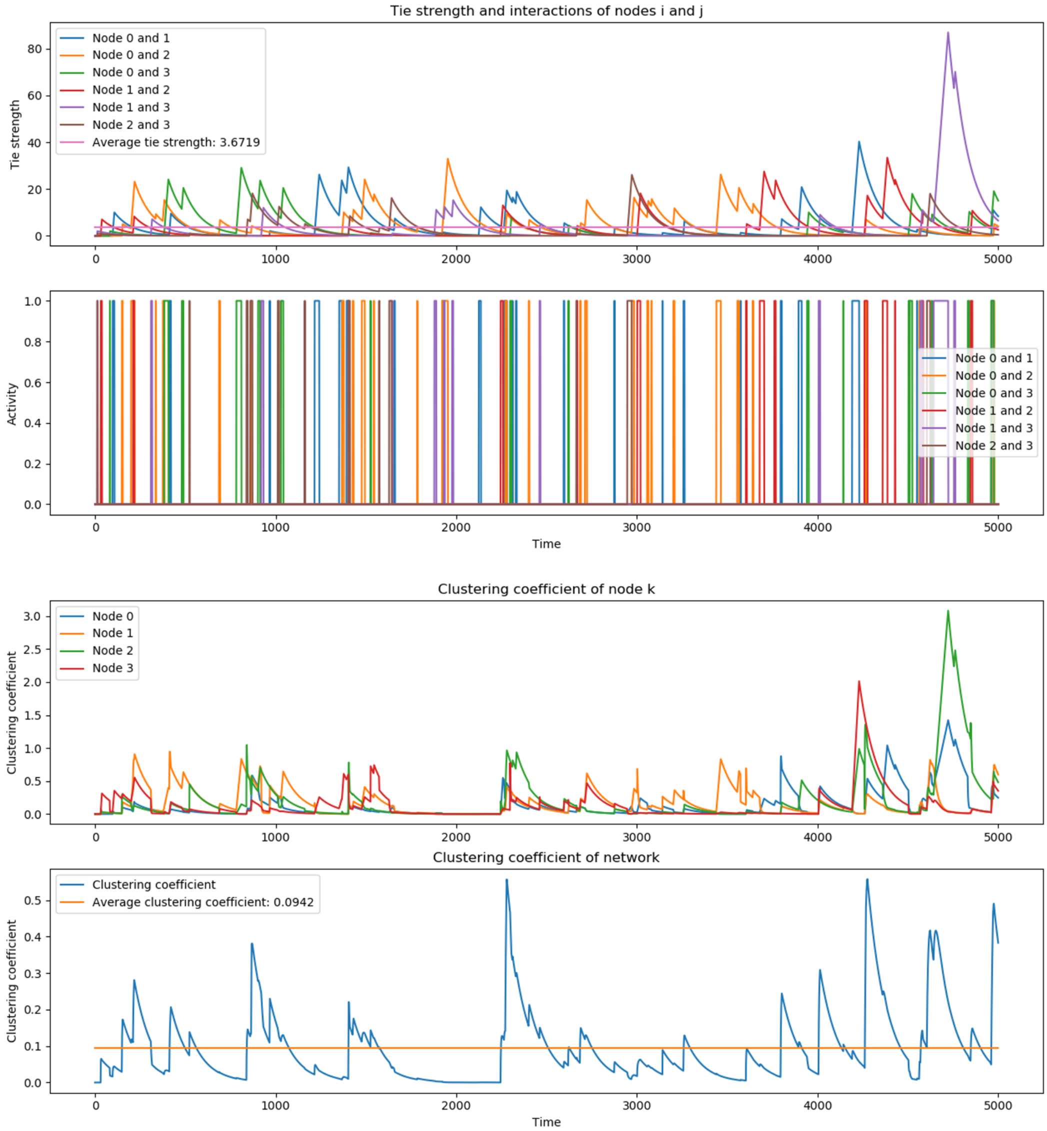

FIGURE 10: Tie strength, interactions, clustering coefficient per node and clustering coefficient of the network for a network with $N = 4$, $p = 0.9$, $q = 0.0031$, $\alpha = 0.01$ and $\beta = 150$.

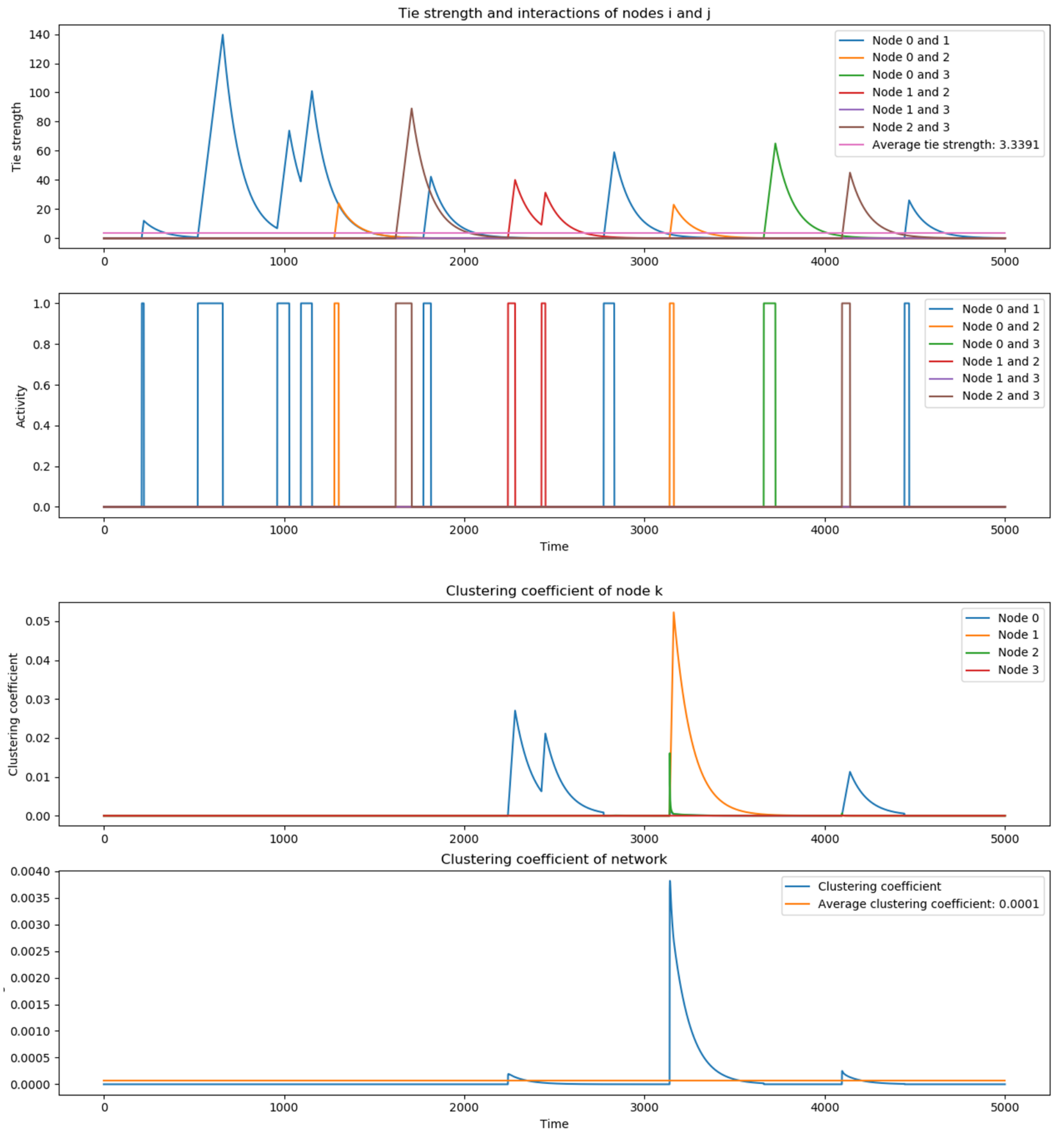

FIGURE 11: Tie strength, interactions, clustering coefficient per node and clustering coefficient of the network for a network with $N = 4$, $p = 0.98$, $q = 0.0006$, $\alpha = 0.01$ and $\beta = 150$.