

BSc Thesis Applied Mathematics

Influence of popularity on tie strength in tie-decay networks

Jop Zwienenberg

Supervisor: Dr. C. Stegehuis

June 26, 2020

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

UNIVERSITY OF TWENTE.

Preface

This paper was written to fulfill the graduation requirements of the bachelor Applied Mathematics at the University of Twente. The research was performed from April 20, 2020 up to June 30, 2020. I would like to thank my supervisor Dr. C. Stegehuis for her essential support and guidance during the entire time of my bachelor thesis.

Influence of popularity on tie strength in tie-decay networks

Jop Zwienenberg

June 26, 2020

Abstract

The concept of the strength can be used to analyze interactions. The tie strength describes the strength of your friendship with one of your friends. It increases when an interaction takes place, and then decays continuously over time if no new interaction takes place. This concept can be very useful to find the most important connections in a social network. This paper investigates the influence of popularity on the tie strength in such tie-decay networks. To model popularity, we use an interaction probability p_{ij} between persons i and j at every time step, which depends on the popularities of persons i and j, the number of people in the network and the average popularity among all people in the network. Consequently, the expected value of the long-term tie strength between person i and a randomly chosen other person given i's popularity can be calculated. For this it is assumed that the popularity of all the other people in the network follows a power law distribution. The results of this analysis are used to analyze real life data of interactions: face-to-face contact data during 20-second time intervals in different contexts. By analyzing this data, the distributions of real-world tie strength are determined. Then, by using several network data statistics with, among other things, the skewness, the differences between these distributions can be explained.

Keywords: tie strength, popularity, interaction probability, people, friends, data, distribution, temporal networks

1 Introduction

In most networks, interactions appear and disappear over time. Over the past decade more and more people became interested in the complex 'connectedness' of modern society. This connectedness is found in the growth of the Internet and the Web; there is a growing amount of ways to communicate online through instant message networks such as WhatsApp. The fast spreading of news and information as well as epidemics and financial crises around the world are other illustrations of this connectedness. Motivated by these developments in the world, there has been an increasing interest to investigate highly connected networks.

The interactions in these networks can be analyzed using 'tie strength'. Usually, interactions in social networks follow 'bursty' patterns. That means that you may interact very frequently during a small time period, but it is also possible that the time between two interactions is very long. This is the reason why it is good to have a measure for the connectedness between two people. In this paper, the tie strength describes the strength of your friendship with one of your friends. When you interact with one of your friends, the tie strength increases. When you do not interact for a certain time period, the tie strength of your friendship decreases. So the more interactions between two people, the higher the tie strength between them. In this way the concept of tie strength can be useful to find the most important connections in a social network.

In a network some people are more popular than others. We will define the following definition for popularity: "In sociology, popularity is how much a person, idea, place, item or other concept is either liked or accorded status by other people. Liking can be due to reciprocal liking, interpersonal attraction, and similar factors. Social status can be due to dominance, superiority, and similar factors" [1]. For example, a friendly person may be considered likable and therefore more popular than another person, and a rich person may be considered superior and therefore more popular than another person. In general, at a certain time, a popular person is more likely to interact with another person than a less popular person. This is because, according to the previously mentioned definition of popularity, a popular person is more liked or accorded status by other people than a less popular person than with a less popular person. In these assumptions lies the basis of our models.

In earlier models for the strength [2, 3, 4], these differences in popularity between people are not considered: it is assumed that every person has an equal popularity. This is in contrast to the differences in popularity between people we will assume in our models. We will model this by assuming that popularities are distributed according to a power law. A power law is very unequally distributed: a relatively small fraction of people has a large amount of popularity, which is a good representation of reality.

In this paper we investigate the mathematical properties of tie strength when some people in the network are more popular than others. Consequently, the major research question is: What happens to the tie strength when some people in the network are more popular than others?

In order to model the increase and decrease in tie strength, we use the tie-decay model of Ahmad [2] as our first model: during each time step, two people can have an interaction or not. This model can be found in Subsection 2.1. If there is an interaction, then the tie strength increases by 1. If there is no interaction, assuming α is the decay parameter, then the tie strength decays exponentially. We will look at the effect of different values of α on the tie strength values in this model. The second model we will use is the tie-decay model of Jin [3] (Subsection 2.2): if there is an interaction, then the tie strength goes up to 1. If there is no interaction, then the tie strength goes up to 1. If there is no interaction, then the tie strength goes up to 1. If

We then analyze the long-term expected tie strength given popularity at time step t (for example: t = 2 means at the end of the second time step from t = 1 to t = 2) in both models by expanding these models to the situation that some people in the network are more popular than others. For this reason, the long-term expected tie strength derived in the work of Zuo [4] will include the probability of an interaction between two persons i and j (p_{ij}) at time step t. We will not use an equal and stationary interaction probability for all people, but one that depends on popularity and is unequal and stationary. This is taken from the work of van der Hofstad [5].

The results of this analysis are used to analyze real life face-to-face contact data, during 20-second time intervals, in different contexts in Section 3. This data is obtained through

the SocioPatterns sensing platform [6]. Then, using the network data statistics for these interaction networks on the Network Data Repository [7], the shape of the histogram plots, in which the distribution of the tie strength is represented, can be explained.

To conclude, the goal of this paper is to investigate the influence of popularity on the long-term expected tie strength in tie-decay networks. Using the resulting relation, conclusions about the distribution of tie strength are drawn from real life interaction data.

2 Behaviour of tie strength

In this section we will investigate the influence of popularity on the long-term expected tie strength in tie-decay networks. The underlying time in these networks is continuous and non-negative, which is measured in small increments δt . The tie strength between two people depends on the history of interactions. That is why the tie strength between each pair of people at t = 0 is considered as being 0. It increases when an interaction takes place, and then decreases over time if no new interaction takes place. We assume that, during a single time step, two people either have one interaction or no interaction. We suppose that the growth and decay pattern of each pair of people is independent of all other pairs, so we independently consider each pair during each time step.

2.1 Tie strength in the tie-decay model of Ahmad

As mentioned in the introduction, we use the tie-decay model of Ahmad [2] for our first model. We will denote this model as model 1 for future reference. During a time step of length δt , if there is an interaction between two people, which occurs with a varying probability $0 \leq p_{ij} \leq 1$ (more about the exact value later), then the tie strength of the friendship between these two people increases by 1. If there is no interaction, which occurs with complementary probability $1 - p_{ij}$, then the tie strength between them decays by a factor of $e^{-\alpha \delta t}$.

It is important to examine multiple values of $\alpha \delta t$ [8]. In this paper, we focus on the decay and boosting behavior of ties between two people. Therefore, it is more meaningful to examine the product of the time step (δt) and the decay rate (α) , instead of studying them individually. For simplicity, we take $\delta t = 1$ in this model, which is in line with [4]. Consequently, time will be discrete and non-negative. Later, in Subsection 2.1.1, we will see the effect of different values of α (and thus different values of $\alpha \delta t$) on the tie strength values for different popularities in the theoretical model 1. In model 2 this effect will be similar. In the next section about real life interaction data in model 1 (Section 3), we will see the effect of these different values on the distribution of the tie strength. We then try to explain this effect by using the effect of different values of α on the tie strength in model 1.

We can rewrite the aforementioned model in a single expression so that it enables us to calculate the tie strength at a certain time step t:

Definition 2.1. If z_t is a Bernoulli random variable (so $z_t = 1$ if there is an interaction during time step t and $z_t = 0$ if there is no interaction) and assuming the length of a time step $\delta t = 1$, then we can write the tie strength s_t between two persons i and j at a certain time step t as:

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

In Figure 1, we show an illustrative example of the model's tie-decay dynamics [4]. The tie strength between two people increases by 1 when there is an interaction during a time step, and it decays exponentially when there is no interaction:



FIGURE 1: An illustration of dynamics in the tie-decay model of Ahmad [2]. In this simulation, we have N = 1000 people, a decay rate of $\alpha = 0.01$, a homogeneous interaction probability of p = 0.003, and t = 1000 time steps. The vertical axis shows the tie strength between one pair of people. Six interactions occur between these two people.

In the work of Zuo [4] this same model is used and subsequently the long-term expected tie strength is calculated for one interaction probability (p_{ij}) at time step t. This leads to the following theorem:

Theorem 2.2 (Equation (3) of [4]). The long-term expected tie strength between two persons *i* and *j* with interaction probability p_{ij} and decay parameter α is given by:

$$E[s] := \lim_{t \to \infty} E[s_t] = E[s]e^{-\alpha}(1 - p_{ij}) + (E(s) + 1)p_{ij} = \frac{p_{ij}}{(1 - e^{-\alpha})(1 - p_{ij})}.$$
 (2)

Proof. Note that we use p_{ij} as probability of success (interaction probability) of the Bernoulli random variable z_t . As a consequence, by the law of total expectation and using equation (1):

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

= $E[z_t] + E[e^{-\alpha(1-z_t)}s_{t-1}|z_t = 1]P(z_t = 1) + E[e^{-\alpha(1-z_t)}s_{t-1}|z_t = 0]P(z_t = 0)$
= $p_{ij} + E[s_{t-1}]p_{ij} + e^{-\alpha}E[s_{t-1}](1-p_{ij}) = p_{ij}(1+E[s_{t-1}]) + (1-p_{ij})e^{-\alpha}E[s_{t-1}].$

Now in the paper of Zuo [4] it is verified that we reach a stationary state as $t \to \infty$. In this state it holds that $\lim_{t\to\infty} E[s_{t-1}] = \lim_{t\to\infty} E[s_t] = E[s]$. Taking $t \to \infty$, we obtain the following:

$$E[s] := \lim_{t \to \infty} E[s_t] = p_{ij}(1 + \lim_{t \to \infty} E[s_{t-1}]) + (1 - p_{ij})e^{-\alpha}\lim_{t \to \infty} E[s_{t-1}]$$
$$= p_{ij}(1 + E[s]) + (1 - p_{ij})e^{-\alpha}E[s].$$

From here we isolate E[s] and obtain:

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

2.1.1 Long-term expected tie strength given popularity

We can now obtain the long-term expected tie strength given popularity, which depends on the probability of an interaction between two persons i and j (p_{ij}) at every time step t. The definition of p_{ij} in the paper of van der Hofstad [5] is now used, which is in the generalized random graph model. Here $h = (h_i)_{i \in [n]}$ are the node weights of nodes $[n] = \{1, 2, ..., n\}$ and $\sum_{i \in [n]} h_i$ is the total weight of all nodes. Then the probability that there is an edge between two nodes i and j equals:

$$p_{ij} = \frac{h_i h_j}{\sum_{i \in [n]} h_i + h_i h_j}.$$
(3)

The interaction probability is proportional to the product of the given weights of nodes i and j: $h_i h_j$. This implies the following: the higher the product of the node weights of a certain pair of nodes, the higher the probability that there is an edge between these nodes. In [5] they have the following explanation for using equation (3) to determine this probability: nodes with high weights have a higher probability to have many neighbors than nodes with small weights. Nodes with extremely high weights could act as the 'hubs' [5] observed in many real-world networks. Hubs can be understood as centers of these real-world networks. Based on the probability in equation (3), we will use the following formula for our model. It is important to note that H_i is a random variable and h_i a realization of this random variable:

Definition 2.3. If every person *i* in the network has a popularity $H_i = h_i$, *N* is the number of people in the network and $\langle h \rangle$ is the average popularity among all people in the network, then the probability that two persons *i* and *j* interact at every time step is:

$$p_{ij} = \frac{h_i h_j}{N \langle h \rangle + h_i h_j}.$$
(4)

In order to be able to obtain an illustration of the long-term expected tie strength given $H_i = h_i$ and $H_j = h_j$ against the popularity h_i , which we will derive later in equation (7), we need values for α , h_j , $\langle h \rangle$ and N. The values of h_j and $\langle h \rangle$ will be based on the assumption that the popularity of all the people in the network other than person *i* follows a power-law distribution. A power law is very unequally distributed: a relatively small fraction of people has a large amount of popularity. Below we have the mathematical definition of a power-law distribution using [9]. To be more precise, $p(h_j)$ is here the probability density function (pdf) of the distribution of h_j for $h_j \geq h_{j_{min}}$:

$$p(h_j) = \frac{\beta - 1}{h_{j_{min}}} \left(\frac{h_j}{h_{j_{min}}}\right)^{-\beta}.$$
(5)

Consequently, the first moment of h_i [9] is defined as:

$$\langle h \rangle = h_{j_{min}} \left(\frac{\beta - 1}{\beta - 2} \right). \tag{6}$$

This is only defined for $\beta > 2$. When $1 < \beta < 2$, the mean as well as the variance are infinite and when $2 < \beta < 3$, the mean of the popularity is finite, but the variance is infinite, which is the case for many real-world networks. When $\beta > 3$, the mean and the variance are both finite. This is the reason we will choose $\beta = 2.5$ in case of performing simulations. Now $h_{j_{min}}$ has to be chosen: the minimum popularity of person j. We will choose $h_{j_{min}} = 1$, so that equations (5) and (6) become easier to read. At this point, we have enough knowledge to derive the long-term expected tie strength given popularities $H_i = h_i$ and $H_j = h_j$ in equation (7) and given only popularity $H_i = h_i$ in equation (8): **Theorem 2.4.** The long-term expected tie strength between two persons i and j given their popularities $H_i = h_i$ and $H_j = h_j$, using equations (2) and (4) is:

$$E[s|H_i = h_i, H_j = h_j] = \frac{h_i h_j}{(1 - e^{-\alpha})N\langle h \rangle}.$$
(7)

Then, consequently, the long-term expected tie strength between a person i and a randomly chosen other person given only popularity $H_i = h_i$ becomes:

$$E[s|H_i = h_i] = \frac{3h_i}{(1 - e^{-\alpha})N\langle h\rangle}.$$
(8)

Proof. Using Equation 2 in combination with Equation 4, we obtain the following:

$$\begin{split} E[s|H_i = h_i, H_j = h_j] &= \frac{\frac{h_i h_j}{N \langle h \rangle + h_i h_j}}{(1 - e^{-\alpha})(1 - \frac{h_i h_j}{N \langle h \rangle + h_i h_j})} \\ &= \frac{h_i h_j}{(1 - e^{-\alpha})(N \langle h \rangle + h_i h_j - h_i h_j)} = \frac{h_i h_j}{(1 - e^{-\alpha})N \langle h \rangle}. \end{split}$$

Then, using the law of total expectation and the formula for expectation: $E[f(h_j)] = \int_1^\infty f(h_j)p(h_j)dh_j$, with $p(h_j)$ from equation (5) with $\beta = 2.5$ and $h_{j_{min}} = 1$, we acquire the following:

$$E[s|H_i = h_i] = E\left[E[s|H_i = h_i, H_j = h_j]\right] = E\left[\frac{h_i h_j}{(1 - e^{-\alpha})N\langle h\rangle}\right]$$
$$= \frac{h_i}{(1 - e^{-\alpha})N\langle h\rangle} E[h_j] = \frac{h_i}{(1 - e^{-\alpha})N\langle h\rangle} \int_1^\infty h_j \frac{1.5}{h_j^{2.5}} dh_j$$
$$= \frac{1.5h_i}{(1 - e^{-\alpha})N\langle h\rangle} \int_1^\infty h_j^{-1.5} dh_j = \frac{-3h_i}{(1 - e^{-\alpha})N\langle h\rangle} [h_j^{-0.5}]_1^\infty = \frac{3h_i}{(1 - e^{-\alpha})N\langle h\rangle}.$$

Now, values for h_j can be obtained by generating power-law distributed random numbers. The formula for this is taken from the work of Clauset [10]. Here r is uniformly distributed in the interval $0 \le r < 1$:

$$h_j = h_{j_{min}} (1-r)^{-\frac{1}{\beta-1}}.$$
(9)

In the following plots we will use equations (6) and (9) to plot equations (7) and (8) against the popularity of person i (h_i). In line with [4], we will use $\alpha = 0.01$ in both equations. We will generate values for h_j 10 times according to the power-law distribution. In other words, the value of r is determined 10 times. In Figures 2a and 2b the different values of h_j can be distinguished by different line colors. For each h_i in the range from 1 to 100, the corresponding long-term expected tie strength is determined for respectively N = 100 and N = 1000. We expect that the higher the popularity of person i, the higher the expected long-term tie strength. And for N = 1000 (so 1000 people in the network), in general, the expected long-term tie strength is anticipated to be lower at the same popularity compared with the situation N = 100. More people in the network implies a lower chance of interacting with a specific person j according to equation (4), so also a lower expected long-term tie strength between persons i and j. The results of plotting equation (7) against the popularity of person i, which are simulated in MATLAB, can be seen in Figure 2a and 2b:



FIGURE 2: Expected long-term tie strength given popularity of person i (h_i) and person j (h_j) as a function of popularity of person i for 100 people in Figure 2a and 1000 people in Figure 2b for $\alpha = 0.01$.

Now we will also plot the long-term expected tie strength given only $H_i = h_i$ in equation (8) against the popularity h_i . We expect this will result in only one increasing line, since now the expectation does not depend on the random variable h_j (as in equation (9)) anymore, as can be seen in equation (8).



FIGURE 3: Expected long-term tie strength given popularity of person i (h_i) as a function of popularity of person i for 100 people in Figure 3a and for 1000 people in Figure 3b for $\alpha = 0.01$.

These results are exactly in line with our expectations: the more popular a person i, the higher the expected long-term tie strength with a random other person j in the network. Also, the expected long-term tie strength for N = 1000 is indeed, in general, lower at the same popularity compared with the situation N = 100. Mathematically, this can be explained as follows: if we have a look back at equation (7) and (8), then we see that a higher value of N implies a lower value for the expected tie strength.

As a side note, we will now look at what impact a lower and a higher value of α has on the plots of the long-term tie strength against popularity in Figures 2a and 3a for N = 100. We do not consider N = 1000, since the effect of different values of α on the tie strength values for different popularities will be similar. First, we plot the long-term tie strength against popularity for a lower value of α ($\alpha = 0.0001$):



(A) Given popularity of person $i(h_i)$ and $j(h_j)$ (B) Given popularity of person $i(h_i)$

FIGURE 4: Expected long-term tie strength given popularity of person i (h_i) and person j (h_j) in Figure 4a and given popularity of person i (h_i) in Figure 4b as a function of popularity of person i for N = 100 and $\alpha = 0.0001$.

If we compare Figure 2a with Figure 4a, then we see that the positive slope of the line increases in Figure 4a as α decreases. The same goes for Figure 3a compared with Figure 4b. This implies that the ratio between the long-term tie strength values of relatively popular people and less popular people is way higher in the case of a lower value of α , due to the higher long-term tie strength values of relatively popular people. Mathematically, this can be explained as follows: if we have a look back at equations (7) and (8), then we see that a lower value of α implies a higher exponent of $e^{-\alpha(1-z_t)}$, thus a higher value of $e^{-\alpha(1-z_t)}$, which implies a lower decrease in tie strength if two persons *i* and *j* do not have an interaction in a certain time interval. This means that the long-term expected tie strength attains higher values and that the positive slope of the line is larger as well, since the line is linear. The line represents the long-term expected tie strength at different values for h_i .

Second, we plot the long-term tie strength against popularity for a higher value of α ($\alpha = 1$):



(A) Given popularity of person $i(h_i)$ and $j(h_i)$ (B) Given popularity of person $i(h_i)$

FIGURE 5: Expected long-term tie strength given popularity of person i (h_i) and person j (h_j) in Figure 5a and given popularity of person i (h_i) in Figure 5b as a function of popularity of person i for N = 100 and $\alpha = 1$.

If we compare Figure 2a with Figure 5a, then we see that the positive slope of the line decreases in Figure 5a as α increases. The same goes for Figure 3a compared with Figure 5b. This implies that the ratio between the long-term tie strength values of relatively popular people and less popular people is way lower in the case of a higher α . Looking back at equations (7) and (8), we see that a lower value of α implies lower values for the long-term expected tie strength and thus a smaller positive slope of the line.

2.2 Tie strength in the tie-decay model of Jin

We will now have a look at a different tie-decay model than the model of Ahmad [2], namely the model of Jin [3] for our second model. We will denote this model as model 2 for future reference. We will see if it results in a similar plot of the expected tie strength against popularity. During a time step of length δt , if there is an interaction between two people, which occurs with a varying probability p_{ij} , then the tie strength of the friendship between these two people goes up to 1, as opposed to the model of Ahmad [2] in equation (1) in which the tie strength increased by 1 after an interaction. If there is no interaction, which occurs with complementary probability $1 - p_{ij}$, then the tie strength still decays by a factor of $e^{-\alpha \delta t}$.

We also take $\delta t = 1$ in this model. Consequently, time will be discrete and non-negative. We can rewrite the aforementioned model in a single expression so that it enables us to calculate the tie strength at a certain time step t:

Definition 2.5. If z_t is a Bernoulli random variable (so $z_t = 1$ if there is an interaction during time step t and $z_t = 0$ if there is no interaction) and assuming the length of a time step $\delta t = 1$, then we can write the tie strength s_t between two persons i and j at a certain time step t as:

$$s_t = z_t + (1 - z_t)e^{-\alpha(1 - z_t)}s_{t-1}.$$
(10)

In Figure 6, we show an illustrative example of the model's tie-decay dynamics [4]. The tie strength between two people resets to 1 when there is an interaction during a time step, and it decays exponentially when there is no interaction:



FIGURE 6: An illustration of dynamics in the tie-decay model of Jin [3]. In this simulation, we have N = 1000 people, a decay rate of $\alpha = 0.01$, a homogeneous interaction probability of p = 0.003, and t = 1000 time steps. The vertical axis shows the tie strength between one pair of people. Four interactions occur between these two people.

In the work of Zuo [4] this same model is used and subsequently the long-term expected tie strength is calculated for one interaction probability at time step t. This leads to the following theorem:

Theorem 2.6 (Equation (5) of [4]). The long-term expected tie strength between two persons *i* and *j* with interaction probability p_{ij} and decay parameter α is given by:

$$E[s] := \lim_{t \to \infty} E[s_t] = p_{ij} + E[s_{t-1}]e^{-\alpha}(1 - p_{ij}) = \frac{p_{ij}}{1 - e^{-\alpha}(1 - p_{ij})}.$$
(11)

Proof. Note that we use p_{ij} as probability of success (interaction probability) of the Bernoulli random variable z_t . As a consequence, by the law of total expectation and using equation (10):

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

= $E[z_t] + E[(1 - z_t)e^{-\alpha(1 - z_t)}s_{t-1}|z_t = 1]P(z_t = 1)$
+ $E[(1 - z_t)e^{-\alpha(1 - z_t)}s_{t-1}|z_t = 0]P(z_t = 0) = p_{ij} + e^{-\alpha}E[s_{t-1}](1 - p_{ij}).$

Now in the paper of Zuo [4] it is verified that we reach a stationary state as $t \to \infty$. In this state it holds that $\lim_{t\to\infty} E[s_{t-1}] = \lim_{t\to\infty} E[s_t] = E[s]$. Taking $t \to \infty$, we obtain the following:

$$E[s] := \lim_{t \to \infty} E[s_t] = p_{ij} + e^{-\alpha} \lim_{t \to \infty} E[s_{t-1}](1 - p_{ij})$$

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

From here we isolate E[s] and obtain:

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

2.2.1 Long-term expected tie strength given popularity

In order to be able to obtain an illustration of the long-term expected tie strength given $H_i = h_i$ in equation (13) against the popularity h_i , we will use equations (5), (6) and (9) again. For this we first need to obtain the long-term expected tie strength given popularities $H_i = h_i$ and $H_j = h_j$, which includes the probability of an interaction between two persons i and j (p_{ij}) at time step t. For this we again use equation (4).

Theorem 2.7. The long-term expected tie strength between two persons i and j given their popularities $H_i = h_i$ and $H_j = h_j$, using equations (11) and (4) is:

$$E[s|H_i = h_i, H_j = h_j] = \frac{h_i h_j}{N \langle h \rangle (1 - e^{-\alpha}) + h_i h_j}.$$
(12)

Then, consequently, the long-term expected tie strength between person i and a randomly chosen other person given only popularity $H_i = h_i$ becomes:

$$E[s|H_i = h_i] = -\frac{3h_i \left(\sqrt{\frac{h_i}{N\langle h \rangle (1 - e^{-\alpha})}} \arctan\left(\sqrt{\frac{N\langle h \rangle (1 - e^{-\alpha})}{h_i}}\right) - 1\right)}{N\langle h \rangle (1 - e^{-\alpha})}.$$
(13)

Proof. Using Equation 11 in combination with Equation 4, we obtain the following:

$$E[s|H_i = h_i, H_j = h_j] = \frac{\frac{h_i h_j}{N \langle h \rangle + h_i h_j}}{1 - e^{-\alpha} (1 - \frac{h_i h_j}{N \langle h \rangle + h_i h_j})}$$
$$= \frac{h_i h_j}{N \langle h \rangle + h_i h_j - e^{-\alpha} (N \langle h \rangle + h_i h_j - h_i h_j)} = \frac{h_i h_j}{N \langle h \rangle (1 - e^{-\alpha}) + h_i h_j}.$$

Then, using the law of total expectation and the formula for expectation: $E[f(h_j)] = \int_1^\infty f(h_j)p(h_j)dh_j$, with $p(h_j)$ from equation (5) with $\beta = 2.5$ and $h_{j_{min}} = 1$, we acquire the following:

$$E[s|H_{i} = h_{i}] = E\left[E[s|H_{i} = h_{i}, H_{j} = h_{j}]\right] = E\left[\frac{h_{i}h_{j}}{N\langle h\rangle(1 - e^{-\alpha}) + h_{i}h_{j}}\right]$$
$$= h_{i}E\left[\frac{h_{j}}{N\langle h\rangle(1 - e^{-\alpha}) + h_{i}h_{j}}\right] = h_{i}\int_{1}^{\infty}\frac{h_{j}}{N\langle h\rangle(1 - e^{-\alpha}) + h_{i}h_{j}}\frac{1.5}{h_{j}^{2.5}}dh_{j}$$
$$= 1.5h_{i}\int_{1}^{\infty}\frac{1}{N\langle h\rangle(1 - e^{-\alpha})h_{j}^{1.5} + h_{i}h_{j}^{2.5}}dh_{j}.$$

Now substitute $b = N \langle h \rangle (1 - e^{-\alpha})$, since it is a constant. Then we acquire:

$$E[s|H_i = h_i] = \frac{3h_i}{2} \int_1^\infty \frac{1}{h_i h_j^{2.5} + bh_j^{1.5}} dh_j$$

By performing the substitution $u = \sqrt{h_j}$ (so $\frac{du}{dh_j} = \frac{1}{2\sqrt{h_j}}$) we acquire:

$$E[s|H_i = h_i] = 3h_i \int_1^\infty \frac{u}{h_i u^5 + bu^3} du = 3h_i \int_1^\infty \frac{1}{u^2(h_i u^2 + b)} du.$$

Now, by performing partial fraction decomposition:

$$E[s|H_i = h_i] = 3h_i \int_1^\infty \left(\frac{1}{bu^2} - \frac{h_i}{b(h_i u^2 + b)}\right) du$$
$$= \frac{3h_i}{b} \int_1^\infty \frac{1}{u^2} du - \frac{3h_i^2}{b} \int_1^\infty \frac{1}{h_i u^2 + b} du.$$

By performing the substitution $v = \frac{\sqrt{h_i}u}{\sqrt{b}}$ (so $\frac{dv}{du} = \frac{\sqrt{h_i}}{\sqrt{b}}$) we acquire:

$$\begin{split} E[s|H_i = h_i] &= \frac{3h_i}{b} \left[-\frac{1}{u} \right]_1^{\infty} - \frac{3h_i^2}{b} \int_{\sqrt{\frac{h_i}{b}}}^{\infty} \frac{\sqrt{b}}{\sqrt{h_i}(bv^2 + b)} dv \\ &= \frac{3h_i}{b} - \frac{3h_i^{1.5}}{b^{1.5}} \int_{\sqrt{\frac{h_i}{b}}}^{\infty} \frac{1}{v^2 + 1} dv = \frac{3h_i}{b} - \frac{3h_i^{1.5}}{b^{1.5}} [\arctan(v)]_{\sqrt{\frac{h_i}{b}}}^{\infty} \\ &= \frac{3h_i}{b} - \frac{3h_i^{1.5}}{b^{1.5}} \left(\frac{\pi}{2} - \arctan\left(\sqrt{\frac{h_i}{b}}\right) \right) = -\frac{3h_i \left(\sqrt{\frac{h_i}{b}} \left(\frac{\pi}{2} - \arctan\left(\sqrt{\frac{h_i}{b}}\right)\right) - 1\right)}{b} \\ &= -\frac{3h_i \left(\sqrt{\frac{h_i}{b}} \arctan\left(\sqrt{\frac{b}{h_i}}\right) - 1\right)}{b}. \end{split}$$

Now by undoing the substitution $b = N \langle h \rangle (1 - e^{-\alpha})$:

$$E[s|H_i = h_i] = -\frac{3h_i \left(\sqrt{\frac{h_i}{N\langle h \rangle (1 - e^{-\alpha})}} \arctan\left(\sqrt{\frac{N\langle h \rangle (1 - e^{-\alpha})}{h_i}}\right) - 1\right)}{N\langle h \rangle (1 - e^{-\alpha})}.$$

For the situation N = 100 the results can be seen in Figure 7a and 8a and for N = 1000 in Figure 7b and 8b using $\alpha = 0.01$:



FIGURE 7: Expected long-term tie strength given popularity of person i (h_i) and person j (h_j) as a function of popularity of person i for 100 people in Figure 7a and for 1000 people in Figure 7b.



FIGURE 8: Expected long-term tie strength given popularity of person i (h_i) as a function of popularity of person i for 100 people in Figure 8a and for 1000 people in Figure 8b.

2.3 Comparing the plots of long-term tie strength of the two models

Since in model 2 the tie strength increases to a value of 1 after an interaction, also the expected long-term tie strength in Figures 7 and 8 will never reach a value above 1. This is different in model 1, in which the tie strength can reach very high values due to the fact it increases with 1 after an interaction. This can be seen in Figures 2 and 3 in which the expected long-term tie strength with a random other person j in the network is increasing with the popularity of a person i. Also, for higher N the expected long-term tie strength with situations of lower N. However, we see differences in the way the lines are increasing. In Figures 2 and 3 we see a linear growth of the expected tie strength for increasing popularity of person i. This can be explained by the linear nature of the formulas for the long-term expected tie strength in equations (7) and (8). In Figures 7 and 8 we see a decreasing growth of the expected tie strength for increasing popularity of the expected tie strength for increasing growth of the expected tie strength for increasing growth of the expected tie strength in equations (7) and (8). In Figures 7 and 8 we see a decreasing growth of the expected tie strength for increasing popularity of person i.

We now explain the nature of the graphs in Figures 2 and 3 of model 1 against the graphs in Figures 7 and 8 of model 2. Model 2 differs from model 1 in that the tie strength does not increase with 1 if there is an interaction, but increases to a value of 1. The consequences of this can be seen in the graphs of model 2 in Figures 7 and 8: we see that for increasing popularity the increase in expected tie strength is lower than in the graphs of model 1: Figures 2 and 3. To explain this we need the following reasoning: a higher popularity of a person i implies a higher probability of having an interaction with a random other person j during a certain time step, thus a higher amount of interactions on the long term with person j. Considering model 2, for more popular people, it will more often happen that the tie strength increases from an already relatively high value (close to 1) to 1 compared with less popular people. For less popular people the tie strength will more often increase from a relatively low value to 1. On the long run this means that in model 2, the ratio between the expected tie strength for popular people and less popular people is smaller than this ratio in model 1, in which the tie strength increases with 1 after each transaction, which happens more often for more popular people.

3 Real life interaction data

3.1 Interaction data applied to model 1

In this section we will take a closer look at real life data of interactions. We will look at interactions as being face-to-face contacts. For this we use face-to-face contact data during 20-second time intervals obtained through the SocioPatterns sensing platform [6]. We will use 6 data sets. The contexts in which these data were collected are: a workplace, with data collected in two different years (InVS13, InVS15), a hospital (LH10), a scientific conference (SFHH), a primary school (LyonSchool) and a high school (Thiers13). The data files represent the active contacts during 20-second intervals of the data collection. Each line has the form '*ijt*'. Here *i* and *j* are the anonymous IDs of the people in contact, and the interval during which this contact was active is [t - 20s, t] (with *t* in seconds).

We will use this data to simulate the tie strength between all people in the data set, according to the definition of tie strength in model 1 in equation (1), in MATLAB. In other words, if there is an interaction between persons i and j during a 20-second time window, then the value of the tie strength increases with 1 compared with the last window. If there is no interaction, then the tie strength decreases with a factor $e^{-\alpha}$. After all the simulations have been performed for the data available, we will acquire a tie strength matrix in which each entry (i, j) represents the tie strength between persons i and j at the last time at which data is available. It is important to note that we will not also perform a simulation according to the definition of tie strength in model 2 in equation (10). Since in this model the tie strength resets to 1 when there is an interaction, only the last interaction will contribute to the final value of the tie strength, which is opposed to model 1, in which all interactions are important for the final value.

Then a histogram plot can be made in which the distribution of the tie strength is represented. The x-axis provides the tie strength s on a linear scale and the y-axis provides the relative frequency of the strength values on a log scale. The log scale enables to see bars with widely varying heights and also more bars in the histogram. The relative frequency of the tie strength is defined as the fraction of the total number of ties that take certain tie strength values in a bin. It is important to note that we used 50 bins, which turned out to be right balance between seeing as many bars as possible and seeing all the different bars clearly. By using the relative frequency instead of the total frequency, the number of different ties in a network as in Table 1 does not influence the histogram. In this way, the histograms can be better compared. In the calculations of the long-term expected tie strength given popularity of person i as in equation (8) we assumed that the popularity of all the people in the network other than person i follows a power-law distribution. As we only computed the expected tie strength in equation (8), we do not necessarily expect the long-term tie strength in all the considered data to be exactly following a power-law distribution as well, but we do expect the data to be right skewed with a heavy tail, just as the power-law distribution. A possible suggestion for future research is to determine the best-fit distributions for the histograms below. This will be further explained in the conclusion and future work section (Section 4).

We will compare the plots of distribution of tie strength in similar contexts: the workplace in 2013 and 2015 and the primary and high school. Then we try to give (possible) explanations for the differences. We will also try to explain the plots of the hospital and scientific conference on their own, because they have nothing in common with the other contexts. We will do this by using - among other things - the concept of skewness, which is a measure of the asymmetry of the data around the sample mean. In the table below we will already summarize other data using the network data statistics for all data sets on the Network Data Repository [7], which we will use in analyzing the histogram plots. Here the start-up period is defined as the time it takes before at least two interactions take place in the same time window. All periods are measured in seconds.

Context	Participants	Different	Measuring	Starting	Start-up
		\mathbf{ties}	period [s]	time	period [s]
Workplace	92	755	987620	08:00	1240
2013					
Workplace	232	4274	993540	08:00	0
2015					
Hospital	75	1139	347500	13:00	540
Conference	403	9889	114300	09:00	880
Primary	242	8317	116900	08:40	0
school					
High school	329	5818	363560	08:20	0

TABLE 1: Network data statistics for different interaction networks

We will now focus on the skewness and the measures that indicate the height of the value of the skewness, namely the sample mean and the maximum of all the tie strength values. If skewness is negative, then the data spreads out more to the left of the mean than to the right. If skewness is positive, then the data spreads out more to the right. So the more positive the skewness, the more the data spreads out to the right. The skewness of every perfectly symmetric distribution is zero. It computes a sample version of this population value. In the formula of the biased skewness, μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of the quantity t:

$$s = \frac{E(x-\mu)^3}{\sigma^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}\right)^3},\tag{14}$$

In our case x is the matrix in which each entry (i, j) represents the tie strength between persons i and j at the last time at which data is available. As an addition to the histogram plots of the distribution of the tie strength in different contexts, we will first list the sample mean and the maximum of all the tie strength values together with the skewness of the tie strength values in the table below. In this way it becomes more evident that all the distributions are right skewed (all the skewness values are positive) and that the positive (right) skewness values differ between the different contexts.

Context	Sample mean of	Maximum of tie	Skewness of tie
	tie strength	${f strength}$	${f strength}$
Workplace 2013	0.2106	68.88	20.90
Workplace 2015	0.3589	651.0	43.16
Hospital	0.3977	76.96	18.95
Conference	0.09275	435.3	94.65
Primary school	0.4660	85.14	16.54
High school	0.1655	149.6	40.85

TABLE 2: Skewness for different interaction networks

3.2 Long-term tie strength in a workplace in two different years

In this section we will simulate the long-term tie strength in a workplace by using a data set of a temporal network of contacts between individuals measured in an office building in France, from June 24 to July 3, 2013, which was described and analyzed in [11], and in 2015, described and analyzed in [12]. In the figure below the distribution of the tie strength, using $\alpha = 0.01$, for these data-sets in 2013 and 2015 is represented in the form of a histogram:



FIGURE 9: Distribution of the tie strength s in a workplace in France in 2013 in Figure 9a and in 2015 in Figure 9b for $\alpha = 0.01$.

As expected, both above histogram plots are right skewed with a heavy tail, just as the power-law distribution for the long-term expected tie strength that is predicted in Figure 3 by equation (8). Comparing the two histograms, we see the same relative distribution: a relatively large fraction of the people has a relatively low value of tie strength. What differs is that the data of 2015 results in much higher values for the tie strength in general (in particular a higher maximum long-term tie strength as can be seen in Table 2). This, in combination with a similar low sample mean of tie strength, then leads to an even stronger right skewed distribution: in 2015 the plot has an even longer right tail than in 2013 as can be seen in Figure 9. Another cause of this stronger right skewness has to do with start-up effects. If we take a closer look at the start-up periods in both years in Table 1, then we see that the interactions in 2015 are much faster at a high level compared with those in 2013, while the measuring period and starting time are similar. This has probably to do with stricter regulations regarding being on time at work at 8:00 in 2015. In 2013 interactions

really started to build up only around 8:20. We can also mathematically substantiate this stronger right skewed distribution in 2015 by calculating the skewness as in equation (14). We acquire a value of 20.90 for the data of 2013 and 43.16 for 2015, which confirms our thoughts.

As a side note, we will now look at what impact a lower and a higher value of α has on the shape of the histogram plots as in Figure 9. We will only do this for the case of the workplace, since the effect of a change in α in other contexts will be similar. We then try to explain the different shapes of the corresponding histograms by using the effect of different values of α on the tie strength in the theoretical model 1 as in Subsection 2.1.1.

First, we plot the distribution of the tie strength for these data-sets in 2013 and 2015 for a lower value of α ($\alpha = 0.0001$):



FIGURE 10: Distribution of the tie strength s in a workplace in France in 2013 in Figure 10a and in 2015 in Figure 10b for $\alpha = 0.0001$.

If we compare the distribution of tie strength in the workplace in 2013 and 2015 for $\alpha = 0.01$ in Figures 9a and 9b respectively and for $\alpha = 0.0001$ in Figures 10a and 10b respectively, then we see that for a lower value of α the relative frequency of higher tie strength values is higher. This means that less mass of the distribution is concentrated to the left of the mean, thus the right skewed distribution is weaker. Consequently, more bars can be seen in the histogram plot. Because of this, the heavy tail of the right skewed distribution can be seen more clearly. In Figures 10a and 10b it can be clearly seen that the higher the tie strength values, the lower the bars (the lower the relative frequency of the tie strength values). This was already the case for $\alpha = 0.01$ in Figures 9a and 9b, but there only eight and five bars were visible respectively in contrast to respectively 15 and 24 bars for $\alpha = 0.0001$ in Figures 10a and 10b.

We do expect higher tie strength values in general for a lower value of α if we have a look back at the definition of tie strength in model 1 in equation (1) that we used for these simulations. The lower the value for α , the higher the tie strength values in general. We also saw this relation in the plots we made for model 1 for N = 100 in Figures 2a and 3a for $\alpha = 0.01$ compared with Figures 4a and 4b for $\alpha = 0.0001$ respectively.

Second, we plot the distribution of the tie strength for these data-sets in 2013 and 2015

for a higher value of α ($\alpha = 1$):



FIGURE 11: Distribution of the tie strength s in a workplace in France in 2013 in Figure 11a and in 2015 in Figure 11b for $\alpha = 1$.

If we compare the distribution of tie strength in the workplace in 2013 for $\alpha = 0.01$ in Figure 9a and for $\alpha = 1$ in Figure 11a, then we see that for higher α the relative frequency of lower tie strength values is higher. This means that more mass of the distribution is concentrated to the left of the mean, thus the right skewed distribution is stronger. Consequently, less bars can be seen in the histogram plot. Because of this, the heavy tail of the right skewed distribution can be seen less clearly. In Figure 11a it can be seen less clearly, compared with Figure 9a, that the higher the tie strength values, the lower the bars.

We do expect lower tie strength values in general if we have a look back at the definition of tie strength in model 1 in equation (1). The higher the value for α , the lower the tie strength values in general. We also saw this relation in the plots we made for model 1 for N = 100 in Figures 2a and 3a for $\alpha = 0.01$ compared with Figures 5a and 5b for $\alpha = 1$ respectively.

If we compare the distribution of tie strength in the workplace in 2015 for $\alpha = 0.01$ in Figure 9b and for $\alpha = 1$ in Figure 11b, then we do not see clearly that for higher α the relative frequency of lower tie strength values is higher, what we would expect if we look at the definition of tie strength in equation (1). We only see that the bar around a tie strength value of 40 in Figure 9b is included in the large bar around 0 in Figure 11b. We can explain the fact that the 3 remaining bars for tie strength values higher than 200 do not take lower values for a higher value of α . The value of α only influences the value of tie strength between two persons *i* and *j* if they do not have an interaction for a certain time. Since two people with a very high tie strength almost have had an interaction in every time interval, their final value of tie strength has not been negatively influenced much by a higher value of α , so the bars did not move to a lower bin.

3.3 Long-term tie strength in a hospital

In this section we will simulate the long-term tie strength in a hospital ward in Lyon, France by using a data set of a temporal network of contacts between patients, patients and health-care workers (HCWs) and among HCWs from Monday, December 6, 2010 at 1:00 pm to Friday, December 10, 2010 at 2:00 pm. The study included 46 HCWs and 29 patients and is described in [13]. In the figure below the distribution of the tie strength for this data-set is represented in the form of a histogram:



FIGURE 12: Distribution of the tie strength s in a hospital ward in Lyon

This histogram plot is right skewed with a heavy tail, just as the power-law distribution for the long-term expected tie strength that is predicted in Figure 3 by equation (8). We see in Figure 12 that a relatively large fraction of the people in the hospital has a very low value of tie strength. What strikes us, is the weak right skewed distribution compared with the other contexts in Figures 9, 13 and 15 (except for the primary school in Figure 14). This is also confirmed by the second lowest value of skewness in Table 1 of 18.95. This is caused by a very low maximum long-term tie strength in combination with a relatively high sample mean of tie strength as can be seen in Table 2.

3.4 Long-term tie strength of participants to a scientific conference

In this section the long-term tie strength of participants to a scientific conference is simulated using a data set that describes the face-to-face interactions of 405 participants to the 2009 SFHH conference in Nice, France (June 4-5, 2009). It was released as data set in [12]. In the figure below the distribution of the tie strength for this data-set is represented in the form of a histogram:



FIGURE 13: Distribution of the tie strength s of participants to a scientific conference

This histogram plot is right skewed with a heavy tail, just as the power-law distribution for the long-term expected tie strength that is predicted in Figure 3 by equation (8). We see in Figure 13 that a relatively large fraction of the participants to the conference has a very low value of tie strength. What strikes us, is the very strong right skewed distribution compared with all the other contexts in Figures 9, 12, 14 and 15. This is also confirmed by the highest value of skewness in Table 1 of 94.65. This is caused by a very high maximum long-term tie strength in combination with a very low sample mean of tie strength as can be seen in Table 2. We also see in Table 1 that the conference had the lowest measuring period (114300 seconds). This implies that the tie strength are less likely to reach high values which leads to very unevenly distributed tie strength values.

3.5 Long-term tie strength in a primary school

We will now simulate the long-term tie strength in a primary school by using a data set of a temporal network of contacts between children and teachers. It is used in the study published in BMC Infectious Diseases 2014 [14] and in Plos One [15]. In the figure below the distribution of the tie strength for this data-set is represented in the form of a histogram:



FIGURE 14: Distribution of the tie strength s in a primary school

Again this histogram plot is right skewed with a heavy tail. More explanation will follow in Subsection 3.7.

3.6 Long-term tie strength in a high school

Now we will simulate the long-term tie strength in a high school using a data set that gives the contacts of the students of nine classes in Marseilles, France, during 5 days in December 2013. It is analyzed in the publication [16]. In the figure below the distribution of the tie strength for this data-set is represented in the form of a histogram:



FIGURE 15: Distribution of the tie strength s in a high school

Again this histogram plot is right skewed with a heavy tail.

3.7 Comparing the distribution of the tie strength of the primary school with the one of the high school

As expected, both histogram plots in Figures 14 and 15 are right skewed with a heavy tail, just as the power-law distribution for the long-term expected tie strength that is predicted

in Figure 3 by equation (8). Comparing the two histograms, we see the same relative distribution: a relatively large fraction of the people has a relatively low value of tie strength. What differs is that the data of the high school results in higher values for the tie strength in general (in particular a higher maximum long-term tie strength as can be seen in Table 2). This, in combination with a lower sample mean of the strength (0.1655 for the high school against 0.4660 for the primary school), then leads to an even stronger right skewed distribution: for the high school the plot has an even longer right tail than for the primary school as can be seen in Figures 14 and 15. Another possible cause for this is also similar to one of those mentioned for the workplace in two different years in Subsection 3.2: the interactions for the high school are faster at a higher level compared with the primary school. Assuming both schools started at the same time measuring interactions as the time school nearly starts (which is plausible looking at the starting times in Table 1), a possible explanation for this can be given. From own experience, while attending high school you have long interactions rather than many interactions. You have close friends rather than many friends, so you will contact these close friends faster than a student in primary school would contact a 'less close' friend. This, in combination with the longer measuring period (363560 for high school against 116900 for primary school), leads to a higher chance for tie strength in high school to reach higher values at the end of the measuring period. So the data spreads more out to the right. We can also mathematically substantiate this stronger right skewed distribution for the high school by calculating the skewness as in equation (14). We acquire a value of 40.85 for the data of the high school and 16.54 for the primary school, which confirms our thoughts.

When looking at the differences between the number of participants and number of different ties for both schools in Table 1, we see that the number of participants is higher for the high school (329) compared with the primary school (242), while the number of different ties formed is much lower for the high school (respectively 5818 and 8317). This implies that the number of different ties formed per student is lower in the case of the high school. Using this data, the previously mentioned possible explanation for the stronger right skewed distribution for the high school - while attending it you have long interactions rather than many interactions - becomes even more realistic.

4 Conclusions and recommendations

In this section we will return to the research question and goals as formulated in the introduction in Section 1. We will also formulate new insights and ideas that the research led to that still need to be worked out.

The concept of tie strength can be used to analyze interactions. The tie strength describes the strength of your friendship with one of your friends. It increases when an interaction takes place, and then decreases over time if no new interaction takes place. This concept can be very useful to find the most important connections in a social network.

Our research question was: What happens to the tie strength when some people in the network are more popular than others? We saw in Section 2 that the more popular a person, the higher is the long-term expected tie strength between this person and a randomly chosen other person. We even saw in model 1 that this relation is linear in Figures 2 and 3. This was concluded by assuming that the popularity of all the people in the network other than person i follows a power-law distribution, which is a right skewed distribution

with a heavy tail: a relatively small fraction of people has a large amount of popularity. In model 2 the relation between popularity and long-term expected tie strength is not linear and more complicated than in model 1 as can be seen in Figures 7 and 8.

We also saw that the chosen value of the decay rate α has a large effect on the tie strength values for different popularity in the theoretical model 1 and on the distribution of the tie strength for which real life interaction data is used. A higher value for α implies lower tie strength values in general. Then, more mass of the distribution is concentrated to the left of the mean and the right skewed distribution is stronger. In our analysis of the workplace the heavy tail in the distribution of the tie strength was the most clearly visible for $\alpha = 0.0001$. When choosing higher values for α , more tie strength values are close to 0.

We saw that the distribution of tie strength of real life interaction data, according to the definition of tie strength in model 1 in equation (1), in different contexts is very right skewed (which implies high values for the skewness), similar to the assumed power-law distribution of popularity in both models. This right skewed data is also what we expect following the linear relation between popularity and long-term expected tie strength in model 1 in equation (8).

A first suggestion for future research is to develop a good procedure to choose the right value for the decay rate α to model a certain data-set, so that the differences between all tie strength values at the end of the measuring period are a good representation of the differences in the strength of all the friendships in reality. Therefore the value of α should maybe depend on the length of the measuring period of the certain data-set or on another specific network data-set statistic as in Table 1. It is also possible that it should depend on the mean number of interactions per time window over the whole measuring period of the certain data-set.

A second suggestion is to take into account the department, class and/or status of each individual in each context for plotting and analyzing the histograms. In this paper, we did not distinguish between interactions between people from the same department, class and/or status and those between different ones. For example, in the hospital a distinction can be made in interactions between patients, patients and health-care workers (HCWs) and those among HCWs. Then by counting the number of interactions in each category a more detailed explanation can be given for the low maximum long-term tie strength and the relatively high sample mean. Also the shape of the histogram can be better explained. A similar approach can be used for other contexts for which a distinction can be made in department, class and/or status of the individuals. For each context, except for the scientific conference (since all the individuals in the data-set are participants), the IDs and their corresponding department, class and/or status are available through the SocioPatterns sensing platform [6]. For our research it was only important to investigate whether or not the distribution of the tie strength of real life interactions in a certain context, regardless of the department, class and/or status of the individuals, does really follow a right skewed distribution with a heavy tail, just as the power-law distribution.

Determining the best-fit distribution is our third recommendation for future work. We did conclude that all the histogram plots were right skewed with a heavy tail, but this does not necessarily imply that all the distributions are power-law distributions. We did not

investigate which distribution(s) best fits to the histograms of the tie strength in different contexts in Figures 9, 12, 13, 14 and 15. These can be worked out in further research by comparing different distributions.

References

- [1] Popularity Wikipedia.
- [2] Walid Ahmad, Mason A Porter, and Mariano Beguerisse-Díaz. "Tie-decay temporal networks in continuous time and eigenvector-based centralities". In: (2018).
- [3] Emily M Jin, Michelle Girvan, and M. E.J. Newman. "Structure of growing social networks". In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 64.4 (2001), p. 8.
- [4] Xinzhe Zuo and Mason A Porter. "Models of Continuous-Time Networks with Tie Decay, Diffusion, and Convection". In: (2019).
- [5] Remco van der Hofstad. Random graphs and complex networks. Vol. 1. 2016, pp. 1– 321.
- [6] Datasets « SocioPatterns.org.
- [7] Ryan A. Rossi and Nesreen K. Ahmed. "An Interactive Data Repository with Visual Analytics". In: ACM SIGKDD Explorations Newsletter 17.2 (Feb. 2016), pp. 37–41.
- [8] Rajmonda Sulo, Tanya Berger-Wolf, and Robert Grossman. "Meaningful selection of temporal resolution for dynamic networks". In: *Proceedings of the Eighth Workshop* on Mining and Learning with Graphs - MLG '10. New York, New York, USA: ACM Press, 2010, pp. 127–136.
- [9] Aaron Clauset. "Inference, Models and Simulation for Complex Systems". In: (2011).
- [10] Aaron Clauset, Cosma Rohilla Shalizi, and M. E.J. Newman. Power-law distributions in empirical data. 2009.
- [11] Mathieu Génois et al. "Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers". In: *Network Science* 3.3 (2015), pp. 326–347.
- [12] Mathieu Génois and Alain Barrat. "Can co-location be used as a proxy for face-to-face contacts?" In: *EPJ Data Science* 7.1 (2018), p. 11.
- [13] Philippe Vanhems et al. "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors." In: *PloS one* 8.9 (2013).
- [14] Valerio Gemmetto, Alain Barrat, and Ciro Cattuto. "Mitigation of infectious disease at school: Targeted class closure vs school closure". In: *BMC Infectious Diseases* 14.1 (Dec. 2014), p. 695.
- [15] Juliette Stehlé et al. "High-resolution measurements of face-to-face contact patterns in a primary school". In: *PLoS ONE* 6.8 (2011).
- [16] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. Ed. by Cecile Viboud. Sept. 2015.

5 Appendix

This section will include source MATLAB code we used to plot our results.

First, we have the code of plotting equation (7) against the popularity of person i as can be seen in Figure 2 plotted by 'figure(1)'. Then we also have the code of plotting equation (8) against the popularity h_i as can be seen in Figure 3 plotted by 'figure(2)':

```
1 % Long-term tie strength against popularity plot
2 M = 10; %Number of times values for h_j are determined
3 N = 1000; %Number of people in the network
4 h_j = (1-rand(M,1)).^(-1/1.5); %Popularity of person j
5 h_m = 3; %Expectation of h_j
6 h_i = 1:0.00001:100; %Range of popularity of person i
7 alpha = 1; %Decay rate
8 Eshi = (3.*h_i)./((1-exp(-1*alpha)).*h_m.*N);
9 %Expected tie strength given popularity of i
10 hold on
11
12 figure(1)
13
  for k = 1:10
  %Calculates expected tie strength given popularity of i and j
14
          Es = (h_{i.*}h_{j(k)})./((1-exp(-1*alpha)).*(N.*h_m));
15
          txt = ['h_j = ', num2str(h_j(k))];
16
          plot(h_i,Es,'LineWidth',1.5,'DisplayName',txt)
17
18 end
19 xlabel('h_i: Popularity of person i')
20 ylabel('E[s|h_i,h_j]: Expected tie strength given popularity of i and j')
21 title('Long-term tie strength against popularity')
22 axis fill
23 hold off
24 legend('show','Location','northwest')
25
26 figure(2)
27 plot(h_i,Eshi,'LineWidth',1.5)
28 xlabel('h_i: Popularity of person i')
29 ylabel('E[s|h_i]: Expected tie strength given popularity of i')
  title('Long-term tie strength against popularity')
30
31
  axis fill
  hold off
32
```

The code of plotting equations (12) and (13) against the popularity of person i, in Figures 7 and 8 respectively, is similar to the previously mentioned code. The only difference is that 'Es' is now equal to equation (12) and 'Eshi' equal to equation (13).

Finally, we have the code for plotting the distribution of the tie strength, using $\alpha = 0.01$, for different data-sets in the form of a histogram. The following code is used to plot the histogram of the workplace in 2013, as in Figure 11a. The histograms of all the other data-sets can be plotted as well, just by changing the argument of '*importdata*'.

```
1 data = importdata('ia-workplace-contacts.edges'); %Imports data-set
2 T = size(data,1); %Number of interactions
3 m = max(max(data(:,1:2))); %Maximum ID
4 S = NaN(m);
5 %Stores the tie strength between person i and j at a certain time step
6 alpha = 0.01; %Decay rate
```

```
7 pl = 0; %Time = 0
8 for t = 1:T %Initialize all data-entries at 0
9
       S(data(t,1), data(t,2)) = 0;
10 end
11 for t = 1:T %Updates the tie strength during every time window
     if data(t,3) ≠p1
12
         S = S + exp(-1 + alpha);
13
14
          p1 = data(t,3);
      end
15
      S(data(t,1), data(t,2)) = S(data(t,1), data(t,2))/exp(-1*alpha)+1;
16
17 end
18 S(isnan(S)) = []; %Remove all NON data-entries from the matrix
19 NumberTS = numel(S); %Number of different ties
20 [N, edges] = histcounts(S, 50);
21 %Number of tie strength within a certain bin of in total 50 bins
22 idx = find(N); %Find indices of nonzero elements of N
23 disp(['Alpha =',num2str(alpha),':']) %Display alpha
24 for i = idx %Display bin with count
       disp(['Bin = ', '(',num2str(edges(i)),'-',num2str(edges(i+1)), '), ...
25
           count = ', num2str(N(i))])
26 end
27
28 figure(1)
29 histogram(S, 50, 'Normalization', 'probability')
30 %Plot normalized histogram with 50 bins
31 MeanTS = mean(S); %Mean value of all values of matrix S(i,j)
32 MaxTS = max(S); %Max value of matrix S
33 Skewness = skewness(S); %Skewness of S
34 set(gca, 'XScale', 'linear')
35 set(gca, 'YScale', 'log')
36 xlabel('s: tie strength')
37 ylabel('relative frequency')
38 title('Distribution of the tie strength')
```