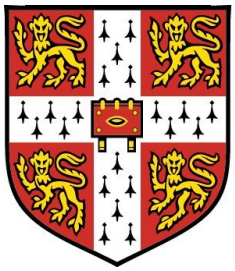
Error and Attack Vulnerability of Temporal Networks

Stojan Trajanovski
Fitzwilliam College



A dissertation submitted to the University of Cambridge in partial fulfillment of the requirements for the degree of Master of Philosophy in Advanced Computer Science

University of Cambridge
Computer Laboratory
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
e-mail: st508@cam.ac.uk

June 16, 2011

Declaration

I, **Stojan Trajanovski** of **Fitzwilliam College**, **University of Cambridge** being a candidate for the M.Phil in Advanced Computer Science, hereby declare that this report and the work described in it are my own work, unaided except as may be specified below, and that the report does not contain material that has already been used to any substantial extent for a comparable purpose.

Total word count: **13,360**

Signed:

Date:

to my parents and my sister
на мама, тато и Андреа

Abstract

The aim of this dissertation is to investigate the robustness of time-varying communication networks under different kind of failures. The work relies on the widely accepted temporal network analysis and network robustness. Both approaches have never been used together for performance evaluation of a networked systems under intelligent attacks. The work considers several temporal theoretical models and data from real world. Temporal robustness is evaluated, measuring relative change in performance metric before and after sustaining several intelligent attacks and random failures.

The results show that temporal networks, where some nodes are more dominant or central, are more affected by intelligent attacks than random errors. Moreover, different intelligent attacks show similar effect on the temporal robustness, because the same nodes are recognized as "important entities" in the temporal network. Contrarily in temporal networks where all the nodes have similar properties, random failures and all the intelligent attacks cause the same effect. This work also defines and determines *robustness range* for all intelligent attacking strategies, which gives all possible temporal robustness values respect to choice of attacked nodes. These conclusions suggest better protection of important entities or decentralized network architecture, which leads to more robust design of the system.

Acknowledgments

It is a great pleasure to thank everyone who have made this dissertation possible. I am grateful to my supervisor, Dr. Cecilia Mascolo, for her guidance throughout my research. I also would like to thank Salvatore Scellato and Dr. Ilias Leontiadis from the Computer Laboratory, University of Cambridge, for valuable suggestions and tutorials during the work. Finally, my gratitude goes to my course advisor Dr. Richard Gibbens and my friends Duncan Roberts, Éireann Leverett and Thomas Piachaud for their technical suggestions and initial proofreading.

Contents

List of Figures

List of Tables

List of Algorithms

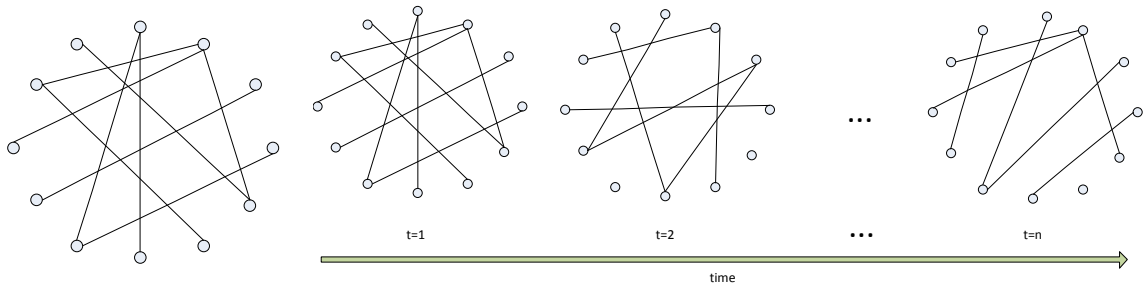
Chapter 1

Introduction

This dissertation explores the effects of several error and attack strategies on real systems and theoretical models. The main goal is to understand how time-varying networks react to random errors and targeted attacks.

1.1 Motivation

The functionalities of real world systems are based on the communications and interactions of participating entities. Those entities cannot perform their actions individually and independently. The required input for one entity depends on the outcome that some other produces. Naturally, those systems can be modelled as networks. Communication between our friends, transport network in a certain area, interactions and contractions between brain cells have fascinated scientists for decades.



(a) Static network (b) Temporal network

Figure 1.1: Network topologies

In a famous example from 1967 "six degree of separation" [Mil67] by Stanley Milgram, a message initiated by one person needs at most 6 intermediate people to reach everyone on the planet. Using a straightforward approach one can model the entities in the system as a static network topology (Figure 1.1a). Although people are still present in the

system, their ability of communication is not constant. Due to the environmental, the technological or even subjective nature the communication between them is not consistent over a given period. People may decide to transfer the information later or communication devices that they use to have some technical problems which again delays the spreading. People interactions are time-varying and fast evolving during the time. Accordingly, the network representation should express the feature of time-dependent and ordered interactions. The last leads us to *temporal networks* (Figure 1.1b), where interactions are subject to changes during the time. In the other words, this representation may be considered as a collection of several static representations.

Complex network analysis is a driving force behind the research in many real systems and people interactions. It represents every system as a network topology with non-trivial features, where the entities (participants) are represented by the set of *vertices* and their interactions [BLM+06] as a pairs of vertices or *links*. More naturally, complex networks are recognized by real-world examples such as the Internet and World Wide Web [FFF99, HA99], transport networks (road, air or train traffic), biological, brain networks [VCAL11] and protein relations, human interactions (scientists' collaborations, memberships or social networks) and many others. Complex networks provide substantial tools for analyzing the interactions inside the network that can answer many interesting research questions.

The common approach includes network analysis in a fixed time. This means that only the interactions at a given moment are taken into consideration, at the same time neglecting the evolution and the temporal properties of the network [Kos09, KKK00]. In some cases, where the network evolves more slowly, for instance road networks, static analysis still makes sense. On the other hand, the changes in the network may have a significant role. In cases where interactions are prone to fast changes, static analysis leads to wrong results [TMML09, TSM+10]. For example, collaboration networks between scientists can evolve rapidly during a conference or communication between moving taxi cabs equipped with wireless devices is prone to fast change due to their mobility and in-stability of wireless connection. The main weakness of static network representation is overestimation of temporal connectivity: two entities can interact or communicate at one moment, but this may not be a case at the next. For instance, discussions between scientists at a conference are more intensive during the presentation sessions than late at night. This shows us the importance of the chosen moment, when a network's representation is taken. Hence the most relevant information for the network's interactions or particular network metric (e.g. average distance between nodes) could be obtained considering the network as a dynamic system.

For this reason I employ *temporal network analysis*. For each pair of nodes we consider the time required to spread a message from one to the other. This time defines *temporal distance*. In order to characterize the whole temporal network, temporal distances between all pairs participate in an averaged sum that defines the performance metric for evaluation.

On the other hand, investigating the performance of a real networked system after it has sustained structural damage has been extensively considered [AJB00, CNSW00]. Several measures of network robustness have been proposed and used to evaluate the change in performance under random errors or intelligent attacks. Both failures are motivated from real examples. Random errors are usually initiated by some internal fault, expired and old parts in the system or unattended change. Intelligent attacks may be regarded as malicious external damage, such as malware software or planned attack on important hubs in a computer network, attacking influential people or targeting important objects like power plants in some country.

Current research in network robustness usually considers static characteristics, neglecting the time-changes and the time-ordering in real systems. By adopting a static approach, only performance of the system at a certain moment is measured, ignoring the probability that at the previous moment the same value was significantly lower or higher. For this reason the dissertation proposes two important changes for evaluating robustness. Firstly, I extend the definition of robustness for temporal networks, which defines *temporal network robustness*. Secondly, both random errors and planned intelligent attacks targeting important nodes are considered in the evaluation.

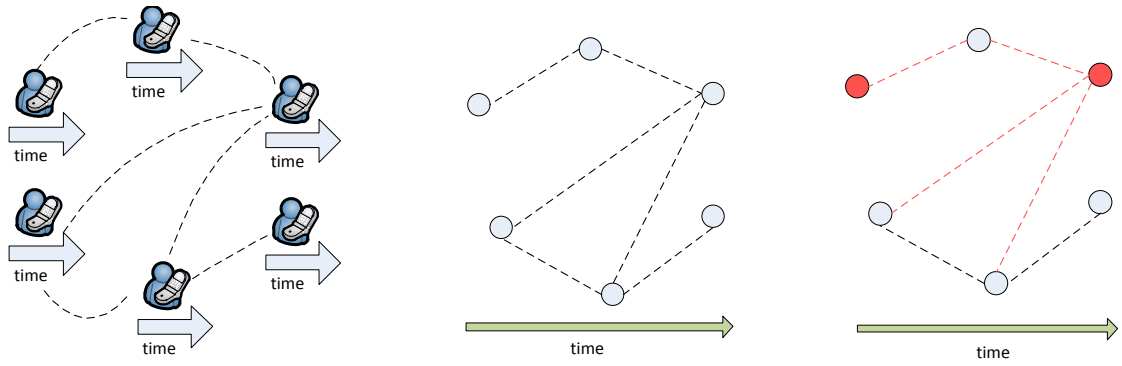
1.2 Problem Description

In order to evaluate the performance of the network after a disturbance, we should measure the relative change in performance compared with the state before it. The measure of relative change in network performance is *robustness*. Robustness may be thought as a measure of "goodness", an indicator of *how well the network performs after sustaining damage*. Denoting robustness by value R, $R = 0$ means that the network is totally destroyed and cannot perform any more. Robustness value $R = 1$ reflects the same performance as previously. In general, it is a value in the interval $[0, 1]$ and values closer to 1 report for satisfactory performance after some structural damage. In the robustness evaluation temporal properties and the dynamic of a system should be taken into account.

This robustness metric for both static graphs and temporal networks should be as simple as possible, interpretable and feasible, which means suitable for all types of finite networks [VMDW$^+$10]. In order to calculate temporal network robustness we need to agree which performance metric will be considered. At a certain moment two nodes in the network may not be connected. A single message could be transferred between any two nodes if there is a path between them. However, if there is no path between the sender and the destination, the message is re-sent by the nodes that already have it. At a time later, if there is a path to some node that has the message, the destination receives it. The average time required to spread the information between two nodes is *temporal efficiency*. Temporal efficiency is employed as a performance metric whose relative change defines

temporal robustness.

The mapping of a real system into temporal network representation and robustness evaluation is shown in Figure 1.2. Nodes that sustain structural damage and links incident to them are coloured red in Figure 1.2c.



(a) Real system (b) Before damage (c) After damage

Figure 1.2: Robustness is relative change of the network performance

Structural damages in temporal networks differ in nature. In general, they can be divided into two classes: *random errors* and *intelligent attacks*. In order to determine the severity of damage that a network sustains we define a *probability of error/attack*. Probability of error/attack determines the number of removed nodes, which means that increased probability of error/attack implies more nodes sustain errors/attacks. The main difference between random errors and intelligent attacks lies in the decision which nodes are removed. Random error strategy randomly picks nodes and removes them from the network. Intelligent attacks strategies target the nodes that have a certain node's temporal property (e.g. average degree of a node).

The work in this dissertation considers random error and several attacking strategies evaluated on different temporal networks. Two general groups are used: theoretical temporal models and real temporal data sets.

The main findings of this dissertation can be summarized as follows:

- Temporal robustness shows graceful decrease in both random errors and intelligent attacks, unlike the behaviour of the robustness in static networks.
- For each intelligent attack strategy, *robustness range* is evaluated; that is, expected deviation of temporal network robustness by different choice of removed nodes. The robustness range can determines all possible values of robustness. If temporal robustness is larger, it means that there are some nodes, considered as "important hubs" and their failure will significantly affect the performance; otherwise all the nodes contribute similarly in the general system performance.

- The influence of different intelligent attacks strategies is similar for a particular temporal model or real temporal network, since same nodes are considered as important in all the strategies. This finding expresses the behaviour of important entities to be: highly connected, most of the paths traverse through them and contribute at most in information spreading.

- In real-world temporal networks and mobility models the influence of intelligent attacks is higher than in the case of random errors, due to the presence of important pre-dominant nodes. Temporal robustness value can drop by 50% to 75% if important entities are attacked, rather than randomly chosen nodes. This means that more robust design can be achieved in two ways. Firstly, important entities in the system should be better protected. Secondly, by considering changes in the architecture and, modification into more decentralized (e.g. P2P or multiple mini data-centres) design is recommended.

- In the case of equally distributed temporal networks (e.g. Erdős-Rényi or Markov temporal networks), temporal robustness is similar for intelligent attacks and random error strategy, because all the nodes have similar temporal property (e.g. average node degree, temporal closeness). In this case, temporal robustness can be determined irrelevant to attacking strategy and choice of attacked nodes.

1.3 Outline

The remaining part of the dissertation is organized as follows. Chapter 2 presents the concepts of temporal metrics and network robustness, including performance metrics and implementation issues. In Chapter 3 details of theoretical temporal network models are given. Chapter 4 evaluates temporal network robustness on different models and discusses the results. Case studies that consider temporal network data sets from real world and robustness evaluation are presented in Chapter 5. Chapter 6 states conclusions and suggests directions for future work.

Chapter 2

Temporal Metrics and Robustness

Networked real world systems could be modelled as network topologies or, as is commonly accepted, as networks. For instance, in a computer network they can be processing/computational nodes or just routers and hubs that re-route/transmit a networking packet. The entities are not isolated single elements and performing their service, they interact (e.g.exchanging information or communicating). Their interactions are represented as links between the nodes that communicate.

The interactions in one system are subject of changes during a certain period. The intensity of interaction between two entities varies at all the time moments. In order to measure performance of time-varying network, we should consider suitable metric that will express the system performance at all time moments. Static metrics consider the network at a certain moment. In this case measured performance is only determined by a certain system state, neglecting all previous states which lead us to wrong conclusions. For this reason, temporal metrics that count all time steps of the temporal network are employed in the analysis.

Apart from network topology, the system taken in general as a unit is responsible for performing some service. In a road network, it is the quality of the transport, in banking system it is a financial transaction, in social network the speed of information spreading. In order to quantify the service performance we need to define a metric. Current research already has reported measures [BLM+06, dFCRTVB07] like average shortest path, quality of service or size of the giant component. The metric should be descriptive enough to present the performance in general, apart from previously mentioned temporal property.

2.1 Background and Related Work

Robustness of networks have been intensively studied in the last few decades. However, due to the complexity of communication networks such as multiple layers, neglecting service aspects or very often dynamic nature of the networking systems, it is

difficult to generalize robustness metric. Some surveys of network robustness in general were done in [CMH+07, VMDW+10]. The earliest studies were on network reliability [SP78, Wil72, RA78] investigating connectivity of the network. These works mainly focus on preserving connectivity after failures. Other work uses reliability polynomials [PP94] for robustness evaluation. One shortcoming of reliability is unavailability to characterize irregular failures. The recent study emphasizes power-law networks reliability [HKYH02], because this well modelled real world examples (e.g. World Wide Web [FFF99]). Investigating the network performance after sustaining different types of failures, including random error or intelligent attacks was studied by Albert et. al. [AJB00]. Considered networks vary from static random models to scale-free networks with predominant nodes. A more theoretical direction is followed by [CNSW00], considering network behaviour undergo nodes and/or edges removal. This work was extended by investigation of what kinds of error and attacks are more disruptive for real systems and models [HKYH02]. However, this work is mostly based on static network representation.

On the other hand, temporal network analysis aims to upgrade the model of static networks. The first recorded attempt to express temporal network property was made by Kempe et. al. [KKK00], considering time labels of links. The authors proposed an algorithm, but this approach neglects temporary disconnected nodes. Temporal correlation of human-interactions was considered in [CE07]. In the same work periodical behaviour of the systems is studied. The concept of temporal networks as a collection of static network topologies taken in suitable time resolution was proposed in [Kos09]. This temporal representation is used as a base in this work. Recently, the idea of temporal metrics, namely temporal paths and temporal length were given in [TMML09, TSM+10].

Network robustness and temporal network analysis have rarely been used together in performance evaluation of time-varying systems. Recently, by [SLM+11] the concept of temporal network robustness has been defined with a framework suitable for temporal models and real data-set. The way that robustness is evaluated considers random errors on real systems and theoretical models. The dissertation extends this approach considering network vulnerability under several intelligent attack strategies apart from random errors and proposes the concept of *robustness range*. This variable reports for possible value of robustness respect to the choice of attacked nodes.

2.2 Temporal Network Analysis

This section defines the concepts of Temporal Network Analysis, including temporal networks and temporal metrics used as performance metric of robustness evaluation.

2.2.1 Temporal Networks

Networks are commonly used for representation of entities and their relations and interactions. In this form, nodes are associated with presence of entities (e.g. communication units, human beings) and links express their behaviour, communications, interactions or other type of connection. The static graph and network representations take one state of the network in a certain moment. Static representation (e.g in Figure 1.1a) is one that has been commonly studied, considering different metrics for characterization of a certain node (degree of a node, closeness centrality) or global metrics for all the nodes (e.g. clustering coefficient). This analysis works well in cases where the graph change is slow and it keeps the entities and their relations during the given time. An example taken from traffic network reports more congested communication in some periods of the day than others. In particular, spreading the information between two nodes in the network is more probable when the communication is intensive, rather than less intensive periods. The last shows us importance of the time choice, when a network's representation is taken. Hence, the most relevant information on the network's interactions or a given metric (e.g. average distance between nodes) could be obtained considering the network as a dynamic system.

The limitation of static graph becomes significant in cases, in which we have a sharp changes in the entities' interactions. Because of this, we define *temporal networks* (e.g. Figure 1.1b).

Definition 1. *A temporal network* $G(t) = G(V, E(t))$ *is a sequence of* n *undirected static network representations* $\{G(t_i)\} = \{G(V, E(t_i))\}$ $(i = 1, 2, \ldots, n-1, n)$.

Temporal network may be thought as sequence of consecutive static graphs. During the time evolution of the network the set of nodes $V = V(t) = const$, where $|V| = N$ is kept constant and only the set of links in the graph is subject to changes. The set of nodes is fixed, which means that new nodes cannot be added to the graph nor nodes can be removed. However, even in the case where the network sustains structural damage (error or attack) the set of nodes is preserved and each node is counted in the final evaluation, but attacked nodes do not contribute in the performance metric after the error/attack. Formally, attacked nodes are counted in the averaging, although their contribution in the performance function (temporal efficiency in this case) is zero.

2.2.2 Temporal Path Length

Temporal distance does not appear in static graphs. In static cases the condition for spreading the information between two particular nodes is their connectivity or presence of a path between these nodes. In temporal networks the information (message) travels during the given time. At every time moment, a node receives the message, if it is

connected by a path with at least one node that has already received it. For temporal networks, the concept of temporal path is defined by following "flooding mechanism":

At a certain moment t, a node b receives a message from node a if and only if there is a path between a and b at that moment t. We define a *temporal path* between nodes a and b by a flooding concept: a message sent by a at the moment t_0 is received by n_1 at the moment $(t_0 + 1)$; the message sent by n_1 is received by n_2 at the next moment $(t_0 + 2)$; and so on. In general a message sent by n_i is received by n_{i+1} at the next moment $(t_0 + i)$ for $i = 0, 1, ..., d - 1$, where $a \equiv n_0$ and $b \equiv n_d$. *Temporal length* of this path is d, the time required for a message sent by a to be received by b.

In general, between two nodes there are more than 1 *temporal path*. At this point we can define *temporal distance*.

Definition 2. *Temporal distance* $d_{ij}(t_1, t_2)$ *between nodes* i *and* j *is the smallest length among all the temporal paths between* i *and* j *in the time interval* $[t_1, t_2]$.

In the following example descriptive details are given of how temporal distance is determined. To find a temporal distance between two nodes a searching algorithm is used, which is very descriptive. For larger temporal networks, the more efficient method of Lamport clocks [Lam78] is used, as shown in Section 2.5.

Example 1. *Temporal network in* 5 *time moments as depicted in Figure 2.1. Calculate temporal distance between nodes* 1 *and* 5.

Solution. A straightforward method of calculating temporal distance between nodes 1 and 5 is to perform a Depth First Search (DFS) algorithm.

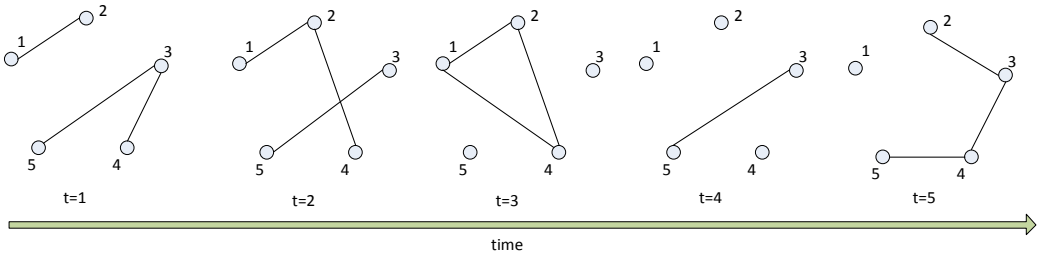


Figure 2.1: Temporal Network

We have two structures R and D for each node i: is reached (true/false) and the distance from sender node at the moment (an integer). In the initial moment, all the distances are set to infinity and all R variables to $false$, except the sender's which distance is 0 and its R is $true$.

At the first two moments nodes 2 and 4 are reached and the arrays D and R are updated. At the third we have a link between nodes 3 and 5. However, 3 and 5 have not received

Table 2.1: D and R values per node in d_{15} calculation

$t=1$	1	2	3	4	5
D	0	1	∞	∞	∞
R	T	T	F	F	F

$t=2$	1	2	3	4	5
D	0	1	∞	2	∞
R	T	T	F	T	F

$t=5$	1	2	3	4	5
D	0	1	5	1	5
R	T	T	T	T	T

information yet, so they are not connected with discovered nodes and the arrays R and D remain as previously. There are no updates at the third and fourth moments. In the fifth moment, nodes 3 and 5 are discovered by 2 (or 4) and we have the final distances. The distance between nodes 1 and 5 is $d_{15} = 5$. ■

Temporal distance is a metric between two nodes. Average temporal path length extends the concept of temporal distance for the whole network. It gives averaged information for the network behaviour in general.

Definition 3. *In a given time interval* $[t_1, t_2]$ *the average temporal distance over all pairs of nodes defines the temporal length*

$$L_G(t_1, t_2) = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}(t_1, t_2)$$

of a temporal network $G(t)$ *in time interval* $[t_1, t_2]$.

For average temporal path length the following theorem holds.

Theorem 1. *In temporal network* $G(t)$*, considered in time interval* $[t_1, t_2]$*, average temporal length* $L_G(t_1, t_2) \geq 1$.

Proof. It follows from the fact that temporal distance $d_{ij}(t_1, t_2) \geq 1$, $i \neq j$. $L_G(t_1, t_2)$ is averaged sum over all pairs of nodes. □

The last means that average temporal length has a lower bound which can be obtained if and only if we have a connected network in the first moment of temporal network. The lower average temporal length means that there is good communication between the entities and the information will be transferred more quickly than where the average temporal length is a larger value. In the case when is possible to transmit a message between each two nodes, average temporal length is a finite value. In the case where during the time it is not possible to spread the message between two particular nodes i and j temporal distance is not defined. In this case, we accept that temporal distance is infinity. If at least one pair of nodes has temporal distance infinity, average temporal length also becomes infinity. This is noted as a drawback of temporal length as a metric and in this case we can not characterize the whole temporal network, even the rest part of the temporal network (except disconnected pairs) has a short temporal length.

2.2.3 Temporal Efficiency

Temporal efficiency is another metric that resolves the problem with disconnected pairs. In order to resolve those cases it is better to consider inverse value of temporal efficiency. This means that a smaller value of temporal length results in larger value for the inverse function. The later clarifies the meaning of the term *efficiency*.

Definition 4. *In the time interval* $[t_1, t_2]$*, temporal efficiency between two different nodes* i *and* j *is the inverse value of their temporal distance* $e_{ij}(t_1, t_2) = \frac{1}{d_{ij}(t_1,t_2)}$.

Temporal efficiency is a local metric for one particular pair of nodes and it differs from temporal length because it is ranged from both sides. For temporal efficiency we have upper and lower bounds. Based on the definition and the fact that $d_{ij} \geq 1$, temporal length is a value in the interval $[0, 1]$.

Using temporal efficiencies for pairs of nodes one can define average temporal efficiency as a metric for a whole graph.

Definition 5. *Average temporal efficiency is the averaged sum of temporal efficiencies over all pairs of nodes in time interval* $[t_1, t_2]$

$$E_G(t_1, t_2) = \frac{1}{N(N-1)} \sum_{i,j;i\neq j} e_{ij}(t_1, t_2)$$

For the bounds of average temporal efficiency $E_G(t_1, t_2)$ it holds:

Theorem 2. *In the time interval* $[t_1, t_2]$*, average temporal efficiency is in the range* $0 \leq E_G(t_1, t_2) \leq 1$*. Left equality holds if and only if there are no links in the temporal network and all the nodes are isolated during the whole period* $[t_1, t_2]$ *and the right equality holds if and only if the temporal graph is connected in the first time moment.*

Proof. Based on the fact that $e_{ij}(t_1, t_2) = \frac{1}{d_{ij}(t_1,t_2)} \leq 1$:

$$\begin{aligned} E_G(t_1, t_2) &= \frac{1}{N(N-1)} \sum_{i,j;i\neq j} e_{ij}(t_1, t_2) \\ &\leq \frac{1}{N(N-1)} \sum_{i,j;i\neq j} 1 \\ &= \frac{1}{N(N-1)} N(N-1) = 1 \end{aligned} \tag{2.1}$$

Equality holds if and only if $e_{ij}(t_1, t_2) = 1$ for $\forall i, j, i \neq j$. This implies $d_{ij}(t_1, t_2) = 1$. If there exists a pair of nodes p, q such that there is no path that connects them in the first moment of temporal network then $d_{pq}(t_1, t_2) > 1$ and $e_{pq}(t_1, t_2) < 1$. The last implies

$E_G(t_1, t_2) < 1$. On the other hand, if all the pairs of nodes are connected in the first moment of temporal network $d_{ij}(t_1, t_2) = e_{ij}(t_1, t_2) = 1$ and $E_G(t_1, t_2) = 1$.

Because $e_{ij}(t_1, t_2) \geq 0$ for $\forall i, j$ it follows that $E_G(t_1, t_2) \geq 0$. The equality holds if and only if $e_{ij}(t_1, t_2) = 0$ for $\forall i, j, i \neq j$ or $d_{ij}(t_1, t_2) = \infty$. If at any moment $t' \in (t_1, t_2]$ there exists a link between some nodes p, q then $d_{pq}(t_1, t_2) = t' - t_1 > 0$. The last means that $e_{pq}(t_1, t_2) = \frac{1}{t'-t_1} > 0$ is a finite value and $E_G(t_1, t_2) > 0$. Hence there are no links in temporal network $G(t_1, t_2)$ at all the time moments. On the other hand, if we have a graph without links at all the moments, then $E_G(t_1, t_2) = 0$. □

When a networked system performs in a "normal" regime, i.e. the temporal network does not sustain structural damage, average temporal efficiency exhibits non-decreasing behaviour. Let us consider temporal efficiency as a function of moment of measuring t, assuming a unique starting moment t_0. Therefore the efficiency function is $E_G(t_0, t)$ in moment t.

Theorem 3. *Average temporal efficiency function $E_G(t_0, t)$ with unique argument t is non-decreasing.*

Proof. The values of the function $E_G(t_0, t)$ at two different moments t_a, t_b $(t_a < t_b)$ are compared. Let us consider the values of the temporal distances $d_{ij}(t_1, t)$ for each pair of nodes i and j. If pair i,j has been connected (directly or via some path) for the first time at the moment t' we have

$$d_{ij}(t_1, t) = \begin{cases} \infty, \ t < t' \\ t' - t_1, \ t \geq t' \end{cases}$$

Therefore, we have several cases

$$\begin{aligned}
&(i) \ \text{if } t_a < t_b < t', \ d_{ij}(t_1, t_a) = d_{ij}(t_1, t_b) = \infty \\
&(ii) \ \text{if } t_a < t' < t_b, \ d_{ij}(t_1, t_a) = \infty \text{ and } d_{ij}(t_1, t_b) = t' - t_1 < \infty \text{ (finite value)} \\
&(iii) \ \text{if } t' < t_a < t_b, \ d_{ij}(t_1, t_a) = d_{ij}(t_1, t_b) = t' - t_1 < \infty \text{ (finite values)}
\end{aligned}$$

Hence, we can conclude $d_{ij}(t_1, t_a) \geq d_{ij}(t_1, t_b)$ or equivalently $e_{ij}(t_1, t_a) \leq e_{ij}(t_1, t_b)$. The last leads us to following result for average temporal efficiency

$$\begin{aligned}
E_G(t_1, t_a) &= \frac{1}{N(N-1)} \sum_{i,j;i \neq j} e_{ij}(t_1, t_a) \\
&\leq \frac{1}{N(N-1)} \sum_{i,j;i \neq j} e_{ij}(t_1, t_b) \\
&= E_G(t_1, t_b)
\end{aligned}$$

Average temporal efficiency is non-decreasing. □

Common behaviour of average temporal efficiency as a time function is depicted in Figure 2.2. One can notice that temporal efficiency requires some transition period before achieving a stationary regime. The length of this period is smaller in "well-connected" temporal networks, where probability of link appearance is higher. Temporal efficiency is an average measure and characterizes a continuously evolving network in the interval $[t-\tau, t]$. Therefore, in order to achieve stationary value for temporal efficiency, the length of the interval τ has to be sufficiently large.

Average temporal efficiency is used as a performance function to define temporal robustness. However, it is worth mentioning that in the case, where temporal network undergo intelligent attack or random error temporal efficiency drops to a smaller value. This is crucial for temporal robustness and more details are given in the next section 2.3.

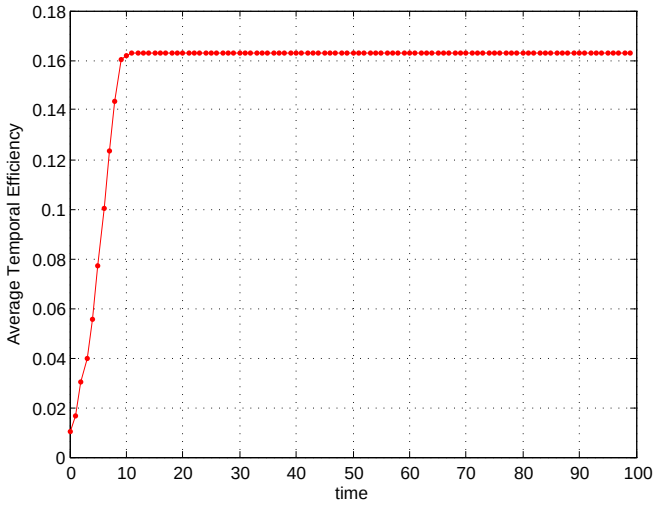


Figure 2.2: Average temporal efficiency (Erdős-Rényi with $N = 100$, $p = 0.01$, $\tau = 100$)

It was mentioned above that not all the nodes in the network are connected from the first moment. This includes direct connection by some link or indirect connections, where there is a path between two certain nodes. We say that a pair of nodes in a temporal network is connected if there is a path at a certain moment between those nodes. This is closely correlated to temporal length and temporal efficiency, where connected pair means that we have finite value for temporal length and non-zero value for efficiency. The number of connected couples is usually compared with the maximum number of couples in a quotient. The maximum number of couples in the temporal network is $\binom{N}{2} = \frac{N(N-1)}{2}$.

2.3 Temporal Network Robustness

A temporal network $G(t)$ can sustain different kinds of damage. Performance is affected by the damage and the system cannot operate at the same level. For instance, a telecommunication provider might be interested in the performance of its network at "peak periods", such as New Year's Eve or national holidays'; similarly an important question for

a military system is communication abilities of smaller and isolated parts, when main installations do not participate in some military training. However, the question of evaluating the *goodness* of the system or ability to perform normally should be generalized in a unique framework. In previous research [BLM+06], different performance metrics have been chosen for robustness evaluation such as the diameter of the network or size of the giant component. In this work, temporal efficiency is taken to be a performance metric. The aim of temporal robustness is to quantify temporal efficiency after sustaining structural damage. After the evaluation of temporal robustness, further actions to improve protection or de-centralizing the network architecture should be conducted.

The name *structural damage* is used as a general term for random errors and intelligent attacks. The change in temporal efficiency is investigated, after occurrence of structural damage. Temporal efficiency before and after structural damage are both taken in a stationary regime for correct robustness evaluation.

Definition 6. *Temporal network robustness is the relative change of efficiency value after a structural damage. If the temporal network after structural damage D is $G_D(t)$ and the efficiency is E_{G_D} temporal robustness is expressed by*

$$R_G(D) = \frac{E_{G_D}}{E_G}$$

where E_G is the efficiency of temporal network $G(t)$ before the damage.

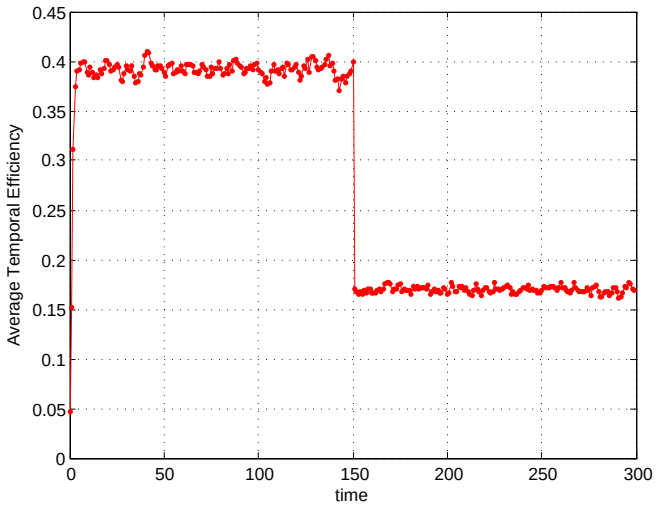


Figure 2.3: Efficiency behaviour under error at the moment $\frac{\tau}{2} = 150$

If structural damage appears, then we have a drop in the temporal efficiency. Common behaviour is depicted in Figure 2.3. Based on Definition 6, the efficiency change $\triangle E(G, D)$

contributes in temporal robustness by the following expression

$$\begin{aligned} R(G,D) &= \frac{E_G - \triangle E(G,D)}{E_G} \\ &= 1 - \frac{\triangle E(G,D)}{E_G} \end{aligned}$$

The value of sliding window τ may significantly change the perception of temporal robustness. Underestimating this factor and considering a scenario, where network sustain structural damage at the moment $\tau/2$ and length of sliding window is insufficiently small τ can result with value of temporal robustness larger than 1. The reason is the fact that time $\tau/2$ is not enough for reaching stationarity and smaller value for efficiency is taken before the damage is sustained. In those cases, efficiency after the damage can achieve greater value. An example of this is given in Figure 4.2 (Chapter 4).

Accordingly, for correct temporal robustness evaluation, both efficiencies before and after the damage have to be taken in stationary regime.

2.4 Error and Attacks strategies

Communications and networked systems experience different types of failures. Apart from the temporal nature of networked system, characterizing of structural damage is also an important issue. Temporal networks sustain random or non-intentional failures in some of the entities (nodes) or could be under intelligent and malicious attacks. This can result in reduced performance, delays in execution of some function or the system might even stop working. The question of whether the system can undergo random failures or intelligent and malicious attacks and still run efficiently becomes crucial. Several structural damage strategies and random error are examined on theoretical models and real data-sets. A general characteristic of error and attacks strategies is that they both target a certain set of nodes. Nodes that are targeted are "isolated" from the remaining part of the network, such that all the links incident to targeted nodes are removed. The number of links that are chosen in both random errors and intelligent attacks is related to *probability of attack*. Higher probability of attack means that more nodes sustain random error or intelligent attacks.

2.4.1 Random Errors

Random error is a type of network damage, where nodes are removed randomly and not related to some static or temporal property. Following random error approach, each node can undergo structural damage with the same probability and this is only related to the probability of error. Finally, the probability of error defines the number of attacked nodes

which turns to be the portion of total number of nodes ($N_{\text{attacked}} = P_{\text{error}} \times N$). Probability of error 0 means that there are no nodes that sustain error and the network performs as previously. On the other hand, probability of error 1 results with removing of all the nodes in the network. Temporal efficiency, temporal length and number of connected couples are all equal to 0 and the new temporal network is completely destroyed and non-operational.

The set of nodes that are selected is random and in two different random error simulations the sets of affected nodes are not the same. Therefore, for the final results, including temporal efficiency, temporal length or temporal robustness evaluation, the simulations should be repeated several times, followed by averaging. The effects of random errors and intelligent attacks on theoretical models and real data-sets are discussed separately in Chapter 3 and 5, respectively.

2.4.2 Intelligent attacks

The causes of many real systems' damages are usually well planned and targeted attacks. *Intelligent attacks* are those strategies which consider some specific temporal node property, rather than a to random failures. The knowledge of how well a system copes when the *most important* nodes (also named as *hubs*) are damaged can help in the decision for future protection. Malicious external/internal activity always targets most influential entities. Term most important/influential in this context is used in a way that those are the nodes that have maximum in some feature, such as degree in average, number of contacts and others. I investigate the effect on both theoretical models and data-sets under different types of attacks.

In this direction, the choice of attacked nodes will have significant effect on the robustness value. Our concern is to investigate the effect of *worst case scenario*, where most influential nodes are attacked. *The worst case scenario*, a sorted list of all the nodes by some temporal property is kept and first N_{attacked} are attacked. However, fluctuations between *worst case scenario* and choices of attacking the same number but other nodes is also important. *Robustness range* reports for the fluctuations and interval of values that is expected, when not the first nodes in the list are removed.

Definition 7. *For a particular strategy, the difference in robustness values while attacking most influential and less influential nodes, when the same numbers of nodes are removed is called* robustness range.

Intelligent attacks strategies are detailed in the remaining part of this section. The common point of all the strategies is probability of attack P_{attack}, which determines the number of nodes that are removed by $N_{\text{attacked}} = P_{\text{attack}} \cdot N$.

Probability of attack is used as an argument of temporal robustness in evaluation. The question of *which nodes are removed* is answered by the choice of intelligent attack strategies. The effects on temporal network after a particular intelligent attack or random error appears on theoretical models and real data-sets are discussed in Chapter 4 and 5, respectively. In the following part details of particular strategies are given.

Temporal closeness nodes attack

This attacking strategy is closely related to *temporal closeness centrality* of the nodes. Closeness centrality is node's property. It has been initially defined for static graphs.

Definition 8. *In a static graph, closeness of a node i is defined as the average shortest path length to all the remaining nodes in the graph.*

This means that one node with a smaller value for closeness centrality is more central than a node with higher closeness centrality value. Based on the idea of static graphs, the concept is extended in temporal networks.

Definition 9. *In a given time interval $[t_1, t_2]$, temporal closeness centrality $C_i(t_1, t_2)$ of a node i is defined as an average sum of all temporal distances between i and other nodes in temporal network. Formally,*

$$C_i(t_1, t_2) = \frac{1}{N-1} \sum_{j;j \neq i} d_{ji}(t_1, t_2)$$

In the definition is assumed the unit distance between every two consecutive moments of the temporal network. This strategy is based on the values of temporal closeness of the nodes. After I have defined temporal closeness centrality for the nodes in temporal network, the strategy can be presented.

Temporal closeness nodes attack strategy picks the nodes with the lowest temporal closeness centrality. These nodes are considered to be more central than other nodes with higher closeness centrality, therefore they are more "important" in the temporal network. According to already calculated temporal closeness for all the nodes we can find temporal robustness. The sorted list of nodes is formed by the value of temporal closeness and the first N_{attacked} are considered for removing, where N_{attacked} is determined by the *probability of attack* ($N_{\text{attacked}} = P_{\text{attack}} \cdot N$). Finally, we can investigate the network behaviour for different values of probability of attack.

In addition, by evaluating *robustness range* the difference in temporal robustness respect to attacked nodes is determined.

Average nodes degree attack

Node's structural property for importance in this strategy is average node's degree. In the static graph degree of a node i is defined as the number of nodes directly connected to node i. However, in temporal networks appearance of a link between pair of nodes is not constant. For temporal networks we can define average degree node.

Definition 10. *In a given time interval* $[t_1, t_n]$ *and temporal network* $\{G(t_j)\}$ $j = 1, 2, \ldots, n$, *average degree of node* i *is the averaged sum over all degrees of* i *in all the static graphs* $G(t_j)$

$$\deg_G\ (i; t_1, t_2) = \frac{1}{N-1} \sum_{j=1}^{n} \deg_{G(t_j)} (i)$$

As in the previous strategy we have a sorted list of average degrees of all the nodes. The last means that we attack first N_{attacked} nodes with highest degree. *Robustness range* is also evaluated considering deviation of robustness, when less important nodes are attacked.

Nodes number of contacts/updates attack

In the network evolution during the time a node can trigger updates in the other nodes. A "contact" can initiate change in some temporal distances. An update of the distance between two nodes i and j is not necessarily caused by a direct link between i and j. For instance, we say that an update of the distance between nodes i and j also happens, if a node i is connected with other node k, and k is connected with node j. In general, if a link is established between two nodes i and j, then it might result with existence of a path between nodes m and n. This is counted as and update for i and j. The nodes i and j may be different from m and n. In this way, we can obtain the total number of updates that each node triggers in other nodes. The nodes that caused highest number of updates are considered to be most active.

2.5 Implementation issues

In the presented Example 1 for temporal distances, we employ Depth First Search algorithm (DFS). In general other searching algorithm for exploring the neighbours of a certain node, such as Breath First Search (BFS) could be used. Although there are significant differences in many aspects, the time complexity of two algorithms is $O(|E| + |V|)$. Therefore, time complexity depends on number of nodes and number of links. In the worst case ($|E| = \frac{n(n-1)}{2}$ where $|V| = n$), it is $O(n^2)$. Performing DFS for all pair of nodes will lead us to complexity $O(n^4)$ in every single moment. This method works fine for a smaller networks with a few nodes.

In the case of large data-sets some other solutions, using the techniques of Lamport [Lam78] clocks are proposed. In the following part are given the algorithms that are used for temporal metrics and temporal robustness.

2.5.1 Algorithm for temporal metrics

For each node i, we keep a structure $clock[i]$. Every clock contains the shortest temporal distances to all the other nodes. At each time moment, for each link, the *clock* structures are updated.

Algorithm 1: Calculating temporal metrics

Input: temporal network $G(N(t), E(t))_t$, where $t \in T$
Output: temporal metrics: Efficiency, Length, % of connected couples
foreach t *in* T **do**
  foreach $edge(node_1, node_2)$ *in* $E(t)$ **do**
    clock[ ][ ] $\longleftarrow$ updateClock($node_1, node_2, clock$);
    clock[ ][ ] $\longleftarrow$ updateClock($node_2, node_1, clock$);
  [eff,len,coup] $\longleftarrow$calculateTemporalMetrics(t);
  Efficiency $\longleftarrow$Efficiency+eff;
  Length $\longleftarrow$Length+len;
  ConCouples $\longleftarrow$ConCouples+$coup$;
Efficiency$\longleftarrow$Efficiency/$\binom{|N|}{2}$;
Length$\longleftarrow$Length/ConCouples;
ConCouples$\longleftarrow$ConCouples/$\binom{|N|}{2}$;

If at a certain time two nodes are in contact with each other, correspondent clocks are updated. In this way, clocks are updated whenever two nodes are in contact with each other. The clocks provide real-time information for connectivity, but also offer a possibility for easier observation of particular contribution in efficiency and robustness of each node. Moreover, random failures and intelligent attacks could easily be monitored in this algorithm. The algorithm for updating time scheduling *clock* structure is as follows:

Algorithm 2: updateClock

Input: nodes: i and j; time: t; array: clock[ ][ ]
Output: array: clock[ ][ ]
foreach $node$ *in* $N(t)$ **do**
  if $clock[i, node] > clock[j, node]$ **then**
    clock$[j, node] \longleftarrow$ clock$[i, node]$;
clock$[j, i] \longleftarrow t$;

From the clock structure, we can calculate correspondent temporal metrics, such as tem-

poral efficiency, temporal average length and portion of connected couples.

Algorithm 3: calculateTemporalMetrics

Input: time: t; arrays: clock[ ][ ]
Output: temporal metrics in a certain moment: eff, len, coup
eff $\longleftarrow$ 0; len $\longleftarrow$ 0; coup $\longleftarrow$ 0;
foreach i *in* $N(t)$ **do**
  foreach j *in* $N(t)$ **do**
    $distance \longleftarrow t-\text{clock}[i,j]+1$;
    eff $\longleftarrow$eff+$1/distance$;
    len $\longleftarrow$len+$distance$;
    coup $\longleftarrow$coup+1 ;

The benefit of this approach is real-time monitoring, in comparison with previously mentioned searching algorithms has just been mentioned. In addition, the Lamport clocks approach is more time efficient. We can see that in each time iteration we have, the algorithm *calculateTemporalMetrics* iterates through each pair of nodes and time complexity is $O(N^2)$. In Algorithm 1, there is a loop through all the edges in the graph. The last brings a time complexity $O(|E|)$, which in the worst case (complete graph) is $O(N^2)$. In each iteration Algorithm 2 is performed twice for end-nodes of a certain edge. It iterates through all the nodes and compares correspondent clocks with input nodes. This results in its own complexity $O(N)$. Finally, the complexity for the Algorithm 1 is given by $O(N \times |E| + N^2)$. Therefore, this algorithm is more efficient than searching algorithms with complexity $O\left((|E| + |V|) \times N^2\right)$. In the worst case the complexity of Algorithm 1 is $O(N^3)$ and again is better than the one of searching (DFS or BFS) with $O(N^4)$.

In addition to evaluation that is based on the algorithm, for Erdős-Rényi temporal random network, theoretical probabilistic solution is given, which is identical with empirical evaluations.

2.5.2 Algorithm for temporal network robustness

The algorithm for temporal network robustness extends that of temporal metrics explored in the previous section. We need to find relative change of temporal efficiency. There are two approaches for how failures appear and efficiency is calculated. For theoretical temporal models, such as Erdős-Rényi and Markov model, we calculate temporal efficiency in stationary regime, before and after the moment of failure, again in stationary regime. Precisely, random error or intelligent attack appears at the middle ($T/2$) of the whole time windows T. The time interval $(0, T]$ should be long enough, such that stationary regime for temporal efficiency is achieved in the intervals $(0, \frac{T}{2}]$ and $(\frac{T}{2}, T]$. For real temporal network, examined in case studies, we cannot distinguish for stationary regime. In this case, efficiency is calculated twice for the whole time interval T. Firstly, we

calculate efficiency without structural damage. Secondly, structural damage (error or attack) appears at the very first moment before the start of the second evaluation and we calculate the efficiency for damaged temporal network.

In random error strategy, the nodes that are attacked are chosen randomly. In intelligent attacks we use a structure for recording nodes importance (average degree node, average temporal closeness or number of contacts/updates). Nodes with highest importance are removed. For robustness range the difference in temporal robustness is evaluated, when most and less important nodes are attacked, respectively.

Chapter 3

Models

Apart from real world networks and extracted data-sets, I use random network models: a set of artificially generated topologies. However, those models aim to reflect some features, observed in the real world, such as randomness or preferential attachment to high degree nodes. Analyzing theoretical models offers a possibility to investigate particular property on a very effective way, because temporal model are flexible with size of the networks (e.g. number of nodes), instead of real data.

Network models are not new concepts. They were first studied for the first time by Erdős and Rényi [ER59], who proposed two popular models: $G_p(N)$ with fixed probability p, and $G_r(N, L)$ with fixed number of links L. Other examples are Watts and Strogatz model for small-world graph [WS98]; Barabási-Albert preferential attachment to high degree nodes [AB02]; power- law and scale-free graphs [FFF99]. Current research mostly focuses on the static theoretical models, neglecting dynamics and network evolution.

In this chapter, three classes of temporal network models are presented: Erdős-Rényi model, Markov model and mobility models. Previously defined temporal metrics, including temporal efficiency and average temporal length are examined using Erdős-Rényi, Markov and mobility temporal model for better understanding of the metrics.

3.1 Erdős-Rényi model

In this section the concept of Erdős-Rényi random model is presented, firstly starting with static approach. Temporal Erdős-Rényi random model is an extension of the static model.

Definition 11. *Erdős-Rényi random graph $G_p(N)$ has N nodes, where a link between each pair of nodes appears independently with fixed probability p.*

The model is simple, the number of the nodes is constant N and determination of node appearance does not depend on other links or nodes and it is only determined by probabil-

ity p. This model is well studied and most of the features are already known, particularly, degree distribution, which is the probability that node has a degree equal to k is given by [ER60, VM06]

$$\Pr\left[\deg\left(i\right)=k\right]=\binom{N-1}{k}p^{k}\left(1-p\right)^{N-1-k}$$

At this point, we can extend the definition of static to temporal Erdős-Rényi random graphs.

Definition 12. *In time interval $[t_1, t_n]$, Erdős-Rényi temporal random graph $G_p(N,t)$ is a sequence of Erdős-Rényi static random graphs $G_p(N,t_i)$ taken in the moments $t_i \in [t_1,t_n]$, $i=1,2,\ldots,n$.*

Based on the independence of previous state, we can conclude similarly for degree distribution function.

Lemma 4. *Average degree node in Erdős-Rényi temporal random graph is $(N-1)\,p$.*

Proof. Degree distribution of Erdős-Rényi temporal random network is time independent, hence it is the same as in static graph $\binom{N-1}{k}p^{k}\left(1-p\right)^{N-1-k}$. Based on the expectation formula for average degree node we have

$$\begin{aligned}
\overline{\deg_t\left(i\right)} &= \sum_{k=0}^{N-1} k\Pr\left[\deg\left(i\right)=k,t\right]=\sum_{k=0}^{N-1} k\binom{N-1}{k}p^{k}\left(1-p\right)^{N-1-k}\\
&= \sum_{k=1}^{N-1} k\frac{\left(N-1\right)\left(N-2\right)!}{k!\left(N-1-k\right)!}p\cdot p^{k-1}\left(1-p\right)^{N-1-k}\\
&= \left(N-1\right)p\sum_{k=1}^{N-1}\frac{\left(N-2\right)!}{\left(k-1\right)!\left(N-2-\left(k-1\right)\right)!}p^{k-1}\left(1-p\right)^{N-2-(k-1)}\\
&= \left(N-1\right)p\sum_{k=0}^{N-2}\frac{\left(N-2\right)!}{k!\left(N-2-k\right)!}p^{k}\left(1-p\right)^{N-2-k}\\
&= \left(N-1\right)p\left(p+1-p\right)^{N-2}=\left(N-1\right)p
\end{aligned}$$

□

Because of the simplicity of Erdős-Rényi model, it can be analyzed by both theoretical/probabilistic analysis and simulation. The metrics defined in the Chapter 2 are targeted. They consider information spreading and transferring a single message across the network. The number of connected couples in each moment is determined by probability that nodes have received a message by moment t. Temporal length and efficiency can also be obtained based on this result. Let us consider random variable N_t for a number of nodes that have received a single message. The message is owned by only one node at the moment $t=0$. The following Lemma 5 [SLM+11] holds.

Lemma 5. *The probability density function of a discrete event N_t is determined by the following recursive formula*

$$Pr[N_t = k] = \begin{cases} 0, \textit{ for } k > 1 \textit{ and } t = 0 \\ 1, \textit{ for } k = 1 \textit{ and } t = 0 \\ \sum_{m=1}^{k} \binom{N-m}{k-m} \left(1-(1-p)^m\right)^{k-m} (1-p)^{m(N-k)} Pr[N_{t-1} = m], \textit{ o/w} \end{cases}$$

Proof. By definition follows that exactly one node owns the message in the initial moment and $\Pr[N_0 = 1] = 1$. For $t > 0$, one can consider the formula for total probability or decomposition

$$\Pr[N_t = k] = \sum_{m=1}^{k} \Pr[N_t = k | N_{t-1} = m] \Pr[N_{t-1} = m] \tag{3.1}$$

Let us denote by p_m the probability that a single node without a message receives it, if exactly m other nodes already have the message. The message will not be delivered only in the case when all m nodes that have the message do not establish a link. The probability of this event is $(1-p)^m$. If a link is established by at least one node the message will be received. Therefore, $p_m = 1 - (1-p)^m$.

Moreover, in order exactly k nodes to receive the message at the moment t, $(k-m)$ additional nodes have to receive the message if m others already owned it. This event is not time dependent. We have $(N-m)$ nodes without a message and exactly $(k-m)$ have to receive it, but $(N-k)$ still not. Therefore, binomial rule holds

$$\begin{aligned} \Pr[N_t = k | N_{t-1} = m] &= \binom{N-m}{k-m} p_m^{k-m} (1-p_m)^{(N-k)} \\ &= \binom{N-m}{k-m} \left(1-(1-p)^m\right)^{k-m} (1-p)^{m(N-k)} \end{aligned} \tag{3.2}$$

Using (3.2) in (3.1), results with

$$\Pr[N_t = k] = \sum_{m=1}^{k} \binom{N-m}{k-m} \left(1-(1-p)^m\right)^{k-m} (1-p)^{m(N-k)} \Pr[N_{t-1} = m]$$

□

Furthermore, we can obtain the probability that a node receives a message before t time steps.

Theorem 6. *Probability R_t that a randomly chosen node has received a message before the moment t is given by the expression*

$$R_t = \frac{1}{N-1} \sum_{k=1}^{N} (k-1)\, Pr[N_t = k]$$

where $Pr[N_t = k]$ is determined in Lemma 5.

Proof. The probability that k nodes have the message at the moment t is $\Pr[N_t = k]$. Except the source node all the other nodes $(k-1)$ can be chosen randomly out of $(N-1)$ other nodes, by probability $\frac{k-1}{N-1}$. Because k can vary in the set $\{1, 2, \ldots, N\}$ the result follows immediately. □

Because $\Pr[N_0 = 1] = 1$, $R_t = 0$. In addition we can find the probability that a node is reached exactly at time t.

Corollary 1. *Probability that temporal distance is equal to t is given by*

$$d_t = Pr[l = t] = R_t - R_{t-1}$$

Temporal efficiency, temporal length and ratio of connected couples can now be obtained calculating the expectations $E\left[T^{-1}\right], E\left[T^{1}\right]$ and $E\left[T^{0}\right]$, respectivelly, for random variable t, using the probability d_t.

1. Temporal efficiency

$$\begin{aligned} E_G &= E\left[T^{-1}\right] = \sum_{t=1}^{T} t^{-1} \cdot \Pr\left[l = t\right] \\ &= \sum_{t=1}^{T} \frac{1}{t} \cdot d_t = \sum_{t=1}^{T} \frac{R_t - R_{t-1}}{t} \end{aligned}$$

2. Ratio of connected couples

$$\begin{aligned} C_G &= E\left[T^{0}\right] = \sum_{t=1}^{T} t^{0} \cdot \Pr\left[l = t\right] = \sum_{t=1}^{T} 1 \cdot d_t \\ &= \sum_{t=1}^{T} \left(R_t - R_{t-1}\right) = \sum_{t=1}^{T} R_t - \sum_{t=1}^{T} R_{t-1} \\ &= \sum_{t=1}^{T} R_t - \sum_{t=0}^{T-1} R_t = R_T - R_0 = R_T \end{aligned}$$

3. Temporal length

$$\begin{aligned} L_G &= \frac{E[T]}{C_G} = \frac{1}{C_G}\sum_{t=1}^{T} t^1 \cdot \Pr[l=t] \\ &= \frac{1}{C_G}\sum_{t=1}^{T} t \cdot d_t = \frac{1}{C_G}\sum_{t=1}^{T} t\,(R_t - R_{t-1}) \end{aligned}$$

Thus we have the theoretical derivation for temporal efficiency, temporal length and portion of connected couples.

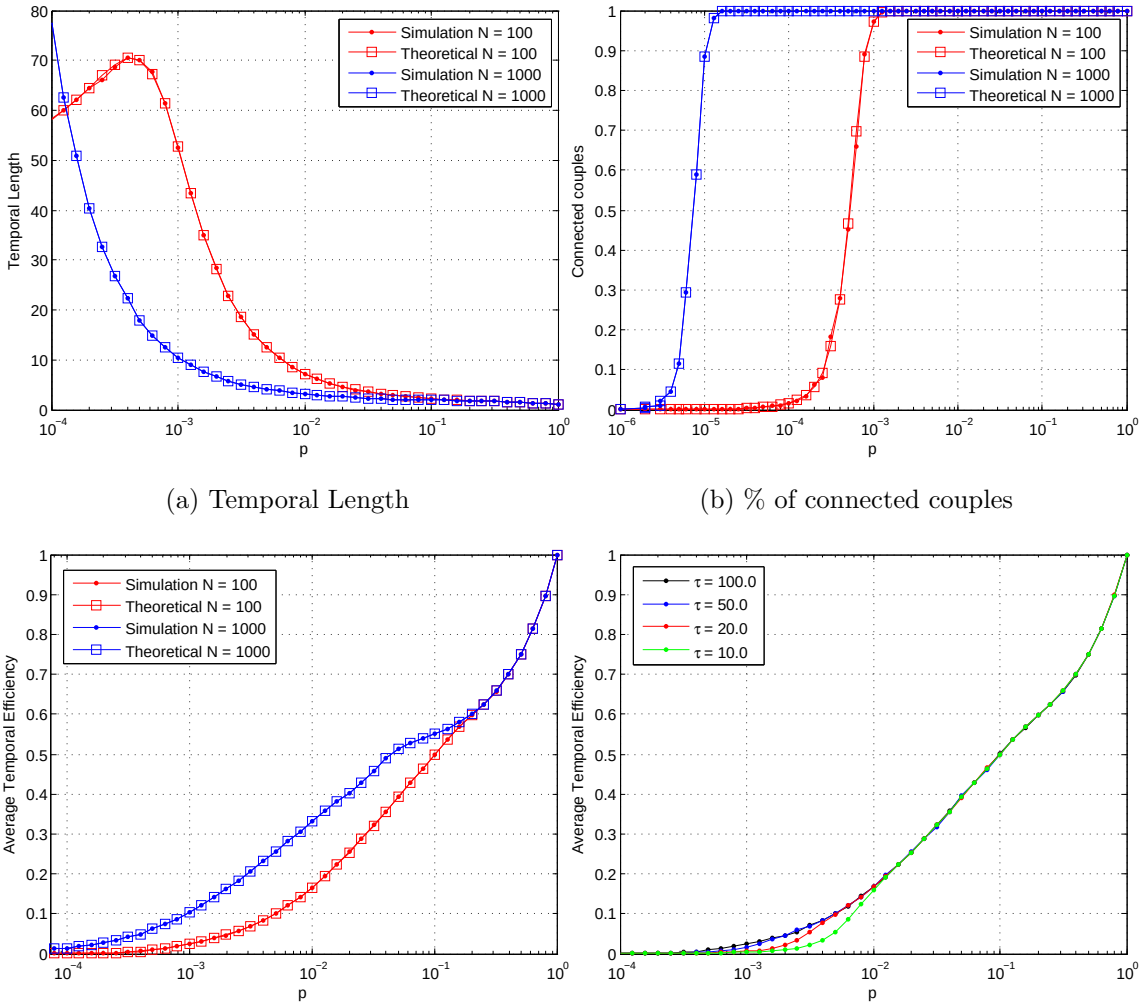
Empirical evaluation

The theoretical results for Erdős-Rényi model are confirmed by empirical simulations. The simulations are in full agreement with the theoretical results and in addition, offer a better understanding of temporal metrics for different values of probability of link appearance p. Erdős-Rényi model curves for temporal metrics as a function of different probability of link appearance p are depicted in Figure 3.1. In particular, temporal length is given in Figure 3.1a; for very small values of p not all pairs of nodes are connected and by increasing p, the number of connected pairs increases faster whereas the distance between already connected pairs decreases. However, after most of the pairs are connected and temporal length has reached the maximum, average temporal length decreases. Both temporal efficiency and the ratio of connected couples show increasing behaviour, because the probability that a particular pair is connected is larger for greater values of p. This is given in Figures 3.1c and 3.1b.

In Figures 3.1a,3.1b and 3.1c the effect of different number of nodes on temporal metrics is also considered. The figures compare $N = 100$ and $N = 1000$. For portion of connected couples (Figure 3.1b), a higher number of nodes means that more links have to be established for all the pairs to be connected. Therefore for $N = 1000$ all pairs will be connected for larger p than it is for $N = 100$.

In the case of temporal length (3.1a), small p means that distances between pairs are long and average temporal length is higher. The connectivity is increased for higher p, it means that temporal distances are shorter and average temporal length decreases. The exception of this behaviour are extremely small values of p (for $N = 100$, $p < 10^{-3}$). For these values, temporal length is taken as an average sum only of the connected couples. The increasing behaviour is because most of the nodes' pairs become connected by increase of p; their temporal distance is higher, but finite and they are included in the average sum. Therefore, they contribute in the increase of the temporal length. After the maximum is reached, temporal length decreases for higher p. In Figure 3.1a, this is shown only for $N = 100$. Higher N means that maximum for temporal length is reached for smaller p.

On the other hand, temporal efficiency (3.1c) increases faster for smaller values of p because the number of possible links increases by N^2 which means more possibilities for a new paths. This results in increased efficiency for larger values of N, when p is small. For larger values of p all pairs of nodes are well-connected and temporal efficiency is not related to number of nodes N.



(a) Temporal Length

(b) % of connected couples

(c) Average temporal efficiency

(d) Average temporal efficiency for different τ

Figure 3.1: Temporal Metrics

The effect of interval length τ is also examined. In Figure 3.1d we can see temporal efficiency curves for different values of the interval τ as a function of probability p. The influence of interval τ is not significant, as long as it lasts long enough to establish a path between all pairs of nodes. This means that influence of temporal efficiency decreases as τ increases, because all paths that have a length longer than τ are not considered, but their contribution is not significant for sufficiently large values of τ, because it contributes to temporal efficiency by the inverse value of temporal distance. Therefore, longer paths than τ can effectively be omitted. In addition, this will speed up the computation.

3.2 Markov model

Temporal correlations and time dependencies at previous moments in temporal network are not provided in Erdős-Rényi model. Markov temporal model provides time dependency on previous state of the link. It is a more general, within which Erdős-Rényi is a special case. The model is based on Markov process evolution.

Let us consider the states of a certain link. Depending on the presence or absence of a link we can clarify two possible state: ON and OFF. Now, considering the current state of a link, it can keep its state or can change it. In this way we can consider a two-state Markov process. The transitions occur with certain probabilities. More precisely we denote a transition probability p that a link present at the moment t will not appear at the moment $(t+1)$; and transition probability q that a link will be added at the moment $(t+1)$, if it was not present at the moment t. Therefore, the probability that the link will remain in the temporal network at the moment $(t+1)$ is $(1-p)$ and the probability that the link will keep absence state at the moment $(t+1)$ is $(1-q)$. The transitions are given on in diagram 3.2.

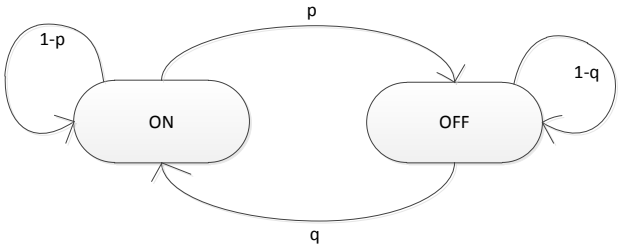


Figure 3.2: Markov model transition states

According to the simple model an the formula for total probability we can calculate the probability for both states ON and OFF

$$\begin{aligned} \Pr[ON] &= \Pr[ON|OFF]\Pr[OFF] + \Pr[ON|ON]\Pr[ON] \\ \Pr[OFF] &= \Pr[OFF|OFF]\Pr[OFF] + \Pr[OFF|ON]\Pr[ON] \\ \Pr[ON] + \Pr[OFF] &= 1 \end{aligned}$$

Switching to probabilities $p, q, (1-p), (1-q)$

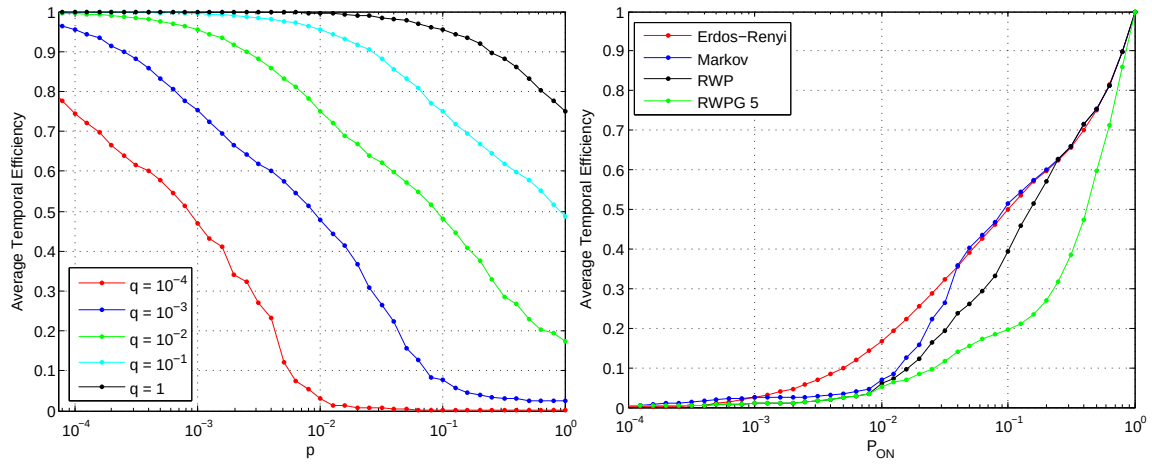
$$\begin{aligned} \Pr[ON] &= q\Pr[OFF] + (1-p)\Pr[ON] \\ \Pr[OFF] &= (1-q)\Pr[OFF] + p\Pr[ON] \\ \Pr[ON] + \Pr[OFF] &= 1 \end{aligned}$$

This gives us

$$\begin{aligned} \Pr[ON] &= \frac{q}{p+q} \\ \Pr[OFF] &= \frac{p}{p+q} \end{aligned}$$

It is worth mentioning that $p + q \neq 1$ in general, which gives time correlation with previous state. In a special case where $p+q = 1$ we have $q = \Pr[ON] = \Pr[ON|OFF] = \Pr[ON|ON]$ and $p = \Pr[OFF] = \Pr[OFF|OFF] = \Pr[OFF|ON]$. Hence, there is no time correlation and we have a fixed probability for link appearance q and absence $(1 - q)$ which corresponds exactly to Erdős-Rényi temporal network.

Having two probabilities for link transition allows us to analyze efficiency from different perspectives. In the first case, shown in Figure 3.3a, for a several fixed values of probability q, the curves for efficiency are given as a function of probability p. We have decreasing functions because higher probability p means more intensive transition for link presence to absence. The common point for analyzing all the temporal models is probability of link appearance P_{ON} and their comparison is made in Figure 3.3b. Particularly, in the figure for Markovian model is given the curve for $q = 10^{-3}$ and different values of $P_{\text{ON}} = \frac{q}{p+q}$. Temporal efficiency curves for the Markov model show similar behaviour to that of Erdős-Renyi, but increases more slowly for smaller values of probability of link appearance P_{ON}.



(a) Markov temporal model ($q = 10^{-3}$) (b) Different temporal models

Figure 3.3: Average temporal efficiency for temporal models

The results for temporal robustness and *robustness* range are discussed in Chapter 4.

3.3 Mobility models

This is a group of theoretical models that aims to simulate the behaviour of mobility. Like the Markov temporal model the mobility model preserves time correlation with the previous state. Unlike the Markov temporal model, in this model the probabilities for changing the state from link presence to absence and vice versa are not constant. The model is as follows [MGBM10]:

Let us imagine a grid with dimensions 1000×1000, such that 100 take one coordinate. The nodes are not static: they move from one coordinate to another. Another variable, denoted as communication range r, determines the probability of link appearance P_{ON}. If the Euclidean distance between two nodes is shorter than r at this moment we have a link between those two nodes in the temporal graph. In this way temporal networks are defined, considering all time moments at which nodes move. However, we can distinguish some details in mobility models: only random movements exist without the preference of some nodes or some nodes being treated as more central.

Two classes of mobility models are considered. A node in Random Waypoint Model (RWP) uniformly picks random location and moves towards this location by speed that is randomly and uniformly chosen in the interval $[5 \text{ mph}, 40 \text{ mph}]$. After a node has reached picked destination, it first waits for a randomly chosen number of moments less than 120 seconds and the procedure starts again by picking a destination and appropriate speed. The benefit of the model is that it provides homogeneous spatial mixing among nodes, but randomness may not express all the aspects of mobility behaviour.

In Random Waypoint Group Model (RWPG) there are two types of nodes: group leaders and followers. Denoting the number of group leaders by M, $(N - M)$ followers are assigned to a group with a unique leader. The size of each group is $\frac{N}{M}$ nodes in total, with 1 leader and $\left(\frac{N}{M} - 1\right)$ followers. The movement rule here is that only the leader of a group picks a destination, as in the RWP model. Followers in a group just follow their leader, such each keeping a distance shorter than a given span (e.g. 100 meters).

Figure 3.3b shows a comparison of the RWP model and RWPG with 5 leaders. One can see that temporal efficiencies for both classes of mobility model increase more slowly than Erdős-Rényi, especially the curve for the RWPG model. This is because there are fewer options for temporal paths and most of them traverse through fewer nodes, particularly RWPG "leaders" nodes.

Temporal robustness and *robustness range* are both considered on several RWP and RWPG models under error and different attack strategies.

Chapter 4

Evaluation of Temporal Robustness

The evaluation of temporal robustness on theoretical models is presented in this chapter. I investigate the effects of temporal network robustness as a function of probability of error/attack for different strategies. Random failures and three attacking strategies: *temporal closeness nodes attack*, *average nodes degree attack* and *nodes number of contacts/updates attack* are considered on Erdős-Rényi, Markov and different mobility models for temporal robustness evaluation.

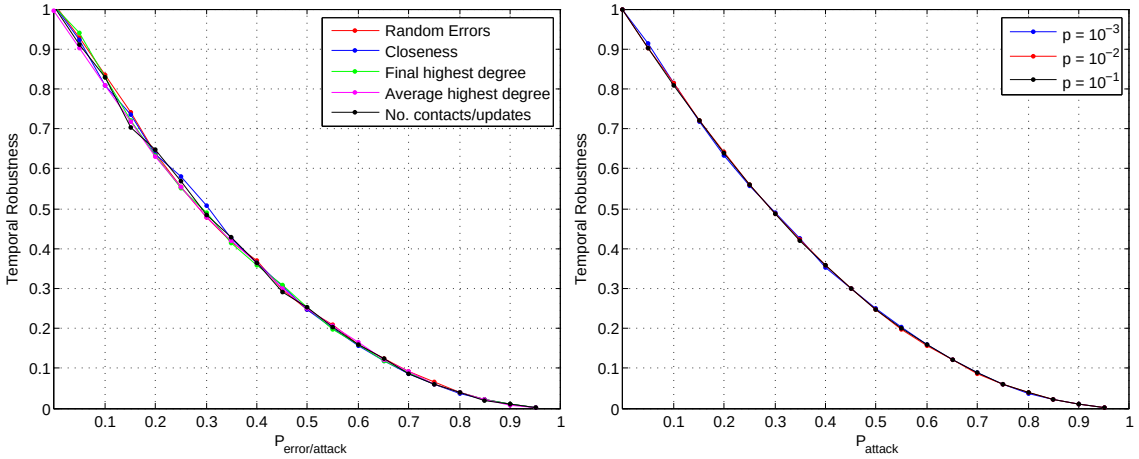
In addition, for each of the strategies, *robustness range* is also examined as a measure of temporal robustness deviation, when less important nodes in the network are attacked, rather than most important. However, in Erdős-Rényi and Markov temporal models the nodes have similar structural properties (e.g. average degree or temporal closeness), which determines their importance. In these cases, this is the reason why the choice of removed nodes does not influence to robustness value and the robustness range is 0. Contrarily, for mobility models we have difference in node's importance, which contributes temporal robustness to decrease faster for intelligent attacks than random errors.

4.1 Erdős-Rényi model

Figure 4.1a gives the values of temporal robustness for Erdős-Rényi temporal network with $N = 100$ nodes and probability of link appearance $p = 10^{-3}$, when it is attacked by different intelligent attacks and random errors. In all the attacking strategies most influential nodes are removed. It can be seen that the value of temporal robustness is the same and irrelevant to the choice of strategy. In addition, identical value for robustness is recorded for all the values of probability of link appearance p. Accordingly, temporal robustness curve is only plotted for one particular p.

Moreover, for each strategy the effect of intelligent attacks and random errors is considered for different Erdős-Rényi temporal networks. This behaviour for *average nodes degree*

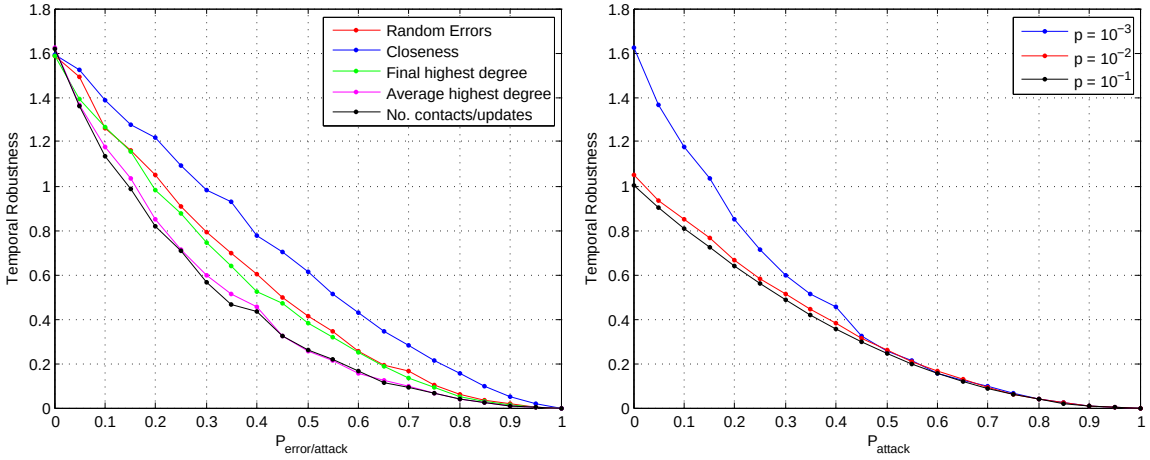
attack strategy is given in Figure 4.1b. Similarly, we have the same robustness value for different Erdős-Rényi temporal networks.



(a) Different strategies ($N = 100$, $p = 10^{-3}$) (b) Average nodes degree attack strategy

Figure 4.1: Erdős Renyi temporal network

It is worth mentioning that previous considerations for temporal robustness are taken in stationary regime for temporal efficiency. In the case where T is small temporal efficiency does not reach stationary value in all the cases. In particular, Erdős-Rényi temporal network with smaller values of p requires larger T to establish the efficiency's stationary regime. In non-stationary cases, temporal efficiency after attack/error can be larger than efficiency before. Therefore temporal robustness is greater than one. This as a weakness of temporal robustness and an example is shown in Figure 4.2 for $T = 20$.



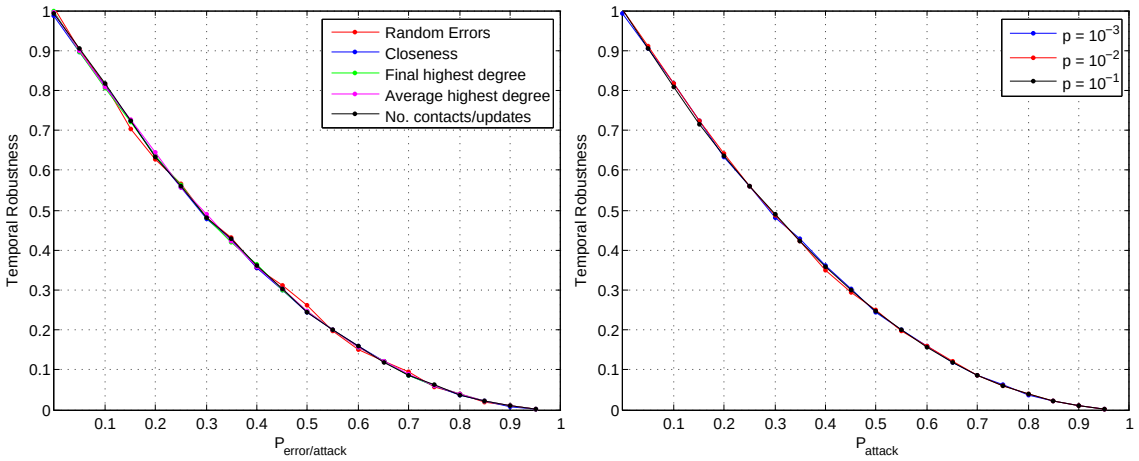
(a) Different strategies ($N = 100$, $p = 10^{-3}$) (b) Average nodes degree attack strategy

Figure 4.2: Erdős-Rényi temporal network (small T effect)

The last means that temporal robustness should be evaluated in cases where efficiency before and after error/attack are both in stationary regime.

4.2 Markov model

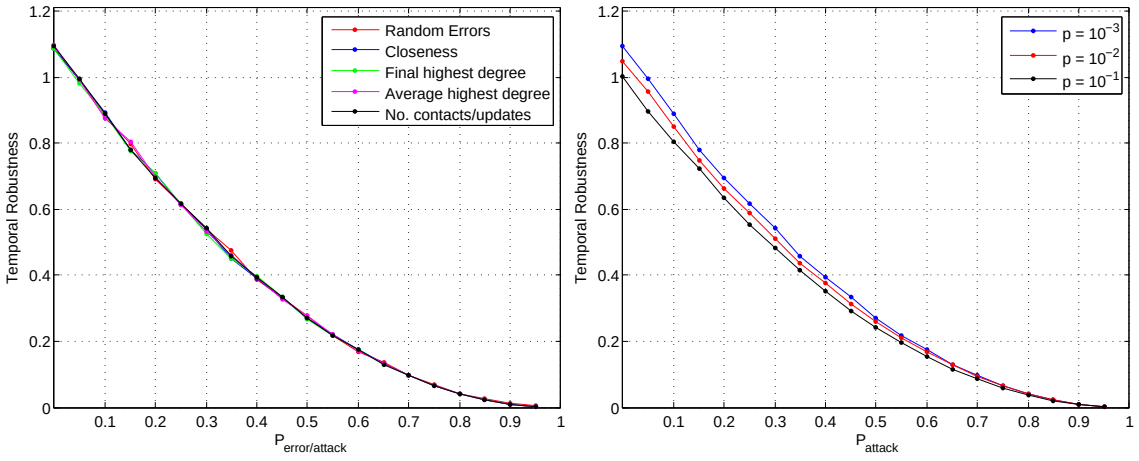
Markov temporal model shows similar features to these of the Erdős-Rényi temporal model. In the evaluation are considered Markov temporal models that differs in probability of link presence $P_{\mathrm{ON}} = \frac{q}{p+q}$ for enough long interval T for reaching efficiency's stationary regime. This is shown in Figure 4.3b for strategy *average nodes degree attack*. We have the same temporal robustness when the model suffers different attacks or error strategies (Figure 4.3a). By this reasoning only one strategy is presented.



(a) Different strategies ($N = 100$, $p = 10^{-3}$) (b) Average nodes degree attack strategy

Figure 4.3: Markov temporal network

Contrarily, when a small T is chosen, the efficiency before and/or after cannot achieve stationary values. This results in robustness greater than 1 in some cases and difference between strategies for a particular model (Figure 4.4).

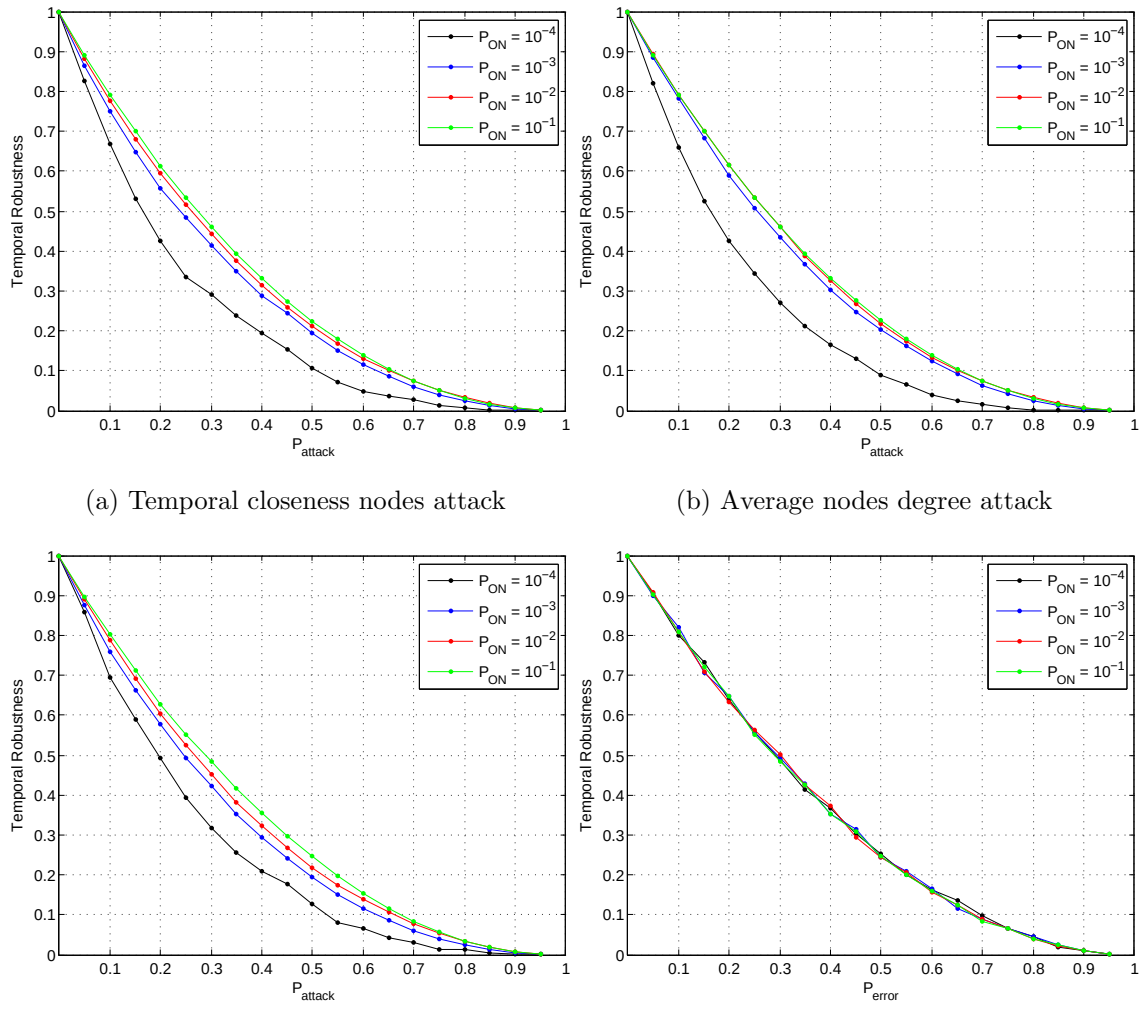


(a) Different strategies ($N = 100$, $p = 10^{-4}$) (b) Average nodes degree attack strategy

Figure 4.4: Erdős Renyi temporal network (small T effect)

4.3 Mobility models

I evaluate temporal robustness for two classes of mobility models. Random Waypoint Model (RWP) and Random Waypoint Group Model (RWPG) with 5 leaders. Descriptions of both models are given in Chapter 3 (section 3.3). In both temporal model classes there are 100 nodes in total. For both classes, different probabilities of link appearance are considered $P_{\mathrm{ON}} = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$.



(a) Temporal closeness nodes attack

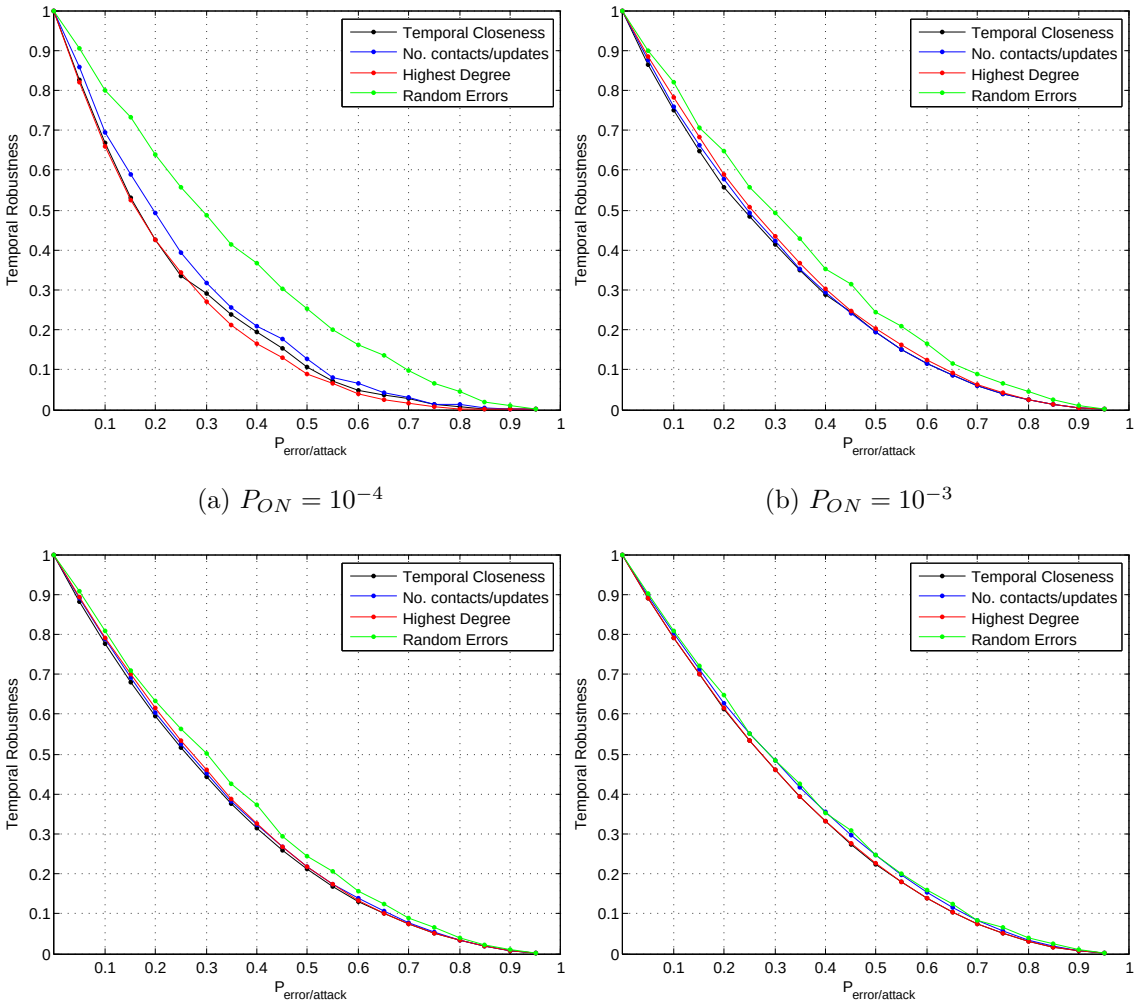
(b) Average nodes degree attack

(c) Nodes number of contacts/updates attack

(d) Random error

Figure 4.5: RWP mobility models

In Figure 4.5 are given 4 figures for a random error and 3 intelligent attack strategies: *temporal closeness nodes attack*, *average nodes degree attack* and *nodes number of contacts/updates attack*. We can see that for all the strategies the model is less affected in well-connected cases (higher P_{ON}). Moreover, for one particular model intelligent attack strategies similarly affect the model. This is shown in Figure 4.6. Temporal robustness for random failures is the same for different RWP mobility models and is less effective than intelligent attacks.



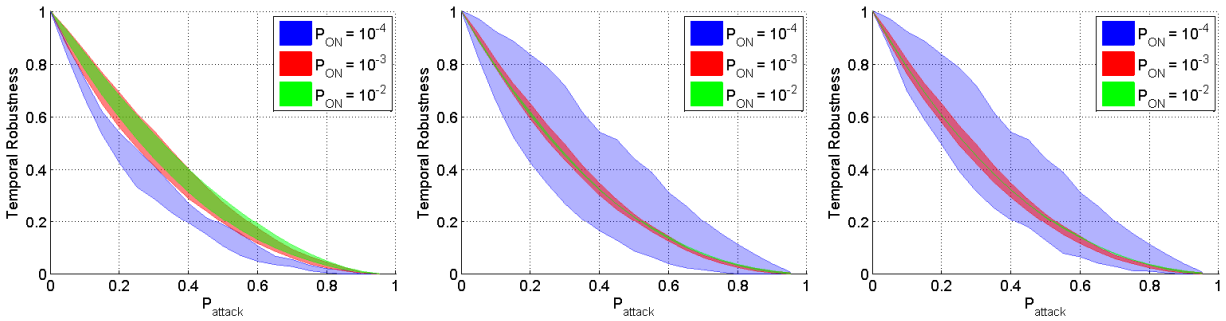
(a) $P_{ON} = 10^{-4}$ (b) $P_{ON} = 10^{-3}$

(c) $P_{ON} = 10^{-2}$ (d) $P_{ON} = 10^{-1}$

Figure 4.6: RWP mobility models

The difference for temporal robustness by the choice of nodes can be evaluated by *robustness range*. In the previous plots the robustness is evaluated when most important nodes are removed. The importance of a node is determined by particular attacking strategy: *temporal closeness nodes attack*, *average nodes degree attack*, *nodes number of contacts/updates attack*. Figure 4.7 expresses the robustness range for different attacking strategies. Robustness range is the area between the temporal robustness curve where less important nodes are attacked and the one when most important are attacked. For smaller values of P_{ON} we have larger *robustness range* area, which indicates that choice of attacked nodes significantly influences on the temporal robustness value. Contrarily, for larger values of P_{ON}, the robustness range area is small, which means that temporal robustness can be determined exactly, irrelevant to the choice of attacked nodes.

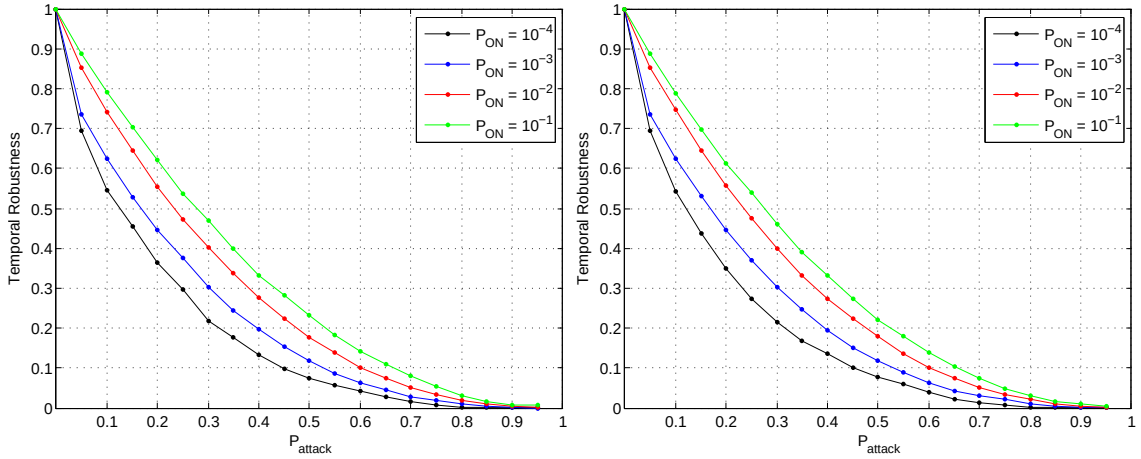
Temporal robustness for RWPG model shows similar behaviour for different probability of link appearance P_{ON}. However, temporal robustness decreases faster than the one of RWP model. Figure 4.8 shows four figures for each attacking strategies and random



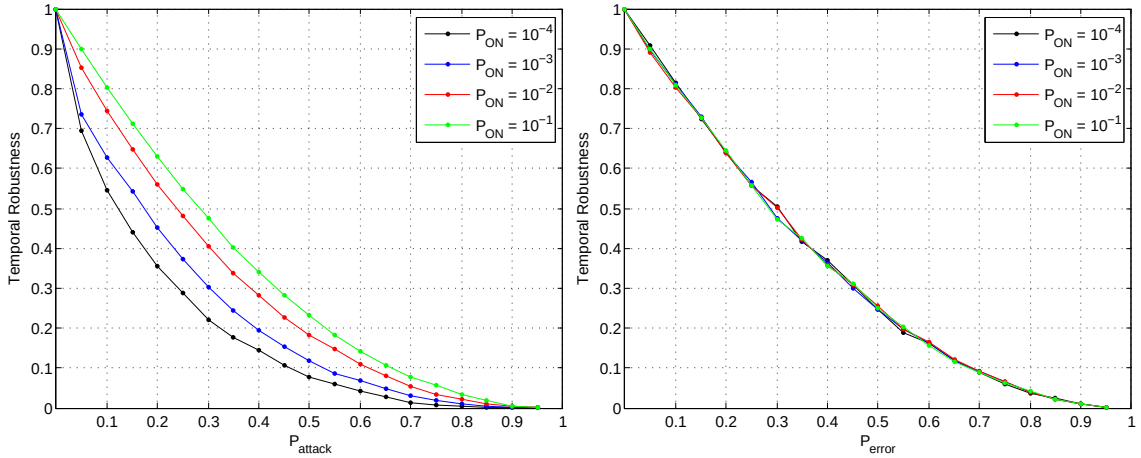
(a) Temporal closeness nodes (b) Average nodes degree (c) Nodes no. of contacts/updates

Figure 4.7: RWP robustness range

errors. In each one, temporal robustness for different P_{ON} is considered.



(a) Temporal closeness nodes attack (b) Average nodes degree attack

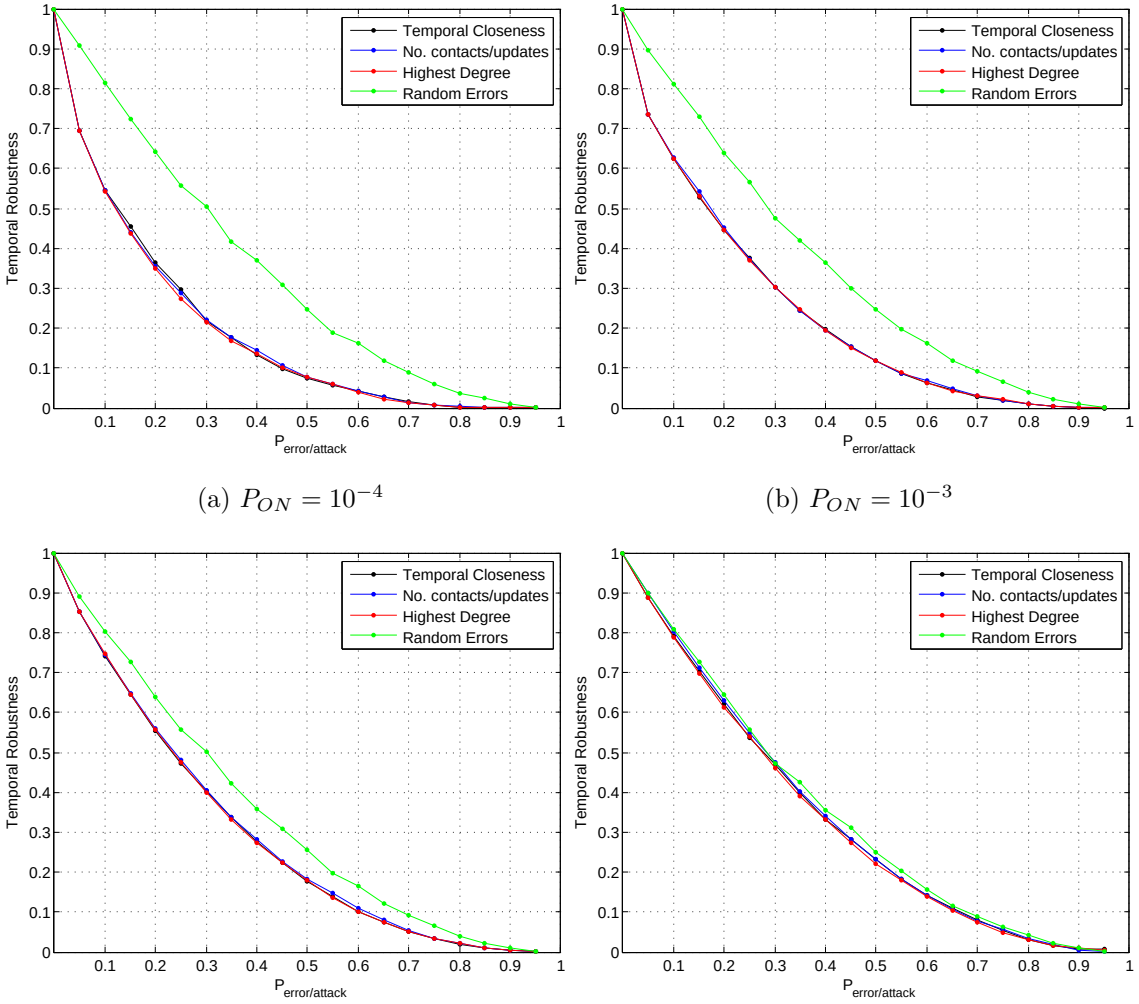


(c) Nodes number of contacts/updates attack (d) Random error

Figure 4.8: RWPG mobility models

In Figure 4.9 we can see that for a fixed P_{ON} the choice of intelligent attack strategy is irrelevant and all are more effective than random errors, particularly for smaller P_{ON}. In

well-connected RWPG ($P_{\mathrm{ON}} = 0.1$), temporal robustness values for intelligent attacks and random failures are levelled.



(a) $P_{ON} = 10^{-4}$ (b) $P_{ON} = 10^{-3}$

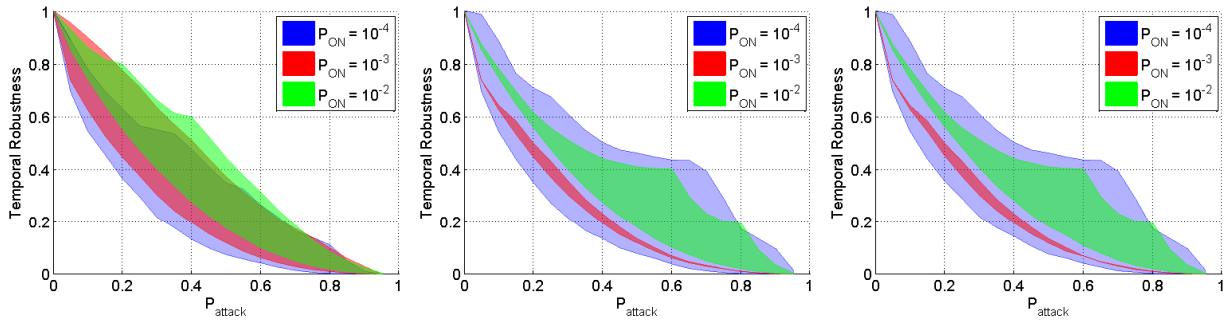
(c) $P_{ON} = 10^{-2}$ (d) $P_{ON} = 10^{-1}$

Figure 4.9: RWPG mobility models

Because of the importance of nodes in the RWPG model, the choice of attacked nodes plays a crucial role and *robustness range* is larger than RWP models. Robustness ranges for different intelligent attacks are shown in Figure 4.10.

4.4 Discussion

Erdős-Rényi temporal network as a structure does not contain predominant nodes or sub-graphs. This is a corollary of the fact stated in Lemma 4, that all the nodes are statistically identical and the average degree of a node is fixed value $(N-1)\,p$ for all the nodes. The last means that in *average nodes degree attack* strategy the choice of the nodes is irrelevant, we have a fixed value for robustness for different choice of nodes and



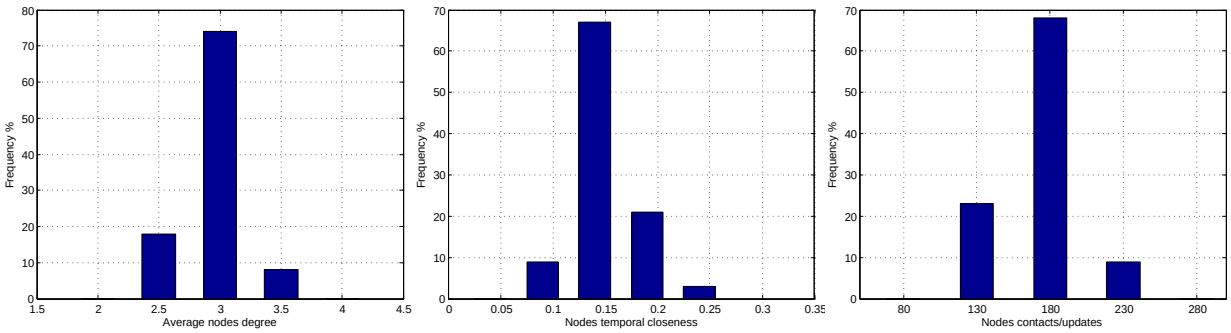
(a) Temporal closeness nodes (b) Average nodes degree (c) Nodes no. of contacts/updates

Figure 4.10: RWPG robustness range

robustness range is zero. For the other attacks strategies, we have that temporal closeness and number of contacts/updates are similar for all nodes, which also implies irrelevance of chosen nodes for the same probability of attack. Therefore, the robustness curve is the same for all the strategies and *robustness range* is 0.

The Markov temporal model differs from Erdős-Rényi because we have transitional probabilities from link appearance to absence. As it is shown in section 3.1, Chapter 3, for different values of p we have different temporal efficiency. However, relative changes of temporal efficiency, after temporal network sustains random error or intelligent attack are the same, which results in the same value of temporal robustness.

In Figure 4.11 are given the histograms for average degree, temporal closeness and number of contact updates in the Markov temporal network. We can see that temporal properties for most of the nodes are similar. This is the main reason for irrelevance of nodes choice and same curve for all random error and attacks strategies. About more than 60% of the nodes have similar values of temporal properties and the remaining have closer values to this.



(a) Average nodes degree (b) Temporal closeness (c) Nodes no. of contacts/updates

Figure 4.11: Histogram for nodes temporal properties (Markov model)

In both Erdős-Rényi temporal network and Markov models we calculate the robustness for small T, when stationary regime for efficiency is not achieved. For temporal models with

lower connectivity (small p or P_{ON}) a longer interval T is required to achieve stationary regime and all pairs to be connected. Otherwise temporal robustness cannot be measured properly and we have value greater than 1 for temporal robustness.

For a fixed model in both classes (RWP and RWPG), when most important nodes are attacked, the robustness is similar for different attacking strategy, because similar nodes are picked. In the other words similar nodes are most important by all the attacking strategies. Mobility models with smaller P_{ON} are more affected by all the strategies, because some crucial temporal distances are more likely to be removed than for larger P_{ON}.

The difference between mobility models has an effect in *robustness range* because leaders in the RWPG model are important hubs and their removal influences on temporal robustness more than on the RWP mobility model, where there are no leading nodes. This results in larger robustness range area for the RWPG mobility model.

Chapter 5

Case studies

This chapter presents two case studies on temporal robustness. Considered data is taken from two real systems. The first case study uses the data collected from taxi cabs that have been moving intensively through whole area around San Francisco, USA. The second case study works with the data gathered from mobile device interactions between participants on scientific conference during the 4 days period. Results for random error and intelligent attacks strategies on both real data sets are given in the section 5.3

5.1 Cab-spotting traces

This case study uses the data from Cab-spotting system for collecting information from taxi movements in San Francisco area. The aim of this project that lasted for 2 years is gathering data about life in the city. All the taxis participating in this project are equipped with GPS sensor devices that periodically send the information for the new location and time stamp to one centralized place. For the case, the data was collected from 24 hours monitoring on 21 May 2008 in a certain area of 20 km $\times$ 20 km around San Francisco. Based on the record, one can reconstruct trajectories of all the taxis. Temporal network is derived on a similar way like mobility models, choosing a communication range of 200 m, which is a common distance for WiFi taxi devices [PSDG09]. Time granularity for sampling is 1 second. In this manner, communication is recorded in a discrete moments. The temporal network is reconstructed by interpolating data in all the moments during the day. It is reported participation of 488 nodes and temporal network consists of about 86,400 graphs. On average duration of a contact lasts for 2 minutes and inter-contact time is 2 hours 30 minutes [SLM$^+$11]. In the following text, the data set will be called shortly *Cab-spotting*.

5.2 INFOCOM traces

The data were collected over 4 days between April 24^{th} and April 27^{th} at the IEEE INFOCOM 2006 conference in Barcelona, Spain. Participants in the experiment were 70 students and researchers who were attending to student workshop. The aim of the experiment was to investigate communication between participants on a particular event. All the participants were equipped with mobile communication devices iMotes [SGC+09], presented in the Appendix A. In total 78 iMotes mobile devices were used by participants and additional 20 stationary iMotes were deployed. Stationary iMotes devices have more powerful battery and extended radio range. The wireless range of mobile iMotes is about 30 meters and that of stationary devices is about 100 meters [SGC+09].

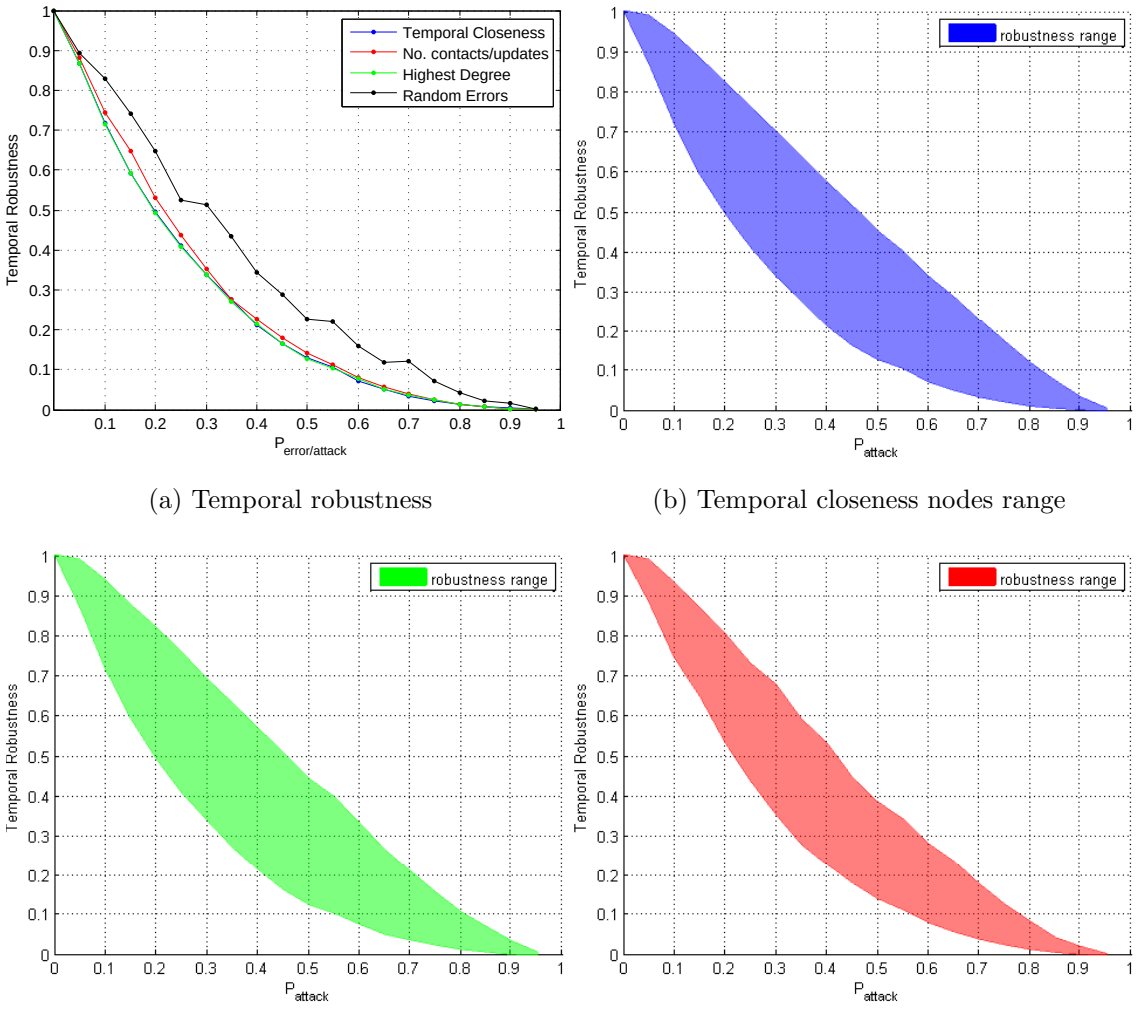
During the 4 days experiment participants were asked to bring the devices which recorded a communication in the range by both mobile and stationary devices. However, the intensity of communication is not the same all the time during the conference. In the overnight periods during the conference's days we have less intensive communication. During peak periods, before and after meetings and discussion sessions, more intensive communication was recorded. At the temporal network, a link appears at a certain time, if there was existence of communication between two entities (participant or stationary device). The time granularity of temporal network is 2 minutes. By the reason of simplicity, the data-set from this case-study is named as INFOCOM.

5.3 Results and discussion

Temporal robustness and robustness range are evaluated for two considered data-sets Cab-spotting and INFOCOM. Figure 5.1 gives four figures for the INFOCOM data-set. Particularly, in Figure 5.1a are plotted robustness curves for 3 intelligent attacks and random error strategies. In the remaining 3 figures (5.1b, 5.1c and 5.1d) robustness range for each particular attacking strategy is given that provides information about robustness deviation by the choice of removed nodes. Temporal robustness can have some value in the robustness range area because it is ranged by robustness curves when less important and most important nodes are attacked, respectively.

Similar analysis is conducted on the Cab-spotting data set. Corresponding temporal robustness curves for different strategies are given in Figure 5.2a and *robustness range* for *temporal closeness nodes attack*, *average nodes degree attack* and *nodes number of contacts/updates attack* in Figures 5.2b, 5.2c and 5.2d, respectively.

Temporal robustness for different temporal networks under *average nodes degree attack* strategy is given in Figure 5.3. It presents a comparison between Cab-Spotting, INFOCOM data-sets and theoretical models: Erdős-Rényi, Markov, RWP and RWPG mobility.



(a) Temporal robustness

(b) Temporal closeness nodes range

(c) Average nodes degree range

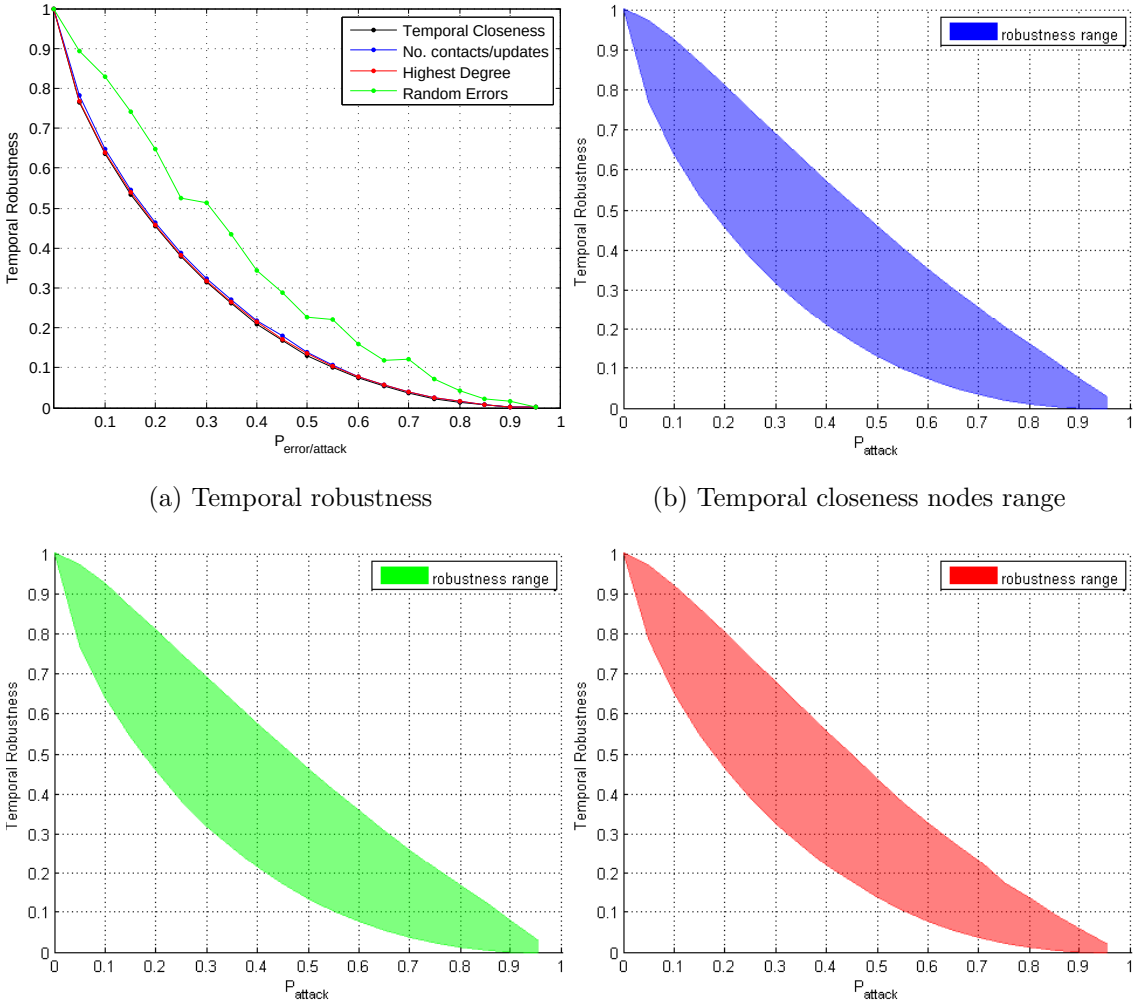
(d) Nodes number of contacts/updates range

Figure 5.1: INFOCOM

Temporal robustnesses for Erdős-Rényi and Markov models are similar and are shown as a unique curve. Because of the difference in temporal robustness for mobility models two curves are given for low and well-connected models ($P_{\mathrm{ON}} = 10^{-4}, 10^{-2}$).

The analysis of the real data-sets shows that different attacking strategies equally affected temporal network. The main reason is that the same nodes are most important according to all three attacking strategies. Because of the presence of more important nodes, the curves for intelligent attacks decrease faster than random error strategy. The robustness range area is similar for all the strategies, which again shows that same nodes are most important by different intelligent attacks.

Comparison of all the real data and temporal network models shows that the effect of attack is most significant in mobility models especially for lower P_{ON}. Temporal robustness also decreases faster in real networks than "balanced" models. As it is shown on Histograms 5.4 and 5.5, nodes in real networks have different temporal properties: average degree, temporal closeness, number of contacts/updates, unlike on the Markov temporal



(a) Temporal robustness

(b) Temporal closeness nodes range

(c) Average nodes degree range

(d) Nodes number of contacts/updates range

Figure 5.2: Cab-spotting

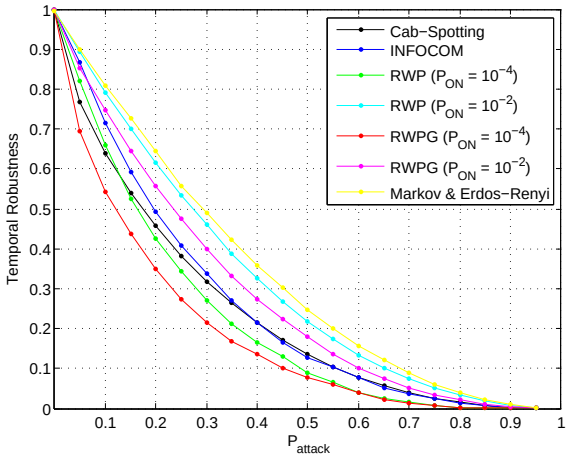
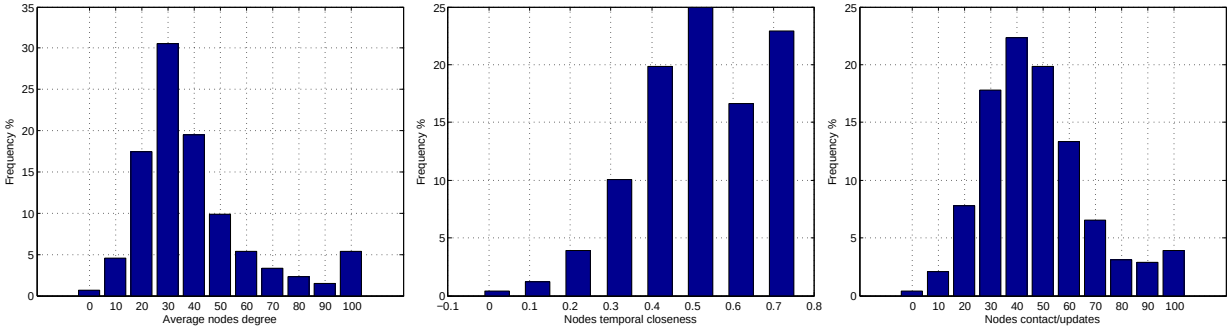


Figure 5.3: Average nodes degree attack strategy

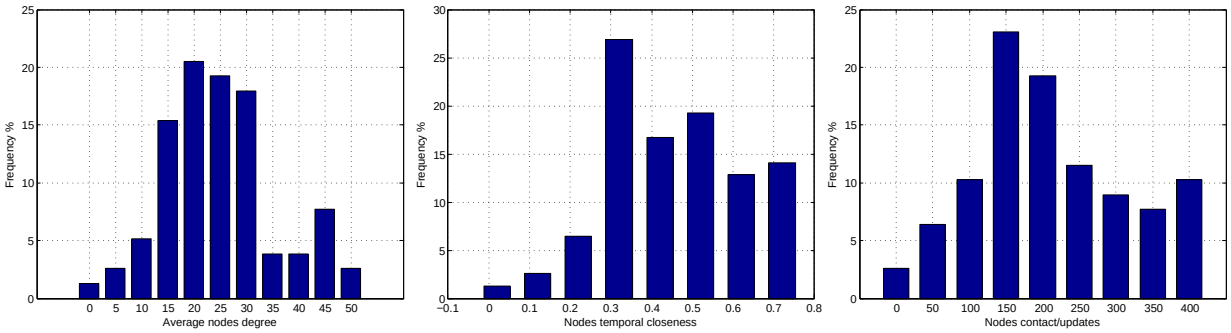
model (Histogram 4.11). Some nodes are more dominant than others in both Cab-spotting and INFOCOM temporal networks. Moreover, the goups of nodes with similar values of

temporal properties is less than 30% of all the nodes in all 3 strategies. Therefore real networks are more affected by intelligent attacks than balanced Erdős-Rényi and Markov model.



(a) Average nodes degree (b) Temporal closeness (c) Nodes no. of contacts/updates

Figure 5.4: Cab-spotting nodes temporal properties



(a) Average nodes degree (b) Temporal closeness (c) Nodes no. of contacts/updates

Figure 5.5: INFOCOM nodes temporal properties

Having a significant robustness range area and sensitivity of intelligent attacks, real temporal networks can be related with mobility models. More precisely, comparing robustness range areas, both INFOCOM and Cab-spotting networks are similar with RWP mobility model for smaller P_{ON}. This points us presence of pre-dominant entities and important hubs. For INFOCOM network the reason may be found in the fact that some scientist is widely recognized and the other tended to communicate with them. On the other hand, for Cab-spotting network some of the taxi cabs used to move in more central places, which resulted that their wireless communication with the other taxi cabs were more intensive. Although, there are important entities in both real networks, they cannot be characterized as "leaders" that are followed by group of other nodes as it is a case with RWPG models, where temporal robustness is most sensitive on "leaders" removal and robustness range area is the widest.

In conclusion, temporal network robustness is more sensitive in the cases of real temporal networks and mobility models, because of the presence of dominant and important nodes.

Thus, in future plans for more robust system/application design it makes sense to follow 2 possible approaches. The first proposes additional protection for important nodes and the second, suggests leveling the importance of all the entities (e.g. “P2P like” instead of centralized architecture).

Chapter 6

Conclusion and Future Research

6.1 Conclusion

This dissertation investigates temporal network robustness for different time-varying networks under several attacking strategies and random error. The evaluation of robustness unites the benefits of temporal network analysis from one side and robustness of network from the other.

Temporal efficiency is chosen as a performance metric as it expresses dynamic of temporal network very effectively. Secondly, the concept of robustness for static networks is extended to temporal networks. Accordingly, I propose a framework for temporal network robustness that measures the temporal robustness under those conditions providing us with more relevant analysis for system dynamic.

In the evaluation of temporal network robustness, temporal efficiency at two crucial moments is measured. The first moment considers temporal efficiency before a group of nodes is removed and the second after that, both taken in stationary regime. However, the choice of nodes that are removed can be crucial in temporal network robustness. For this reason, apart form random error, several attacking strategies are employed in the analysis: *temporal closeness nodes attack*, *average nodes degree attack* and *nodes number of contacts/updates attack*. In all the strategies the nodes are sorted according to importance, removed nodes are most important ones and the number is determined by probability of attack.

In order to understand the deviation when other nodes are removed, the *robustness range* is determined for all the models and real temporal networks, which is a difference between temporal robustness when less important and most important nodes are removed, respectively. In general, temporal robustness can have a values in this range.

I use several theoretical temporal network models and temporal networks from real data. The analysis considers Erdős-Rényi, Markov temporal networks and two classes of mobility

model. Temporal robustness is also evaluated on two case studies: the first one uses the data from mobile devices interactions between scientists on a conference and second one, interactions between taxi cabs wireless devices in San Francisco, USA during one day.

The results for different attacking strategies show similar effect on temporal robustness on all the used data. The reason for this is the fact that almost the same nodes are recognized as important in all the attacking strategies. In Erdős-Rényi and Markovian models, for both low and highly connected cases, the robustness have similar values. It is shown, theoretically (Erdős-Rényi) and by simulations (Markov) that nodes have similar values for average degree, temporal closeness and number of contacts/updates. For these reasons, we can classify these models as "balanced". In those temporal networks, robustness will be equally affected if other nodes are attacked.

On the other hand, for real data and mobility models, especially for RWPG, temporal robustness is more sensitive to intelligent attacks than random errors, because those models contain dominant nodes. RWPG even by definition consists of *leading* nodes that are central for particular group of nodes. Similar behaviour is shown for data from case studies. The last points out that those leading entities in the network, which failure affects further performance, should have better protection. In some cases, additional protection is more expensive and causes additional delays and reduced performance, therefore making the nodes equally important and the architecture decentralized is a better option.

The results suggest two possible applications for more robust design. In a centralized networked systems or software application, it is worth increasing the security level and introduce additional protection, because of their contribution for general performance. Additional protection may be an expensive solution and some processes in the system can be delayed. In these cases some decentralized approach works better. For instance, the system consists of several mini data-centers which keep redundant information is more robust than one huge data center, even if it has redundant data copies and additional protection. Another network architecture design solution is modification of peer-to-peer (P2P) system, where few nodes are more important than others is more robust, rather than a client-server approach.

6.2 Future Research

The results of this dissertation provide a good base and strong foundation for further research. There are two general directions for future work. The first continues in the framework of complex network analysis aware of temporal properties of networks. The second may follow a more practical line, considering the conclusions for better protection and more robust design of important nodes.

Investigating the preferential attachment behaviour [AB02], tendency for connecting to important nodes may show some interesting results for temporal robustness and range.

Moreover, considering another theoretical models like small-world graph or power law graphs may be another interesting field for future research. In the same direction, the effect of community detection over time and the effect of clustering approaches after temporal network sustains structural damage is an area for further research. Analysis will include clustering coefficient and modularity.

The second proposal considers design and real implementation of networked decentralized or P2P modified system with just a few pre-dominant nodes. This system will be evaluated under simulated failures and compared with typical data-center.

Bibliography

[AB02] R. Albert and A.-L. Barabasi. Statistical Mechanics of Complex Networks. *Review of Modern Physics*, 74:47–97, 2002.

[AJB00] R. Albert, H. Jeong, and A.-L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000.

[BLM+06] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, February 2006.

[CE07] A. Clauset and N. Eagle. Persistence and Periodicity in a Dynamic Proximity Network. In *Proceedings of DIMACS Workshop on Computational Methods for Dynamic Interaction Networks*, September 2007.

[CMH+07] P. Cholda, A. Mykkeltveit, B.E. Helvik, O.J. Wittner, and A. Jajszczyk. A survey of resilience differentiation frameworks in communication networks. *Communications Surveys Tutorials, IEEE*, 9(4):32 –55, quarter 2007.

[CNSW00] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468–5471, Dec 2000.

[dFCRTVB07] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *ADVANCES IN PHYSICS*, 56:167, 2007.

[ER59] P. Erdos and A. Renyi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[ER60] P. Erdos and A. Renyi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[FFF99] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In *Proceedings of SIGCOMM '99*, pages 251–262, New York, NY, USA, 1999. ACM.

[HA99] B. A. Huberman and L. A. Adamic. Growth dynamics of the world-wide web. *Nature*, 401(6749):131, 1999.

[HKYH02] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65(5), May 2002.

[KKK00] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. In *J. Comput. Syst. Sci*, page 2002, 2000.

[Kos09] V. Kostakos. Temporal graphs. *Physica A*, 388(6):1007–1023, March 2009.

[Lam78] L. Lamport. Time, Clocks, and the Ordering of Events in a Distributed System. *Commun. ACM*, 21(7):558–565, 1978.

[MGBM10] A. Medina, G. Gursun, P. Basu, and I. Matta. On the Universal Generation of Mobility Models. In *Proceedings of IEEE/ACM MASCOTS '10*, Miami Beach, FL, US, August 2010.

[Mil67] S. Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[PP94] L.B. Page and J.E. Perry. Reliability polynomials and link importance in networks. *Reliability, IEEE Transactions on*, 43(1):51 –58, mar 1994.

[PSDG09] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD trace set epfl/mobility/cab (v. 2009-02-24). Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility/cab, February 2009.

[RA78] S. Rai and K. K. Aggarwal. An efficient method for reliability evaluation of a general network. *Reliability, IEEE Transactions on*, R-27(3):206 –211, August 1978.

[SGC+09] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. CRAWDAD trace cambridge/haggle/imote/infocom2006 (v. 2009-05-29). Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom2006, May 2009.

[SLM+11] S. Scellato, I. Leontiadis, C. Mascolo, P. Basu, and M. Zafer. Understanding robustness of mobile networks through temporal network measures. In *Proceedings of INFOCOM'11*, Shanghai, China, 2011.

[SP78] A. Satyanarayana and A. Prabhakar. New topological formula and rapid algorithm for reliability analysis of complex networks. *Reliability, IEEE Transactions on*, R-27(2):82 –100, June 1978.

[TMML09] J. Tang, M. Musolesi, C. Mascolo, and V. Latora. Temporal Distance Metrics for Social Network Analysis. In *Proceedings of WOSN '09*, Barcelona, Spain, August 2009.

[TSM+10] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora. Small-world behavior in time-varying graphs. *Phys. Rev. E*, 81(5):055101, May 2010.

[VCAL11] M. Vidal, M. E. Cusick, and Barabasi. A.-L. Interactome networks and human disease. *Cell*, 144(6):986 – 998, 2011.

[VM06] P. Van Mieghem. *Performance Analysis of Communications Networks and Systems*. Cambridge University Press, UK, 2006.

[VMDW+10] P. Van Mieghem, C. Doerr, H. Wang, J. M. Hernandez, D. Hutchison, M. Karaliopoulos, and R. E. Kooij. A framework for computing topological network robustness. Technical Report report20101218, Delft University of Technology, 2010.

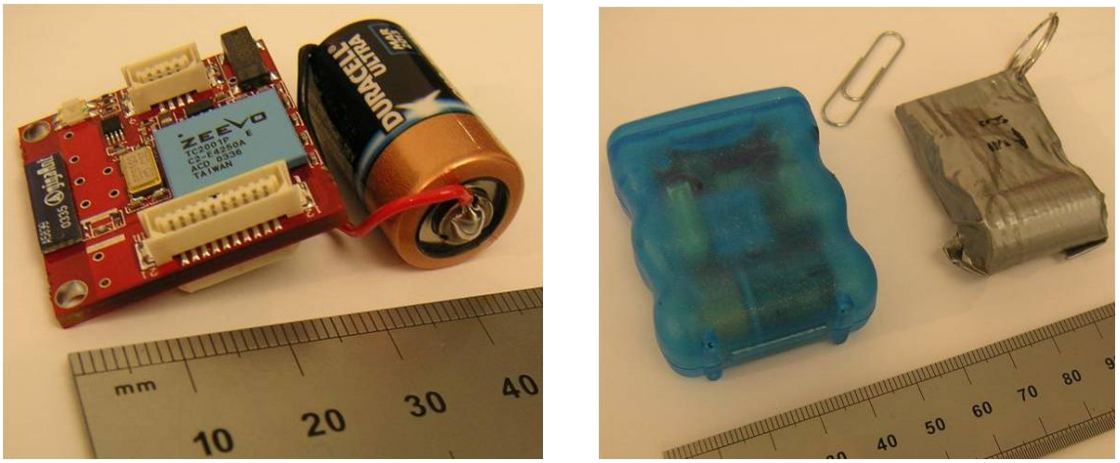
[Wil72] R. Wilkov. Analysis and design of reliable computer networks. *Communications, IEEE Transactions on*, 20(3):660 – 678, June 1972.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of small world networks. *Nature*, (393):440–442, 1998.

Appendix A

iMote wireless device

iMote wireless device used for collecting INFOCOM data is shown in Figure A.1. The width of the device is only 2.5cm.



(a) without packaging (b) with packaging

Figure A.1: iMote device