Modeling Failure Propagation in Large-Scale Engineering Networks

Markus Schläpfer[1] and Jonathan L. Shapiro[2]

[1] Institute of Energy Technology, ETH Zurich,
8092 Zurich, Switzerland
schlaepfer@mavt.ethz.ch
[2] School of Computer Science, University of Manchester
Manchester, M13 9PL, United Kingdom
jls@cs.man.ac.uk

Abstract. The simultaneous unavailability of several technical components within large-scale engineering systems can lead to high stress, rendering them prone to cascading events. In order to gain qualitative insights into the failure propagation mechanisms resulting from independent outages, we adopt a minimalistic model representing the components and their interdependencies by an undirected, unweighted network. The failure dynamics are modeled by an anticipated accelerated "wearout" process being dependent on the initial degree of a node and on the number of failed nearest neighbors. The results of the stochastic simulations imply that the influence of the network topology on the speed of the cascade highly depends on how the number of failed nearest neighbors shortens the life expectancy of a node. As a formal description of the decaying networks we propose a continuous-time mean field approximation, estimating the average failure rate of the nearest neighbors of a node based on the degree-degree distribution.

Keywords: cascading events, failure dynamics, decaying networks, mean field approximation, large-scale engineering systems.

1 Introduction

Infrastructure systems provide essential goods and services to industrialized society including transport, water, communication and energy. A disruption or malfunction often has a significant economical and social impact and potentially propagates to other systems due to mutual interdependencies. Wide-area breakdowns of such large-scale engineering networks are often caused by technical equipment failures which eventually result in a series of fast cascading component outages. Illustrative examples are a number of recent wide-area electric power blackouts and near-misses, where the causal chain of transmission line and generator disconnections became increasingly pronounced during the course of the events (e.g. [1]).

In the field of complex systems, a considerable effort over the past decade has led to an understanding and characterization of infrastructure robustness and

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 2127–2138, 2009.

vulnerability mainly by means of static network analysis [2,3,4] or by including simple nodal load models in order to account for static [5,6,7] or transient [8] load redistribution after a single component failure. One of the main conclusions was that networks with a scale-free degree distribution are more robust in regard to random node failures than networks with an exponential degree distribution, but exhibit a significantly higher vulnerability regarding deliberate attacks on highly connected nodes [9].

This paper addresses the underlying dynamics of cascading events in infrastructure systems by studying and further extending a recent minimalistic model for the spreading of failures in complex networks as introduced in [10]. The model considers independent and stochastic node outages which increase the stress on the remaining nodes as they are assumed to take over the load. The stress on a node, in turn, decreases its life expectancy which is modeled in analogy to an accelerated "wearout" process. Being motivated by cascading events within infrastructure systems, the primary objective of the model is to identify basic features of failure propagation processes within large-scale networks. At this stage of our work it was not the intention to model the actual behavior of a specific engineering system, but rather to capture some important properties which have to be further substantiated by the investigation of real systems. Yet, the model is kept general enough in order to be applicable to a wider range of complex networks where the outage of a node shortens the life expectancy of its neighboring nodes.

The paper is organized as follows: in the next section we describe the basic stress model and present the generalized accelerated "wearout" process. In section 3 we analyze different stress-dependent wearout functions. Section 4 introduces a simple repair process in order to derive stationary states of the networks. In section 5 we describe the decaying process by a continuous-time mean field approximation and compare the numerical results with Monte Carlo simulations. In section 6 we draw conclusions.

2 Degradation Model

2.1 Conceptual Basics

We represent a large-scale engineering network mathematically by an undirected, unweighted graph $\mathbf{G}(N, L)$ with N nodes being interconnected by L links. The graph is described by its $N \times N$ adjacency matrix $\mathbf{A}(G)$ [11].

The conceptual modeling framework for the network degradation process consists of a simple model for the nodal stress and a combined model for both the independent outage of a node and the stress induced shortening of its life expectancy.

2.2 Nodal Stress Function

In this paper, we consider the definition of the nodal stress introduced in [10]. The stress $s_{k_0,i}$ on a node with actual degree i corresponds to the ratio of the number of failed neighboring nodes to the total number of initially connected nodes, i.e. the initial degree k_0. Hence, we calculate $s_{k_0,i}$ as:

$$s_{k_0,i} = \frac{k_0 - i}{k_0} \tag{1}$$

Equation (1) implies the assumption that highly connected nodes have a stronger tolerance with respect to the outage of a neighbor than nodes with lower connectivity. Furthermore, the strength of the influence between two nodes is supposed to be the same for each pair of connected nodes.

2.3 Nodal Failure Rate

At time t=0 the network consists of N_0 nodes. In order to account for stochastic and independent node outages we assign to every node an identical and constant failure rate r_B. Thus, without additional node interaction the failure free time would be exponentially distributed with mean $MTTF = 1/r_B$, where $MTTF$ stands for *mean time to failure* [12] and can be equated to the life expectancy. However, as it is assumed that the stress $s_{k_0,i}$ increases the failure probability of a node in a given (and sufficiently small) time interval $(t+\delta t]$ and thus shortens its expected failure free time, we increase its basic failure rate by a stress-dependent term:

$$r_{k_0,i} = r_B + \Phi(k_0, i) \tag{2}$$

where $r_{k_0,i}$ is the overall nodal failure rate. The second term, $\Phi(k_0, i) = f(s_{k_0,i})$, is an arbitrary function of the actual stress and - in analogy to an accelerated wearout process - is further referred to as the "wearout acceleration function".

3 Analysis of the Wearout Acceleration Function

In order to assess the influence of the stress-dependent wearout on the failure propagation dynamics, two types of functions $\Phi(k_0, i)$ have been selected and analyzed: a linear model and a nonlinear, logistic model. Both models are applied to networks with an initial number of N_0=1000 nodes with either random or scale-free initial degree distribution. The ensemble of random networks is of the Erdős-Rényi type [13] with initial degree distribution $P(k_0) \approx \langle k_0 \rangle^{k_0} e^{-\langle k_0 \rangle}/k_0!$, where $\langle k_0 \rangle$ is the initial average degree. The ensemble of scale-free networks is generated using the preferential attachment method according to Barabási and Albert [14] leading to an algebraic degree distribution $P(k_0) \sim k_0^{-\gamma}$, where γ is the scale-free exponent. The network generation routines were adopted from the implementation in [15]. Due to their different statistical characteristics the two network types have been chosen for capturing the influence of the topology on the failure spreading process. As we aim at being as generic as possible they do not necessarily represent real-life topologies of large-scale engineering networks. Throughout the remainder of the paper the value for the basic failure rate is set to r_B=1 (per time unit).

3.1 Linear Model

The linear model assumes that the failure rate of a node increases linearly with its stress $s_{k_0,i}$:

$$\Phi(k_0, i) = cs_{k_0,i} \tag{3}$$

where c is a linear factor and $s_{k_0,i}$ is calculated according to formula (1). Figure 1 depicts the resulting time sequence of the network degradation process for both Erdős-Rényi random graphs (ER) and scale-free networks (SFN) with different initial average degree.

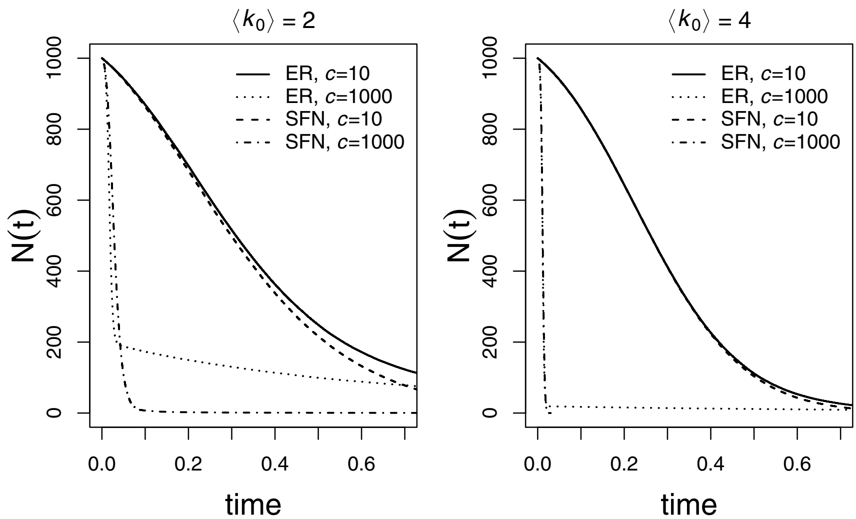

Fig. 1. Time sequence of the degradation process applying the linear wearout acceleration function for networks with N_0=1000 nodes and an initial average degree of $\langle k_0\rangle$=2 (left) and $\langle k_0\rangle$=4 (right). The degradation curves are generated by means of Monte Carlo simulations, averaged over 200 realizations.

According to Fig. 1 the speed of the failure propagation process is only weakly dependent on both the initial average degree and the initial degree distribution. The reason lies in the linear form of the wearout acceleration function $\Phi(k_0, i)$. When a neighbor of a node with a large initial degree k_0 fails, its stress is increased by a much smaller amount than it does for a node with a small k_0. Thus, it is expected that nodes with a large k_0 hold a higher life expectancy than nodes with a small k_0, and that highly connected and heterogeneous networks would therefore be more robust. However, this is compensated to a certain extent by the fact that any of the k_0 neighbors of a node can fail, linearly increasing the rate by c/k_0. As this again is valid for all those neighbors, the overall failure rate becomes only weakly dependent of the inital degree. The abrupt stopping of the cascade in random graphs with $\langle k_0\rangle$=2 corresponds to a total decomposition of the initially connected subgraph that contained the majority of nodes ("giant component").

3.2 Logistic Model

Regarding the nonlinear acceleration function we assume that the stress-dependent part of the failure rate increases according to a sigmoidal function:

$$\Phi(k_0, i) = c\frac{1}{1 + \exp[-\alpha(s_{k_0,i} - \theta)]} \qquad (4)$$

Assuming a sufficiently large value for the gain parameter of α=100 allows approximating a step function. By shifting the threshold θ we determine the level where the stress starts to significantly accelerate the wearout process. Hence, in contrary to the linear model the logistic function strongly weights the robustness of highly connected nodes. Figure 2 depicts the resulting time sequence of the network degradation process for both Erdős-Rényi random graphs and scale-free networks with different initial average degree.

Weighting the robustness of nodes with a high initial degree clearly makes the heterogeneous, scale-free network with an average initial degree of $\langle k_0 \rangle$=2 considerably more robust than the corresponding random graph, which is a homogeneous network with each node having approximately the same degree. The largest difference can be observed at a high influence of the accelerated wearout on the overall failure rate $r_{k_0,i}$. However, at a higher average initial degree of $\langle k_0 \rangle$=4 this effect disappears again. In a scale-free network with $\langle k_0 \rangle$=2 most of the nodes have degree k_0=1 and are connected to nodes with a significantly higher degree, which act as a "barrier" and prevent a further spreading of the

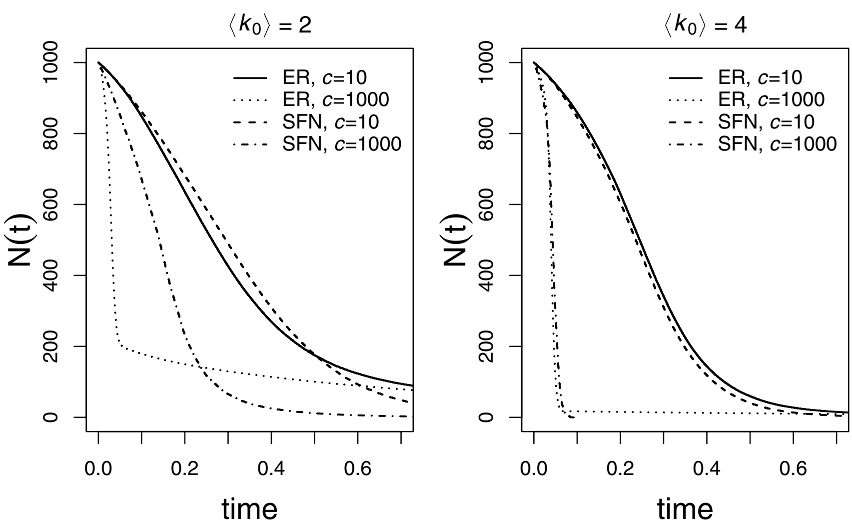

Fig. 2. Time sequence of the degradation process applying the logistic wearout acceleration function for networks with N_0=1000 nodes and an initial average degree of $\langle k_0 \rangle$=2 (left) and $\langle k_0 \rangle$=4 (right). The parameters of the logistic model are set to α=100 and θ=0.3 respectively. The degradation curves are generated by means of Monte Carlo simulations, averaged over 200 realizations.

node failures. In a heterogeneous network with higher connectivity the influence of these stabilizing hubs becomes suppressed as more nodes with lower degree are interconnected, again allowing failures to propagate through the network.

Figure 3 shows the influence of the failure threshold θ on the speed of the failure propagation process in an Erdős-Rényi random graph with initial size N_0=1000 and initial average degree of $\langle k_0 \rangle$=2.

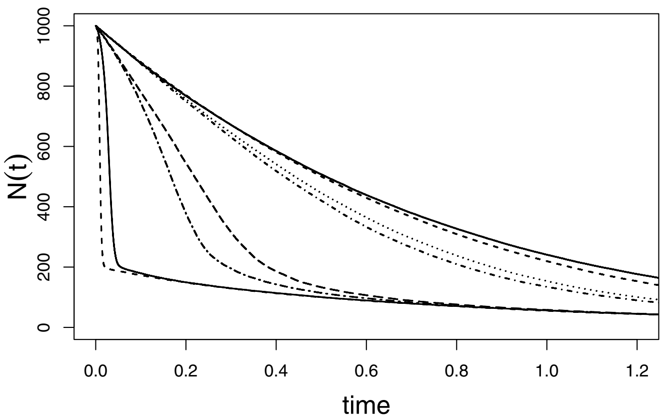

Fig. 3. Influence of the failure threshold θ on the speed of the degradation process for an Erdős-Rényi random graph with N_0=1000 and initial average degree of $\langle k_0 \rangle$=2. The parameters of the logistic model are set to c=1000, α=100 and θ=0.9,0.8,...,0.2 (from slowest to fastest degradation). The degradation curves are generated by means of Monte Carlo simulations, averaged over 200 realizations.

By gradually decreasing the failure threshold θ the speed of the overall network degradation is accelerated in rather large steps, reflecting the underlying degree distribution $P(k)$. The abrupt step between θ=0.6 and θ=0.5 for example corresponds to a highly increased failure rate of nodes with inital degree k_0=2 as they are becoming prone to the stress induced wearout. The increasing speed of the failure propagation between θ=0.4 and θ=0.3 can be explained analogously.

4 Repair of Nodes

We assume that the recovery speed of a failed node is independent from the state of the network. Furthermore, the repair of a node is modeled with a constant rate μ being independent of the initial and actual degree of the node:

$$\mu = \frac{1}{MTTR} \tag{5}$$

where $MTTR$ is the *mean time to repair* and represents the expected time value for the restoration process of a technical component [12]. If a node is repaired it

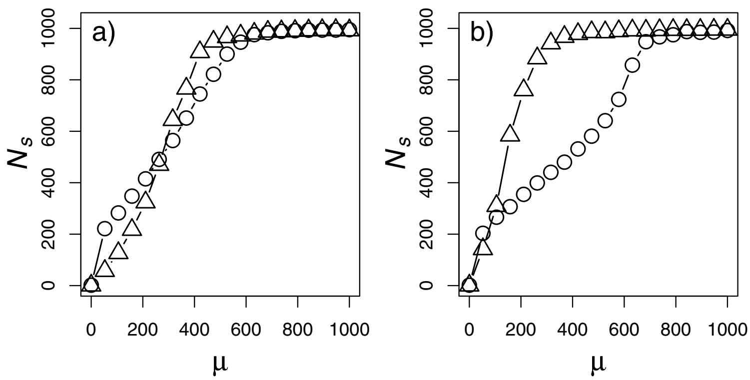

Fig. 4. Steady-state of the number of active nodes N_s as a function of the repair rate μ for ER (circles) and SFN (triangles) with N_0=1000 and $\langle k_0 \rangle$=2 each. a) linear model with c=1000, b) logistic model with c=1000, α=100 and θ=0.3.

becomes reconnected to all the originally neighboring nodes, thus reducing their stress again.

The steady-state of the network as being represented by the number of active nodes, $\lim_{t\to\infty} N(t) = N_s$, is calculated by means of a Monte Carlo simulation. As a stopping rule we use the coefficient of variation:

$$CV(N_s) = \frac{\sqrt{Var(N_s)}}{|E(N_s)|} < \beta \tag{6}$$

When the value of β reaches 0.01 or a maximum number of 150'000 retrials are conducted, the simulation is stopped. Figure 4 plots the steady-state of Erdős-Rényi random graphs and scale-free networks with N_0=1000 and $\langle k_0 \rangle$=2 against the repair rate for the two wearout acceleration models.

The steady-state of the number of working nodes, N_s, again reflects the higher robustness of scale-free networks with a small connectivity in the case of the logistic wearout acceleration function. Whereas for the linear function both networks behave in approximately the same way, a considerably higher repair rate is needed to keep the Erdős-Rényi random graph at the same level of functionality, if the robustness of the highly connected nodes is weighted by the logistic model.

5 Mean Field Approximation

A time-continuous mean field approximation is used to derive a first analytical description of the dynamic network degradation process. In order to account for different network topologies we consider the time dependent degree-degree correlation $P(k'|k,t)$. The degree-degree correlation represents the conditional probability that at time t a node of degree k is connected to a node of degree

k'. Therefore, we combine a master equation describing the evolution of nodes with a master equation describing the evolution of links, whereas the formalism for the derivation closely follows [16]. As the failure of a node is a discrete event we have to base our approach on the following assumption:

Assumption 1. *The total number of nodes $N(t)$ is large.*
The number of nodes is sufficiently large so that every event occurs with the frequency given by the probability of that event.

Consider a network with an arbitrary initial degree distribution $P(k_0)$. At some time t there exist $X_{k_0,i}(t)$ nodes with initial degree k_0 and actual degree i. At the same time the network consists of $L_{i,j}(t)$ links connecting nodes of actual degree i with nodes of actual degree j. The relation between the expected number of nodes with actual degree i, $X_i(t) = N(t)P(i,t)$, and the expected number of links connected to nodes with actual degree i, $L_i(t)$, is given by:

$$L_i(t) = \sum_{j>0} L_{i,j}(t) = i \sum_{k_0>0} X_{k_0,i}(t) = iX_i(t) \tag{7}$$

The total number of active nodes $N(t)$ is then given by:

$$N(t) = \sum_i X_i(t) = \sum_i \frac{L_i(t)}{i} \tag{8}$$

To satisfy the symmetric characteristic of the link-space matrix and in accordance with [16] we double count the number of links in the case $i = j$.

5.1 Node-Space Master Equation

The expected number of nodes $X_{k_0,i}(t)$ can change in three ways:

1. A node with initial degree k_0 and i neighbors can fail. This decreases $X_{k_0,i}(t)$ and occurs at rate $r_{k_0,i}$ according to Eq. 2.
2. One of the neighbors of a node with initial degree k_0 and actual degree i can fail, again decreasing $X_{k_0,i}(t)$. By making use of assumption 1 the expected failure rate of a neighboring node is given by the conditional probability of the actual degree of the neighboring nodes, $P(j|i,t)$, times the conditional probability of the initial degree of that node, $P(k_0|j,t)$, times the resulting stress-dependent failure rate, $r_{k_0,j}$. Hence, the expected (or average) failure rate of a neighboring node becomes:

$$\langle r_{nn}(t) \rangle_i = \frac{1}{L_i(t)} \sum_{j>0} \left[L_{i,j}(t) \frac{1}{X_j(t)} \sum_{k_0>0} X_{k_0,j}(t) r_{k_0,j} \right] \tag{9}$$

   The rate of losing a nearest neighbor is then given by multiplying $\langle r_{nn}(t) \rangle_i$ with the total number of nearest neighbors i.

3. One of the nearest neighbors of a node with actual degree $i+1$ and initial degree k_0 can fail, increasing $X_{k_0,i}(t)$. The expected failure rate of such a node is derived analogously to 2 by calculating the average failure rate of nearest neighbors $\langle r_{nn}(t)\rangle_{i+1}$.

The combination of the three outage modes changing $X_{k_0,i}(t)$ yields the node-space master equation describing the time evolution of the number of nodes with initial degree k_0 and actual degree i:

$$\frac{dX_{k_0,i}(t)}{dt} = -\left[r_{k_0,i} + i\,\langle r_{nn}(t)\rangle_i\right] X_{k_0,i}(t) + (i+1)\,\langle r_{nn}(t)\rangle_{i+1} X_{k_0,i+1}(t) \quad (10)$$

5.2 Link-Space Master Equation

In order to estimate the expected degree of a neighboring node using the two node degree correlation $P(k'|k,t)$ the actual number of links $L_{i,j}(t)$ is needed. The derivation of the link-space master equation follows in similar manner to [16], by introducing the degree dependent failure rate $r_{k_0,i}$. The expected number of links $L_{i,j}(t)$, connecting nodes of actual degree i with nodes of actual degree j can change in three ways being valid for both ends of a given link:

1. If a neighbor of an i+1 node fails and this node, in turn, is connected to a node with actual degree j, a new link $i \leftrightarrow j$ is formed. This is occurring at a rate i times the expected failure rate of a nearest neighbor of a node with actual degree i+1, $\langle r_{nn}(t)\rangle_{i+1}$. Analogously, $L_{i,j}(t)$ is increased if a neighbor of a node with actual degree j+1 fails and this node is connected to a node with actual degree i.
2. One of the neighbors of the two nodes at each end of a link $i \leftrightarrow j$ can fail with average rate $\langle r_{nn}(t)\rangle_i$ and $\langle r_{nn}(t)\rangle_j$, respectively. This decreases $L_{i,j}(t)$. As each of the neighbors may fail, the rates are multiplied by the actual number of connecting links, $i-1$ and $j-1$ respectively.
3. One of the nodes at each end of a link $i \leftrightarrow j$ can fail with average rate $\langle r(t)\rangle_i$:

$$\langle r(t)\rangle_i = \frac{1}{X_i(t)} \sum_{k_0>0} X_{k_0,i}(t) r_{k_0,i} \quad (11)$$

   and $\langle r(t)\rangle_j$ accordingly. This decreases $L_{i,j}(t)$.

The combination of the three outage modes yields the link-space master equation describing the time evolution of the number of links connecting nodes of actual degree i with nodes of actual degree j:

$$\begin{aligned}\frac{dL_{i,j}(t)}{dt} &= iL_{i+1,j}(t)\,\langle r_{nn}(t)\rangle_{i+1} + jL_{i,j+1}(t)\,\langle r_{nn}(t)\rangle_{j+1} \\ &\quad -\Big[(i-1)\,\langle r_{nn}(t)\rangle_i + (j-1)\,\langle r_{nn}(t)\rangle_j \\ &\quad + \langle r(t)\rangle_i + \langle r(t)\rangle_j\Big] L_{i,j}(t) \quad (12)\end{aligned}$$

5.3 Comparison with Monte Carlo Simulation

The system of coupled differential equations (Eq.(10) and Eq.(12)) has been integrated numerically using the standard Matlab ode45 solver [17], whereas for both classes of networks the results are averaged over an ensemble of 20 realizations. The Monte Carlo simulations are averaged over 200 network realizations. Regarding the linear wearout acceleration Fig. 5a) displays the evolution of the total number of nodes, $N(t)$, for Erdős-Rényi random graphs with an initial number of N_0=1000 nodes, an initial average degree of $\langle k_0 \rangle$=4 and different values for c. Figure 5b) shows the results for both scale-free networks and Erdős-Rényi random graphs with an initial average degree of $\langle k_0 \rangle$=2. For the random graphs there is an excellent match between the numerical solution of the mean field approximation and the results of the Monte Carlo simulations. At about 65% decomposition of the scale-free networks the mean field approximation starts to deviate from the Monte Carlo simulation as the speed of the cascade becomes underestimated. With respect to the logistic wearout acceleration function Fig. 5c) depicts the evolution of the total number of nodes for both Erdős-Rényi random graphs and scale-free networks with an initial average degree of $\langle k_0 \rangle$=2.

According to Fig. 5c) the mean field approximation again captures the evolution of the decaying random networks with a rather high accuracy, whereas regarding the scale-free networks it overestimates the speed of the failure propagation when about 30% of the nodes have failed. While the approach might be appropriate to describe heterogeneous networks characterized by a high functionality of the nodes, further model refinements seem to be necessary when it comes to the description of the evolution of a total network collapse.

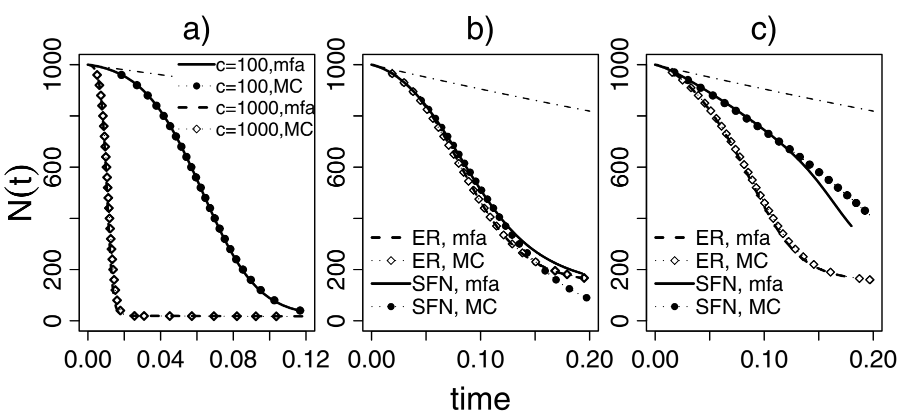

Fig. 5. Comparison of the mean field approximation (mfa) with Monte Carlo (MC) simulations. a) ER, $\langle k_0 \rangle$=4, linear wearout acceleration function with different values for c. b) ER and SFN, $\langle k_0 \rangle$=2, linear wearout acceleration function with c=100. c) ER and SFN, $\langle k_0 \rangle$=2, logistic wearout acceleration function with c=100, α=100 and θ=0.3. The dashed-dotted lines represent the theoretical decay of an empty graph with N_0=1000, i.e. without any node interaction.

6 Conclusions

We presented a minimalistic model based on an accelerated "wearout" process in order to reflect basic patterns of potential failure propagation dynamics within large-scale engineering networks. The influence of the topology on the speed of the overall network degradation highly depends on the particular model for the wearout acceleration due to failing neighbors of a node. Weighting the failure tolerance of highly connected nodes makes networks with scale-free structures significantly more robust than random networks. We further introduced a time-continuous mean field approximation describing the evolution of the decaying networks. The numerical results are in agreement with the Monte Carlo simulations except for the later stages of the degrading scale-free networks, making further model improvements necessary. The presented theoretical insights gained by the wearout analogy remain to be further substantiated by the investigation of real systems, eventually allowing for a derivation of optimal repair and defense strategies so as to prevent cascading events within infrastructure networks.

Acknowledgements. The authors would like to thank EPSRC for funding the Ambleside Summer School on Complex Systems, 2008, where this collaboration started. MS would also like to express his thanks to swiss*electric* research for co-funding the work the present paper is based, and greatly acknowledges the helpful discussions with Roberto Nibali and Sven Dietz (both Institute of Energy Technology, ETH Zurich) and Lubos Buzna (Chair of Sociology, in particular of Modeling and Simulation, ETH Zurich).

References

1. Gheorghe, A.V., Masera, M., Weijnen, M., De Vries, L.: Critical Infrastructures at Risk - Securing the European Electric Power System. Springer, Dordrecht (2006)
2. Sachtjen, M.L., Carreras, B.A., Lynch, V.E.: Disturbances in a power transmission system. Phys. Rev. E 61, 4877–4882 (2000)
3. Albert, R., Albert, I., Nakarado, G.L.: Structural vulnerability of the North American power grid. Phys. Rev. E 69, 025103 (2004)
4. Holmgren, A.J.: Using graph models to analyze the vulnerability of electric power networks. Risk Analysis 26(4), 955–969 (2006)
5. Motter, A.E., Lai, Y.C.: Cascade-based attacks on complex networks. Phys. Rev. E 66, 065102 (2002)
6. Motter, A.E.: Cascade control and defense in complex networks. Phys. Rev. Lett. 93, 098701 (2004)
7. Crucitti, P., Latora, V., Marchiori, M.: Model for cascading failures in complex networks. Phys. Rev. E 69, 045104 (2004)
8. Simonsen, I., Buzna, L., Peters, K., Bornholdt, S., Helbing, D.: Transient dynamics increasing network vulnerability to cascading failures. Phys. Rev. Lett. 100, 218701 (2008)
9. Albert, R., Jeong, A.L.H., Barabási, A.L.: Error and attack tolerance of complex networks. Nature 406, 378–382 (2000)

10. Schläpfer, M., Dietz, S., Kaegi, M.: Stress induced degradation dynamics in complex networks. In: Proceedings of the International Conference on Infrastructure Systems. IEEE, Los Alamitos (2008)
11. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of networks: From biological nets to the internet and WWW. Oxford University Press, Oxford (2003)
12. Birolini, A.: Reliability Engineering, Theory and Practice, 5th edn. Springer, Berlin (2007)
13. Erdős, P., Rényi, A.: On random graphs. Publ. Math. Debrecen 6, 290–297 (1959)
14. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
15. Taylor, A., Higham, D.J.: CONTEST: A controllable test matrix toolbox for MATLAB. University of Strathclyde Mathematics Research Report 13 (2007)
16. Smith, D.M.D., Lee, C.F., Onnela, J., Johnson, N.F.: Link-space formalism for network analysis. Phys. Rev. E 77, 036112 (2008)
17. Matlab, The Language of Technical Computing, Version R2008a