

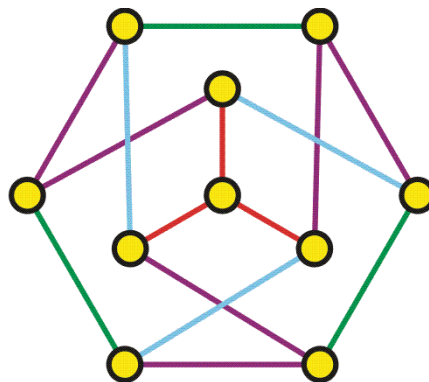
# SCIENTIFIC REPORTS

## WISSENSCHAFTLICHE BERICHTE



Journal of the University of Applied Sciences Mittweida  
Wissenschaftliche Zeitschrift der Hochschule Mittweida

## **E-Mobility and Reliability in Communication Networks**



The Mittweida  
Discrete Mathematics Group

## Impressum:

### Herausgeber:

Hochschule Mittweida  
University of Applied Sciences  
Der Rektor  
Prof. Dr. phil. Ludwig Hilmer  
Der Prorektor für Forschung und  
Entwicklung  
Prof. Dr.-Ing. habil. Gerhard Thiem

### Postanschrift:

Hochschule Mittweida  
University of Applied Sciences  
Referat für Forschung  
Postfach 1457  
D-09644 Mittweida

### Redaktion dieser Ausgabe:

Hochschule Mittweida  
University of Applied Sciences  
Prof. Dr. rer. nat. Peter Tittmann

### Erscheinungsweise: unregelmäßig

**Druck:** WIRmachenDRUCK GmbH

**Schutzgebühr: 15 €**

Die Scientific Reports / Wissenschaftliche Berichte als Wissenschaftliche Zeitschrift der Hochschule Mittweida (FH) University of Applied Sciences lösen die bisherigen Scientific Reports mit allen Volume I-III ab und erscheinen mit Nr. 1, 1998 ab November 1998 in neuem Layout und in neuer Zählung

### Bisher 2015 erschienen:

E-Mobility and Reliability in Communication Networks

Nr. 1, 2015

Für den Inhalt der Beiträge sind die Autoren verantwortlich.

# Preface

*Staying connected* is desired by anyone with respect to her or his own social network. Staying connected can be of vital importance in car-to-car communication networks as fast and reliable information transfer might not only improve fluidity of road traffic but also hinder traffic accidents. Complex dynamic systems such as appearing in e-mobility and telematics require new approaches of mathematical modeling. This report comprises some results from different research projects in the area of network reliability analysis, road traffic simulation, and intelligent power grids for e-mobility.

Chapter 1 of this report provides an introduction to mathematical models of network reliability analysis with emphasis on communication networks. It introduces relevant measures of network reliability and describes the main problems of network reliability analysis.

When a communication network loses its connectivity then most likely a single vertex becomes isolated. Consequently, the calculation of the probability that a network with randomly failing edges possesses isolated vertices is an important problem in reliability analysis. Different approaches to the solution of this problem are presented in Chapter 2.

Car-to-car communication networks are mobile ad-hoc networks without a precisely predictable structure, which renders routing and broadcasting a challenging task. Chapter 3 shows that random broadcasting methods provide mechanisms that alleviate the negative effects of blind flooding – a broadcasting scheme where each node immediately sends any new message to all of its neighbors. It also shows that the mathematical analysis of random broadcasting processes yields a difficult task that can be attacked via models related to the rainbow coloring problem of graphs. Chapter 4 continues the investigation of rainbow coloring problems and provides explicit solutions for some special cases.

Redundancy is a standard method applied in reliability engineering in order to make complex systems resistant against failures of its components. Redundancy is achieved in communication networks by reservation of alternative paths between terminals or more generally by guaranteeing higher connectivity. Chapter 5 is devoted to the calculation of the two-edge connected reliability, which is the probability that there exist two edge independent paths between any two vertices of a graph.

Basic principles for the management of information flow and sharing of data within car-to-car networks are presented in Chapter 6. One application of data sharing between cars is the improvement of navigation via inclusion of collectively aggregated traffic density data. A quite different aspect of e-mobility is the optimal planning and management of power grids supporting car recharging. Chapter 7 deals with the minimization of current fluctuations emerging in low voltage grids with many connected car charging stations.

Finally, I would like to thank all authors of this report and especially Sara Kischnick not only for contributing scientific results but also for carrying out all the  $\text{\LaTeX}$ -work required to produce this book.



# Contents

<b>Reliability Analysis of Car Communication Networks</b>	<b>1</b>
1 Reliability and Safety . . . . .	1
2 Scale and Accuracy of the Model . . . . .	2
3 Errors, Faults, and Failures . . . . .	3
4 Measures . . . . .	4
5 Mathematical Models for Communication Networks . . . . .	7
5.1 Graphs . . . . .	8
5.2 Monotone Binary Systems . . . . .	9
5.3 Markov Chains . . . . .	11
6 Routing . . . . .	11
<b>Isolated Vertices in Random Graphs</b>	<b>15</b>
1 The Extended Cut Polynomial . . . . .	16
1.1 Recurrence Formulae . . . . .	18
1.2 Regular Graphs and Cuts . . . . .	20
1.3 Independence Polynomial and Computational Complexity . . . . .	22
1.4 Edge Covers . . . . .	22
2 Isolated Vertices . . . . .	23
3 Calculation in Special Graph Classes . . . . .	25
4 Conclusion and Open Problems . . . . .	27
<b>Broadcast in Wireless Ad-hoc Networks</b>	<b>29</b>
1 Introduction – Broadcasting in Networks . . . . .	29
2 Random Broadcasting . . . . .	31
3 Rainbow Connection in Graphs . . . . .	34
3.1 Rainbow Independent Sets . . . . .	36
3.2 Rainbow Partitions . . . . .	38
3.3 The Rainbow Polynomial . . . . .	40
3.4 Random Edge Colorings . . . . .	42
4 Conclusions and Open Problems . . . . .	45
<b>Transversals of Random Set Systems</b>	<b>49</b>
1 Introduction . . . . .	49
2 Pseudoforests . . . . .	50
3 The Tutte Polynomial . . . . .	53
<b>Two-Edge Connected Reliability and Reductions</b>	<b>57</b>
1 Introduction . . . . .	57
2 Definitions . . . . .	57
3 General Algorithms for Two-Edge Connected Reliability . . . . .	58
3.1 Complete Enumeration . . . . .	58

3.2	Decomposition . . . . .	59
4	Reductions . . . . .	59
4.1	Degree-Two Reduction . . . . .	59
4.2	Reduction at a Separating Vertex Set of Cardinality Two . . . .	60
4.3	Reduction of Multiple Edges . . . . .	62
5	Conclusion . . . . .	64
<b>Sharing Data in Vehicular Ad-hoc Networks</b>		<b>67</b>
1	Abstract . . . . .	67
2	Introduction . . . . .	67
3	Conceptual Workflow . . . . .	68
4	Routing . . . . .	69
5	Driving . . . . .	69
6	Detection . . . . .	69
7	Sharing . . . . .	70
8	Simulation and Analysis . . . . .	72
8.1	Preprocessing Map Data . . . . .	72
8.2	Routing . . . . .	74
8.3	Sharing . . . . .	74
9	Conclusion . . . . .	75
<b>Selective Compensation of Harmonics in Smart Grid with the Active Filter of a Charging Station for Electric Vehicle</b>		<b>77</b>
1	Introduction . . . . .	77
2	Harmonic Detection . . . . .	78
2.1	Sliding Discrete Fourier Transform . . . . .	78
2.2	Frequency Locked Loop . . . . .	79
2.3	Fundamental Component Filter . . . . .	80
3	Compensation Algorithm . . . . .	80
4	Inverter Control . . . . .	83
5	Laboratory Test Results . . . . .	84
6	Conclusion . . . . .	86

# Reliability Analysis of Car Communication Networks

PETER TITTMANN

This chapter introduces basic notions from reliability theory that are important for the analysis of car communication networks. In addition, it provides an overview about mathematical models in network reliability analysis and presents relevant reliability measures.

## 1 Introduction to Reliability and Safety of Communication Networks

The first economic objective of any network operator is profit maximization. In order to achieve this basic goal, some activities concerning the network dependability emerge naturally. The main goal is to minimize costs that are caused by network failure. This objective in turn implies some of the following measures:

- Minimize the failure time of the system.
- Ensure that in case of a network failure a minimum number of users is impacted.
- Try to maintain as many services as possible when components of the network fail. If network failures cause high temporary peaks of traffic, then attempt to ensure that high-priority services are still guaranteed.
- Ensure that security of services and privacy of user data is maintained even in case of local or global network failures.
- Organize measures in order to achieve a quick recover from network failures. The implementation of this aim requires the development of methods of failure prediction.

The art of design of a system (communication network) that meets the above given requirements is called *reliability engineering*. Figure 1 shows some requirements of communication network design.

The design of a reliable communication network is impossible without proper knowledge about system properties, parameters, and mechanisms that affect network reliability. So, the first necessary step is the search for a network model that reflects its properties with respect to reliability. Then we have to define suitable reliability parameters, assumptions, target functions, and restrictions. The design and implementation of algorithms for the computation of reliability measures permits finally the investigation of concrete networks. We subsume all steps of this process under the term *network reliability analysis*.

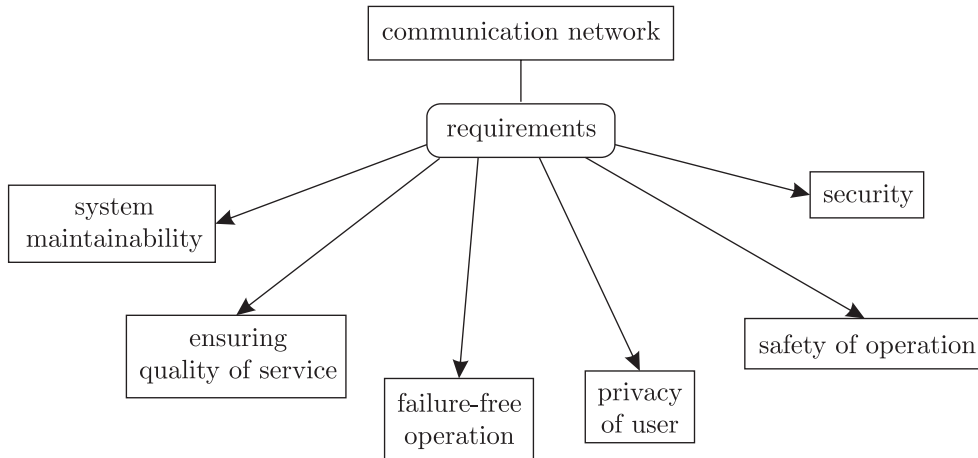


Figure 1: Demands for proper network operation

## 2 Scale and Accuracy of the Model

In order to form a model of the network in question, the definition of a scale of view and a clear boundary of our system is indispensable [21]. Nowadays virtually any communication network is somehow connected to the internet, which does not imply that we have to analyze the whole internet in order to investigate the reliability of a small wireless home network. Usually we *assume* that the Wi-Fi router has connection to an ISP and restrict the reliability analysis to the computers and other devices that are connected to the router. Quite another (often presupposed) assumption is that the power grid delivering electricity for the router and connected devices is well-operating. All these conditions define the *outer boundary* of the network in question.

The *scale of consideration* (called *limit of resolution* in [21]) defines the inner boundary of our system. When we investigate the dependability of VoIP communication, then we can tolerate transmission errors that do not affect the intelligibility. The loss of a single bit does not have any impact on the quality of speech transmission. On the other hand, the successful transfer of a program via the Internet requires that every single byte is transmitted correctly. However, the download of programs is much less dependent on latency as, for instance, video transmission. Hence, the first observation is that the required accuracy of the reliability analysis depends on the quality demands of the service. The scale of consideration has different aspects:

- The *required accuracy* defines tolerance limits for measurements. Here we have to find a balance between desired accuracy of results and realistic assumptions about procurable data.
- The *time resolution* refers to the precision of a measurement with respect to time. A finer time resolution implies a higher analytic effort. There are different ways to reduce time resolution. Fast error correcting mechanisms might be able to repair short-time errors without affecting the functionality of the service. In addition, the user might be willing to accept very short disruptions if they do rarely appear.
- The *space resolution* gives the degree of elaborateness of our system model. Can we consider a server in a computer network as a black box that may be treated as one component in a system? This might be appropriate for reliability evaluation in a “large scale”, for topological measures, or the computation of the connectedness



probability. However, if we investigate problems of throughput or packet delay, then the inner structure of a server becomes important. In this case, the inclusion of queue parameters and network protocol details can be necessary.

For long-term investigation of network reliability, a partition of the process of reliability analysis in different *phases* may be necessary. These phases correspond to steps of a communication protocol. Connection establishment, network service in operation (user communication), and connection termination require different resources of a communication network and hence different approaches of reliability analysis.

Almost all modern communication networks have a *layered structure*. The most prominent model for a layered network structure is the *OSI model* (Open Systems Interconnection). Only the OSI top layers are relevant for reliability analysis of user services. However, a more detailed analysis of selected connection properties may involve even the physical layer.

### 3 Errors, Faults, and Failures

A communication network as well as any other technical system is susceptible to different kinds of failures. *Random failures* are unpredictable failures that have a great variety of possible causes: technical defects of electronic components due to stress, material defects (wear), transmission errors of radio relays caused by bad weather conditions, electricity outage, submarine cable disruption, malfunction due to improper software design, human operator errors, etc. An assumption (or at least a hope) in reliability analysis is that we can somehow predict or estimate probabilities of component failures. A solid foundation for reliability estimation is the long-term observation of equivalent or similar systems and components. There are well-established methods in statistics in order to estimate probabilities from observed frequencies of failures. A quite different cause of failure is a *malicious attack*. We may assume that a clever attacker searches the weakest point of the network or a component whose destruction causes the maximum loss of functionality.

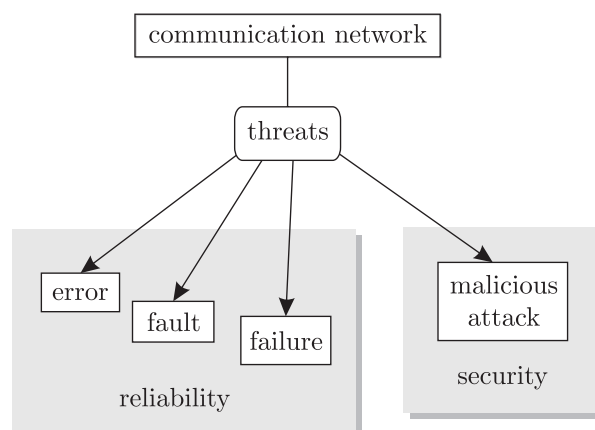


Figure 2: Threats to the operation of a communication network

Figure 2 shows the classification of threats to proper operation of a communication network. Threats that originate from random causes are in the main interest of reliability analysis whereas threats by human attackers with the aim to cause maximum system

damage are in the focus of security analysis. We will show in due course that these two fields, reliability and security, have some interrelations.

Sometimes it may be useful to distinguish between errors, faults, and failures. The term *error* is often used to describe improper behavior of components like erroneous computation results or even single bits that are changed during transmission, but without any influence to system operation due to correction methods or irrelevance. *Faults* are breaks of components that change the system but without any effect to the service in question as they are bridged by proper operating components, for instance via hot redundancy or rerouting. Thus faults need some action from the operator, but perhaps within a normal period of maintenance. The most serious system impact is caused by a *failure*, which implies some loss of functionality.

According to the classification of threats, we distinguish between *reliability measures* and *security measures*. First we have to define what a “properly operating network” is. A special service that we expect to be provided by the communication network is characterized by a set of parameters that may include connection, throughput, delay limits, etc. Thus we can assume that we are able to verify whether the network is in an *operating state* (intact). The *state of the system* depends on the *states of its components*. In the simplest case, a component can be in exactly one of two states, namely in *operating state* or in *failure state*. We assume that we can uniquely determine the state of the system from all its component states. A system that can assume only two states is called a *binary system*. The components of a binary system are also assumed to be in one of the two states. Sometimes we can consider components as binary even if their function degrades continuously. In this case we may define a *threshold* for a component parameter that allows to distinguish the states of proper operation and failure. Figure 3 gives an overview about some deterministic and stochastic reliability measures for communication networks.

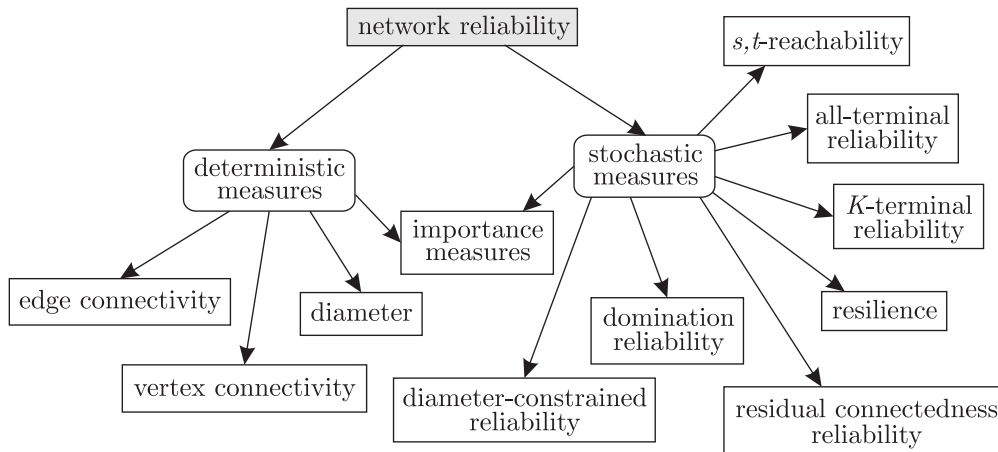


Figure 3: An overview about network reliability measures

## 4 Measures

A nice overview on definition, usage, and interrelations of terms from the area of network dependability is given by Al-Kuwaiti et al. in [1]. As we are mainly concerned with reliability problems of communication networks, the following collection of terms focuses on reliability measures.

## Reliability

The *reliability* of a system (or a component) is the probability that the system remains in operating state during a given period of time. In order to establish a more precise definition, we introduce a *state variable*

$$Z(t) = \begin{cases} 1 & \text{if the system is intact at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

The *lifetime* of a system, denoted by  $T$ , is real-valued random variable that is closely related to state of the system:

$$Z(t) = \begin{cases} 1 & \text{if } t < T, \\ 0 & \text{otherwise.} \end{cases}$$

The *lifetime distribution function*, denoted by  $F(t)$ , is defined by

$$F(t) = \Pr(T < t).$$

The *reliability function* or *survival function* is defined by  $R(t) = 1 - F(t)$ .

## Redundancy

The standard method in order to increase system reliability is *redundancy*. That is the insertion of additional components in a system that can replace failed components. One possibility to ensure redundancy is the parallel operation of equivalent components. A second way is to keep one component in reserve that is switched into operation as soon as the primary component fails. *Voting* is a special kind of redundancy that is employed for sensor and computing devices. A collection of equivalent sensors or processors deliver signals to a voter that decides via majority rules the “correct value”. In communication networks, an even more sophisticated concept of redundancy is realized by failure-redundant routing protocols.

## Availability

The *availability* is the proportion of time that a component is in operating state. Let  $T_{\text{up}}$  be the time in which a component is in operating state and  $T_{\text{down}}$  the time where the component is in failure state. Then the availability is defined by

$$a := \frac{\mathbb{E}T_{\text{up}}}{\mathbb{E}T_{\text{up}} + \mathbb{E}T_{\text{down}}},$$

which means that the definition of availability is suitable for systems with repair. The availability is the probability to find a system in operating state at randomly selected point of time. The reliability is the probability that a system is at time  $t$  still operating on condition it was in operating state at time 0.

## Performance

*Performance* is a measure of how well a system or a component functions [12]. In communication networks, performance refers to a measure of service quality such as

packed failure rate, latency, throughput, available channel bandwidth. A general analytic approach to performance is given by

$$P = \frac{\mathbb{E}X}{\max X},$$

where  $X$  is a non-negative real random variable describing the performance. In network reliability, the edge connectivity of a graph with randomly failing edges could provide a suitable measure of performance.

## Vulnerability

The term *vulnerability* is used to describe the dependence of the operation of a system on random failures of its components or malicious attacks. Let  $F = \{f_1, \dots, f_k\}$  be a set of *failure states* of a given system  $S$  and  $p_i$  the probability that  $S$  is in failure state  $f_i$ . Assume that there is a positive weight  $w_i$  assigned to each failure state  $f_i$ , such that a higher value of a weight corresponds to more severe “damage” of the system functionality. Then the vulnerability of  $S$  is given by

$$v = \sum_{i=1}^k p_i w_i.$$

Communication networks are modeled by directed or undirected graphs. A natural measure of vulnerability of a graph is the minimum number of edges or vertices that have to be removed in order to destroy its connectedness. In [11] the vulnerability of a network with respect to betweenness measures is investigated.

A more general discussion of vulnerability measures is given in [7]. We denote by  $\mathcal{G}$  the set of all finite undirected graphs. (If necessary, we can restrict  $\mathcal{G}$  to a subclass of graphs, we can consider directed graphs, or even more general classes of systems.) Let  $A$  be a nonempty set and  $f \rightarrow A$  be a map. We call  $f$  a vulnerability measure [7] if the following conditions are satisfied:

- The set  $A$  is ordered (or partially ordered). This property permits the comparison of networks and systems.
- The map  $f$  is monotone with respect to failure of components (edges or vertices), that is

$$f(G - x) \leq f(G)$$

for any edge  $x$  (or for any vertex  $x$ ).

- A vulnerability measure should be *sensitive* with respect to failures of any relevant components. As an example, we define

$$f(G) = \begin{cases} 1 & \text{if } G \text{ is connected} \\ 0 & \text{otherwise} \end{cases}$$

for any graph  $G \in \mathcal{G}$ . The codomain,  $\{0, 1\}$ , of this function is an linearly ordered set and  $f$  is monotone. However, this measure is not sensitive as the only edges that causes a change of  $f(G)$  are bridges. One way to improve the sensitivity of this function is to measure the “strength of connectedness”, for instance, by employing the edge or vertex connectivity.

- Two networks in parallel connection should be less vulnerable than the same two networks in a series connection.

## Resilience

The *resilience* is the ability of a networks or system to survive and return to normal operation despite of threats. Resilience describes how quickly a system is likely to recover from failure [13]. A *resilient* network routing infrastructure, for example, is expected to continue operating above a given minimum service level, even under localized failures, disruptions or attacks [4].

## Survivability

*Survivability* is the ability of a system to continue functioning after the failure of some of its components.

## Security

The concept of *network security* is related to protection from unwanted or prohibited actions such as the access to confidential data. Security measures also includes procedures to guarantee confidentiality, integrity, and trustworthiness of network services. The interdependence between reliability and security becomes obvious when we consider possibilities of unauthorized data access caused by system failures. Security analysis deals with the evaluation of damage that can be caused by system intruders, the location of weak points (components) of a network, possible scenarios of attack, and possible security lacks caused by malfunctioning components.

As security concerns in communication networks result generally from possible malicious attacks rather than from random faults, methods of probability theory are less successful in security analysis. The description of security threats has often a more qualitative character.

The report by Neumann, [19], provides a comprehensive survey on survivability with a strong focus on network security measures.

## Safety

The analysis of *safety* is of special importance for car-communication networks. Assume that we have established a car-to-car communication system that permits emergency stopping of cars via warning messages that are transmitted from car to car. Can we ensure that the car is safe in case of transmission errors? Is there a danger of automatically activating the brakes without any existing hazardous situation? The main objective of safety analysis is to discover system states that might imply catastrophic consequences on the user or the environment.

# 5 Mathematical Models for Communication Networks

A natural supposition in network reliability analysis is that the topology of the communication network is known. The set of components can be split into two main classes. The first class comprises servers, routers, user equipment (mobile phones, smartphones, vehicular on-board units in car-communication systems, switches, data bases. Entities of this class of components send and receive messages. They form the *vertices (nodes)* of a network. The second class consists of transmission channels, which may be realized

as fiber optic cables, copper wires, or wireless communication channels. Communication channels are called *simplex communication* and *duplex communication* channels if the communication is in only one direction and in both directions possible, respectively.

## 5.1 Graphs

A suitable mathematical model for the topology of a communication network is an *undirected* or *directed graph*. An undirected graph is a pair  $G = (V, E)$  consisting of a finite *set of vertices* and a finite *set of edges*. In addition, there is an *incidence function*  $\phi : E \rightarrow \{\{u, v\} \mid u, v \in V\}$  that assigns an unordered pair of vertices to each edge of the graph. Figure 4 shows an example graph. For a deeper introduction into graph theory, we recommend the books [22] and [5].

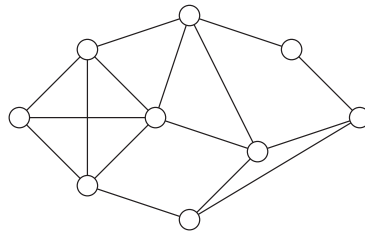


Figure 4: An undirected graph

A *walk* in a graph  $G$  is an alternating sequence,

$$v_1, e_1, v_2, e_2, \dots, v_{k-1}, e_{k-1}, v_k,$$

of vertices and edges of  $G$ , such that the edge  $e_i$  has the end vertices  $v_i$  and  $v_{i+1}$  for  $i = 1, \dots, k - 1$ . A *path* is a walk without any vertex repetitions. Consequently, a path does not traverse the same edge twice. A graph  $G = (V, E)$  is *connected* if there exists a path between any two vertices of  $G$ . A *closed walk* in a graph  $G$  is a walk for which start and terminal vertex coincide. A *cycle* in  $G$  is a closed walk in which no vertex, except start and terminal vertex, is traversed twice. A *tree* is a connected graph without cycles.

In many applications of graphs, we assign *weights* to the vertices or edges of a graph. A *weighted graph* is also called a *network*. Possible weights of vertices are server reliability, server capacity, delay time, cost, and others. An edge of a graph may be weighted with its reliability, length, capacity, or cost.

Two edges  $e = u, v$  and  $f = u, v$  that have the same endvertices are called *parallel*. An edge  $e = v, v$  for which the two endvertices coincide is called a *loop*. A graph  $G$  is called *simple* if  $G$  has neither loops nor parallel edges. As loops usually does not appear in network reliability, we assume that all graph considered here are loop-free. Parallel edges in graphs and networks can often be replaced by a single edge with a suitably chosen weight. Hence we can often restrict our attention to simple graphs.

A graph  $H = (W, F)$  is a *subgraph* of a graph  $G = (V, E)$  if  $W \subseteq V$  and  $F \subseteq E$ . In case  $W = V$  we call  $H$  a *spanning subgraph*. Let  $X \subseteq V$  be a given vertex subset. The *induced subgraph*  $G[X]$  of  $G = (V, E)$  is a subgraph with vertex set  $X$  and all edges of  $E$  that have both endvertices in  $X$ . A *spanning tree* of  $G$  is a connected cycle-free spanning subgraph of  $G$ . Spanning trees are important structures in network reliability as the form minimal subgraphs that guarantee the existence of paths between any pair of vertices.

We introduce some special graph that frequently appear in applications. An *isolated vertex* of a graph  $G = (V, E)$  is a vertex that is not incident to any edge of  $G$ . The *empty graph*, denoted by  $E_n$ , consists of  $n$  isolated vertices. The *complete graph*  $K_n$  is a simple graph in which any two vertices are adjacent. The *path*  $P_n$  has vertex set  $\{1, \dots, n\}$  and edge set  $\{\{1, 2\}, \dots, \{n-1, n\}\}$ . Introducing the additional edge  $\{1, n\}$  in the cycle  $P_n$  results in a *cycle*  $C_n$ .

In a *directed graph* or short *digraph*  $G = (V, E)$ , the edge set  $E$  consists of directed edges, which are also called *arcs*. An arc  $e = (u, v)$  is represented as an ordered pair of vertices, where  $u$  is the *tail* and  $v$  the *head* of  $e$ . In a drawing of a digraph, an arc is depicted as an arrow. Figure 5 shows a digraph with nine arcs.

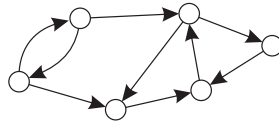


Figure 5: A digraph

A digraph without any directed cycle is called *acyclic*. Acyclic digraphs are important for the construction of routing tables.

## 5.2 Monotone Binary Systems

We consider now a mathematical model for communication networks that is suited for reliability modeling of more general systems than networks, too. We assume that a system, denoted by  $S$ , consists of  $n$  components, which we simply denote by  $1, \dots, n$ . Each component can be in one of two states, either failed or operating. A state variable

$$X_i = \begin{cases} 1 & \text{if } i \text{ is in operating state,} \\ 0 & \text{if } i \text{ is in failure state.} \end{cases}$$

indicates for each component  $i$  the corresponding state. We call  $S$  a *monotone binary system* if the following assumptions are satisfied:

- The components of the system can assume two states, failure state and operating state.
- The systems itself can also assume exactly these two states.
- The state of the system is uniquely defined by the states of all of its components. The *structure function* is a mapping  $\varphi : \{0, 1\}^n \rightarrow \{0, 1\}$  with

$$\varphi(\mathbf{X}) = \varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{if } S \text{ is in operating state,} \\ 0 & \text{if } S \text{ is in failure state,} \end{cases}$$

where  $\mathbf{X} = (X_1, \dots, X_n) \in \{0, 1\}^n$  denotes the state vector. The first three conditions define a *binary system*.

- The system  $S$  is *monotone*, which means that  $\mathbf{X} \leq \mathbf{Y}$  implies  $\varphi(\mathbf{X}) \leq \varphi(\mathbf{Y})$ . Here we use the standard order relation for vectors:

$$\mathbf{X} \leq \mathbf{Y} \iff X_i \leq Y_i \text{ for } i = 1, \dots, n$$



- The system  $S$  does not contain *irrelevant components*: Let  $\mathbf{X} \in \{0, 1\}^n$  be a given state vector. We denote by  $\mathbf{X}_{0,i}$  the vector obtained from  $\mathbf{X}$  by setting  $X_i = 0$ . In the same vein  $\mathbf{X}_{1,i}$  denotes the vector obtained from  $\mathbf{X}$  by setting  $X_i = 1$ . For each  $i \in \{1, \dots, n\}$ , there exists a vector  $\mathbf{X} \in \{0, 1\}^n$  such that  $\varphi(\mathbf{X}_{0,i}) < \varphi(\mathbf{X}_{1,i})$ . As a consequence of this assumption, each single component can change the state of the system (given that the other components are in an appropriate state). Observe that this property together with the monotonicity of  $S$  imply that  $\varphi(\mathbf{0}) = 0$  and  $\varphi(\mathbf{1}) = 1$ . Here  $\mathbf{0}$  and  $\mathbf{1}$  denote the all-0 and the all-1 vector, respectively.

We define the following operations for binary variables:

$$\begin{aligned} x \vee y &= \max\{x, y\} = x + y - x \cdot y \\ x \wedge y &= \min\{x, y\} = x \cdot y \end{aligned}$$

Let  $\mathbf{X} \in \{0, 1\}^n$  be a state vector with  $\varphi(\mathbf{X}) = 1$  and  $A = \{i \mid X_i = 1\}$ . The set of components  $A$  is called a *path set* of the monotone binary system. Consequently, a System  $S$  is in operating state if and only if all components of a path set of  $S$  are in operating state. Now let  $\mathbf{X} \in \{0, 1\}^n$  be a state vector of  $S$  with  $\varphi(\mathbf{X}) = 0$  and  $B = \{i \mid X_i = 0\}$ . Then  $B$  is a *cut set* of  $S$ . We denote by  $\mathcal{P}(S)$  the set of all path sets of  $S$  and by  $\mathcal{C}(S)$  the set of all cut sets of  $S$ . The structure function satisfies (see [3])

$$\varphi(\mathbf{X}) = \bigvee_{A \in \mathcal{P}(S)} \bigwedge_{i \in A} X_i \quad (1)$$

$$= 1 - \bigvee_{B \in \mathcal{C}(S)} \bigwedge_{i \in B} (1 - X_i). \quad (2)$$

Assume now that the components of the monotone binary system fail randomly and independently. Then the state of component  $i$  is given by the random variable  $X_i$ , that is from now on we consider the state variables as random binary variables. Let  $q_i = 1 - p_i$  the failure probability of component  $i$  for  $i = 1, \dots, n$ . Consequently, we obtain

$$\Pr(\{X_i = 1\}) = \mathbb{E}X_i = p_i, \quad i = 1, \dots, n.$$

The *system reliability* of a system  $S$  is the probability

$$R(S) = \Pr(\{X_S = 1\}) = \mathbb{E}X_S = \Pr(\{\varphi(\mathbf{X}) = 1\}).$$

The system reliability can be represented according to Equation (1) by

$$R(S) = \sum_{A \in \mathcal{P}(S)} \prod_{i \in A} p_i \prod_{i \in \{1, \dots, n\} \setminus A} (1 - p_i).$$

The cut set representation follows from Equation (2):

$$R(S) = 1 - \sum_{B \in \mathcal{C}(S)} \prod_{i \in B} (1 - p_i) \prod_{i \in \{1, \dots, n\} \setminus B} p_i.$$



### 5.3 Markov Chains

As in the last Section, we presuppose that the components of a communication network (vertices and edges) can assume certain states. Now we restrict the set of states of component  $i$  not necessarily to  $\{0, 1\}$  but to an arbitrary finite set  $S_i$ . Then the state space of the system is defined by

$$S = \bigotimes_{i=1}^n S_i.$$

We assume here that a change of states occurs only at discrete time steps, which we identify with the set of non-negative integers  $\mathbb{N} = \{0, 1, 2, \dots\}$ . Then a random process  $\{X_i\}_{i \in \mathbb{N}}$ , where the  $X_i$  are random variables with range  $S$  for  $i = 1, \dots, n$ , is a *Markov chain* if

$$\Pr(X_{n+1} = x \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \Pr(X_{n+1} = x \mid X_n = x_n)$$

is valid for all  $n \in \mathbb{N}$  and for all  $x_0, \dots, x_n, x \in S$ . This means that the probability of

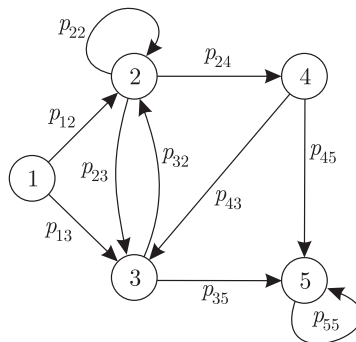


Figure 6: A Markov graph

transition from a state  $x$  at the time  $n$  into a state  $y$  at the time  $n + 1$  does not depend on states of the random Markov chain at points of time earlier than  $n$ . This is the reason why a Markov chain is called a *memoryless stochastic process*. Many classical textbooks in probability theory, such as [9] and [10], provide a good introduction to Markov chains.

The probability

$$p_{ij}(t) = \Pr(X_{t+1} = j \mid X_t = i) \quad i, j \in S, \quad t \in \mathbb{N}$$

is called *transition probability* of the Markov chain. The *transition matrix* of the Markov chain is defined by  $P(t) = (p_{ij}(t))_{i,j \in S}$ . one way to represent a Markov chain is the *Markov graph*, which is a digraph whose vertices correspond to states of the Markov chain. The arcs of the Markov graph represent transitions between states. The are weighted wit the corresponding transition probability. Figure 6 shows the Markov graph of a Markov chain with five states.

## 6 Routing

The classic method of routing in the Internet is *Open Shortest Path First* (OSPF). It is used to rout packets within one autonomous system (AS). The routing between different

autonomous systems is controlled by the *Border Gateway Protocol* (BGP). Based on the IP address of a packet, it uses shortest paths in the network. The assigned length of paths (routing metric) results from costs that depend of achievable transmission rates, path priorities (e.g. internal path first), and measured link quality. The more advanced *Resource Reservation Protocol* (RSVP) permits in addition service-dependent resource allocation. OSPF adapts to link failures or changes in the topology of the network. It provides also routing methods that support broadcast processes. Wireless ad-hoc networks use flooding algorithms in order send messages to a destination point without knowing its location. Flooding is a routing method in which incoming packet are sent along any edge with exception of that one it arrived on.

Assume we want to investigate the reachability of given destination vertex  $t$  in a network from a specified source vertex  $s$ . If the routing between these two vertices is performed in such a way that only shortest paths from  $s$  to  $t$  are used then we can represent the possibilities of forwarding of packets by an acyclic digraph. In this case, the additional routing information simplifies network reliability analysis.

Routing in mobile ad-hoc networks (MANET) is completely different from Internet routing. As there is no fixed structure in a MANET, each host serves also as a router. There can be no constant routing tables as the neighborhood of vertices changes perpetually. Consequently, routing protocols must be adaptive to changing network topology. Path between source and terminal vertices in MANETs are established by *flooding methods*. In order to maintain paths, *multipath routing* methods are proposed, see [18].

We can classify routing protocols for mobile ad-hoc networks into two categories, namely *proactive routing* and *reactive routing*. Proactive routing means that routing paths (shortest path) are computed prior to any message transfer requests. In a reactive routing process, the path from source to destination is established when needed.

In order to establish a reliable routing scheme for safety-critical applications in mobile ad-hoc networks, a classification of priority levels of messages is introduced in [16]:

1. Emergency warning messages have to be immediately transmitted to all cars in reach. This class comprises safety messages that make other cars are aware of hazardous conditions. Messages of this class are transferred to all neighbors via broadcast with a given number of repetitions.
2. Long-range emergency notification messages are messages that require quick and guaranteed delivery as, for instance, vehicle-based road condition warnings. Class-2 messages are sent via one-time multi-hop broadcast.
3. Routine safety messages have to be transmitted regularly but with least priority as the concern more general car measurement reports. This messages are broadcasted once to the neighbors (on-hop).

In vehicle-to-roadside communication, *geometric routing* (also known as *geographic routing* or as *position-based routing*) can be applied. In this case, we assume that a vertex (a car) knows its own position as well as the position of the destination of a message to be sent. In addition, we suppose that each vertex in the network knows besides its own position the positions of all its neighbors, too. The the simplest routing scheme that can be used is *greedy forwarding*, where in each step the message is transferred to a neighbor vertex that is closer to the destination.

Table 1: Routing methods for ad-hoc networks

Routing	Properties	Sources
<i>distance vector</i>	proactive, update via periodic broadcast	[17]
<i>link state</i>	proactive, update: incremental, triggered	
<i>dynamic source</i>	reactive	[14]
<i>geographic</i>	position based	[6, 15, 20]
<i>dominating-set-based</i>	connected backbone	[8, 23]

Table 1 shows some routing methods that are suitable for wireless ad-hoc networks. The hitherto presented routing methods provide only a small part of an existing huge variety of different routing protocols for ad-hoc networks. An overview about different routing methods for ad-hoc networks is given in [2].

An important aspect in modeling routing is that transmission can be unidirectional. Hence in some case digraphs rather than graphs provide the appropriate model for routing processes.

## References

- [1] M. Al-Kuwaiti, N. Kyriakopoulos, and S. Hussein. ‘A Comparative Analysis of Network Dependability, Fault-tolerance, Reliability, Security, and Survivability’. In: *IEEE Communications Surveys & Tutorials* 11.2 (2009), pp. 106–124.
- [2] E. Alotaibi and B. Mukherjee. ‘A survey on routing algorithms for wireless Ad-Hoc and mesh networks’. In: *Computer Networks* 56 (2012), pp. 940–965.
- [3] F. Beichelt and P. Tittmann. *Reliability and Maintenance. Networks and Systems*. 1. Auflage. Boca Raton, Florida: Taylor und Francis Group, 2012.
- [4] M. Bishop, M. Carvalho, R. Ford, and L. M. Mayron. ‘Resilience is more than availability’. In: *NSPW ’11 Proceedings of the 2011 workshop on New security paradigms workshop*. ACM. 2011, pp. 95–104.
- [5] J. Bondy and U. Murty. *Graph Theory*. Springer, 2008.
- [6] P. Bose, P. Morin, I. Stojmenovi’c, and J. Urrutia. ‘Routing with Guaranteed Delivery in Ad Hoc Wireless Networks’. In: *Wireless Networks* 7 (2001), pp. 609–616.
- [7] S. Bulteau and G. Rubino. *Evaluating network vulnerability with mincuts frequency vector*. Research report. 1997.
- [8] B. Das and V. Bharghavan. ‘Routing in ad-hoc networks using minimum connected dominating sets’. In: *ICC ’97 Montreal, Towards the Knowledge Millennium. IEEE International Conference on Communications, 1997*. Vol. 1. 1997, pp. 376–380.

- 
- [9] W. Feller. *An Introduction to Probability Theory and Its Applications: Volume One*. 3rd ed. John Wiley & Sons, 1968.
  - [10] G. Grimmett and D. Stirzaker. *Probability and random processes*. 3rd ed. Oxford University Press, 2001.
  - [11] P. Holme, J. B. Kim, C. N. Yoon, and S. K. Han. ‘Attack vulnerability of complex networks’. In: *Physical Review E* 65 (2002).
  - [12] J. B. Ingram-Cotton, M. J. Hecht, R. J. Duphily, M. Zambrana, T. Hiramoto, and C. O’Connor. *Reliability Program Requirements for Space Systems*. Aerospace Report TOR-2007(8583)-6889. El Segundo, California: Space and Missile Systems Center Air Force Space Command, 2007.
  - [13] S. K. Jain and P. K. Bhunya. ‘Reliability, resilience and vulnerability of a multi-purpose storage reservoir’. In: *Hydrological Sciences Journal* 53 (2 2008), pp. 434–447.
  - [14] D. B. Johnson and D. A. Maltz. ‘Dynamic source routing in ad hoc wireless networks’. In: *Mobile computing*. Springer, 1996, pp. 153–181.
  - [15] F. Kuhn, R. Wattenhofer, and A. Zollinger. ‘Ad-hoc networks beyond unit disk graphs’. In: *DIALM-POMC ’03 Proceedings of the 2003 joint workshop on Foundations of mobile computing*. 2003, pp. 69–78.
  - [16] X. Ma, J. Zhang, X. Yin, and K. Trivedi. ‘Design and Analysis of a Robust Broadcast Scheme for VANET Safety-Related Services’. In: *IEEE Transactions on Vehicular Technology* 61 (1 2012), pp. 46–61.
  - [17] J. M. McQuillan and D. C. Walden. ‘The ARPA network design decisions’. In: *Computer Networks* 1.5 (1977), pp. 243–289.
  - [18] A. Nasipuri, R. Castañeda, and S. R. Das. ‘Performance of Multipath Routing for On-Demand Protocols in Mobile Ad Hoc Networks’. In: *Mobile Networks and Applications* 6 (2001), pp. 339–349.
  - [19] P. G. Neumann. *Practical Architectures for Survivable Systems and Networks: Phase-One Final Report*. Research report. 1999.
  - [20] S. Rührup. ‘Theory and Practice of Geographic Routing’. In: *Ad hoc and Sensor Wireless Networks: Architectures, Algorithms, and Protocols*. Ed. by H. Liu, Y.-W. Leung, and X. Chu. Bentham Books, 2009.
  - [21] W. E. Vesely, F. F. Goldberg, N. H. Roberts, and D. F. Haasl. *Fault Tree Handbook*. Ed. by W. D. U.S. Nuclear Regulatory Commission. 1981.
  - [22] D. B. West. *Introduction to Graph Theory*. 2. Auflage. Upper Saddle River, New Jersey: Pearson Education, 2001.
  - [23] J. Wu. ‘Extended Dominating-Set-Based Routing in Ad Hoc Wireless Networks with Unidirectional Links’. In: *IEEE Transactions on Parallel and Distributed Systems* 13.9 (2002), pp. 866–881.

# Isolated Vertices in Random Graphs

MARKUS DOD<sup>1</sup> AND PETER TITTMANN

Reliability of communication networks has become a key factor in our modern highly interconnected world. Power grids, traffic management systems, enterprise data networks, and public switched telephone networks are examples of communication systems with a high demand for reliability. The mathematical model for the topology of a communication network is an undirected or directed graph. Mathematical network reliability research has been initiated with the pioneering work by Moore and Shannon [7]. Since then the literature on this topic has grown enormously, for some milestones, see [3, 8, 9]. We introduce in this chapter a new reliability measure which is an upper bound for the network reliability. It has nice connections to different counting problems in graphs.

In the following, we assume that the vertices of the graph are perfectly reliable and the edges fail independently with probability  $q$ . If the edge failure probability  $q$  is small (which is the case for many practical applications), then the probability of a single edge failure is often several orders of magnitude higher than the probability of a simultaneous failure of two or more edges. If a random graph with small edge failure probability becomes disconnected, then most likely a single vertex is separated. We denote the probability that a given graph  $G$  has no isolated vertices by  $P_{\text{iso}}(G)$ . Additionally, we denote with  $R(G)$  the probability that the graph  $G$  is connected.

Let  $G = (V, E)$  be a graph of order  $n$  and size  $m$ . The *vertex-induced subgraph* of  $G$  with vertex set  $W$ ,  $W \subseteq V$ , is denoted by  $G[W] = (W, E_G(W))$ , where the edge set is defined by

$$E_G(W) = \{\{u, v\} \in E \mid u \in W, v \in W\}.$$

If the considered graph is known from the context, then we write  $E(W)$  instead of  $E_G(W)$ . The *spanning subgraph*  $G\langle F \rangle$  for a edge subset  $F \subseteq E$  is the graph

$$G\langle F \rangle = (V, F).$$

Let  $\partial_G W$  be the set of edges of  $G$  that have exactly one vertex in  $W$ :

$$\partial_G W = \{\{u, v\} \in E \mid u \in W, v \notin W\}.$$

Again, if  $G$  is clear from the context, then we use the short form  $\partial W$ .

The *open neighborhood*  $N(v)$  of a vertex in  $G$  is defined by

$$N(v) = \{w \in V \mid \{v, w\} \in E\}.$$

The *closed neighborhood* of  $v$  is  $N[v] = N(v) \cup \{v\}$ .

---

<sup>1</sup>This research was supported by the ESF and a scholarship from the State of Saxony.

# 1 The Extended Cut Polynomial

The computation of the probability that there exist at least one isolated vertex is a hard task. To prove some results for this parameter we first study a more general counting polynomial. The *extended cut polynomial* of  $G$  is defined by

$$J(G; x, z) = \sum_{W \subseteq V} x^{|W|} z^{|\partial W| + |E(W)|}. \quad (1)$$

Let  $u_{ij}(G)$  be the number of vertex subsets of  $V$  with cardinality  $i$  such that the induced subgraph has size  $j$ :

$$u_{ij}(G) = |\{W \mid W \subseteq V, |W| = i, |E(W)| = j\}|.$$

The numbers  $u_{ij}(G)$  permit the following representation of the extended cut polynomial:

$$J(G; x, z) = x^n y^m \sum_{i=0}^n \sum_{j=0}^m u_{ij}(G) x^{-i} z^{-j}.$$

The empty graph  $E_n$  satisfies  $J(E_n; x, z) = (1 + x)^n$ , which follows easily by the definition of the extended cut polynomial. A vertex subset  $W$  of cardinality  $k$  in the complete graph  $K_n$  induces a complete subgraph of order  $k$ , which gives

$$|E(W)| + |\partial W| = \binom{k}{2} + k(n - k) = nk - \binom{k+1}{2}.$$

Consequently, we obtain

$$J(K_n; x, z) = \sum_{k=0}^n \binom{n}{k} x^k z^{nk - \binom{k+1}{2}}. \quad (2)$$

In order to prove some interesting properties of the extended cut polynomial, we consider in the following graphs that might have loops. For a graph  $G$  and an edge  $e = \{u, v\}$  of  $G$ , we define  $G \circ e$  as the graph obtained from  $G$  by removal of the vertices  $u$  and  $v$  and attaching loops to vertices in  $G - u - v$ . Each vertex  $w \in V \setminus \{u, v\}$  receives exactly  $|N(w) \cap \{u, v\}|$  new loops. In case  $e$  is a loop of  $G$  attached to vertex  $v$ , we define  $G \circ e$  as the graph obtained from  $G - v$  by attaching one additional loop to each vertex of  $N(v)$ . Let  $v$  be a vertex of the graph  $G = (V, E)$ . The graph  $G \circ v$  is obtained from  $G - v$  by attaching loops to all neighbors of  $v$ , one for each. Figure 1 illustrates these operations.

**Theorem 1** (Loop reduction). *Let  $G = (V, E)$  be a loopless graph and  $G'$  the graph obtained from  $G$  by attaching one loop to each vertex of  $G$ . Then*

$$J(G'; x, z) = J(G; xz, z).$$

*Proof.* We use the definition of the polynomial  $J(G; x, z)$ . Observe that the inclusion of a vertex  $v$  in  $W$  automatically extends  $\partial W$  by the corresponding loop.  $\square$

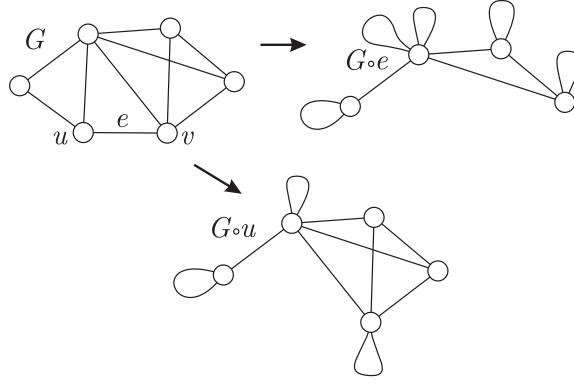


Figure 1: Edge and vertex operations.

**Theorem 2.** Let  $G = (V, E)$  be a graph with  $n$  vertices,  $m = \binom{n}{2}$ , and

$$J(G; x, z) = \sum_{i=0}^n \sum_{j=0}^m a_{ij} x^i z^j.$$

Then the extended cut polynomial of the complement  $\bar{G}$  of  $G$  is

$$J(\bar{G}; x, z) = \sum_{i=0}^n \sum_{j=0}^m a_{ij} x^i z^{\frac{i}{2}(2n-i-1)-j}.$$

*Proof.* By simple counting of edges we obtain

$$\begin{aligned} J(\bar{G}; x, z) &= \sum_{W \subseteq V} x^{|W|} z^{|E_{\bar{G}}(W)| + |\partial_{\bar{G}} W|} \\ &= \sum_{W \subseteq V} x^{|W|} z^{\binom{|W|}{2} - |E_G(W)| + |W|(n-|W|) - |\partial_G W|}, \end{aligned}$$

which gives with  $|W| = i$  and  $|E_G(W)| + |\partial_G W| = j$  the statement.  $\square$

Let  $G = (V(G), E(G))$  and  $H = (V(H), E(H))$  be two vertex disjoint graphs, then the *union*  $G \cup H$  is the graph  $(V(G) \cup V(H), E(G) \cup E(H))$ . The *join*  $G * H$  of two graphs  $G = (V(G), E(G))$  and  $H = (V(H), E(H))$  is the graph union  $G \cup H$  together with all the edges joining  $V(G)$  and  $V(H)$ .

**Theorem 3.** Let  $G = (V(G), E(G))$  and  $H = (V(H), E(H))$  be two graphs, with  $|V(G)| = n$  and  $|V(H)| = m$ . Furthermore, let  $a_i = [x^i]J(G; x, z)$  and  $b_j = [x^j]J(H; x, z)$ . Then

$$J(G * H; x, z) = \sum_{i=0}^n \sum_{j=0}^m a_i b_j x^{i+j} z^{im+jn-ij}.$$

*Proof.* Let  $W \subseteq V(G)$  and  $X \subseteq V(H)$  be two vertex subsets with  $i$  and  $j$  vertices, respectively. Then the term  $z^{im+jn-ij}$  counts the number of edges between these two vertex sets in  $G * H$ . The polynomial  $a_i$  counts the number of incident edges to a vertex subset of size  $i$  in  $G$ . Summing over all possible sizes of the two sets gives the theorem.  $\square$

**Corollary 4.** Let  $G = (V, E)$  be an arbitrary graph with  $n$  vertices. Then

$$J(G * K_1; x, z) = xz^n J(G; x, z) + J(G; xz, z).$$



## 1.1 Recurrence Formulae

In order to derive some recurrence equations for the extended cut polynomial with respect to elementary edge and vertex operations, we need the following statement.

**Lemma 5.** *Let  $G$  be a graph with  $k$  components  $G_1 = (V_1, E_1), \dots, G_k = (V_k, E_k)$ . Then*

$$J(G; x, z) = \prod_{i=1}^k J(G_i; x, z).$$

*Proof.* Using the Definition of the extended cut polynomial yields

$$\begin{aligned} J(G; x, z) &= \sum_{W \subseteq V} x^{|W|} z^{|\partial W| + |E(W)|} \\ &= \sum_{W \subseteq V} x^{|W \cap V_1| + \dots + |W \cap V_k|} z^{|\partial W \cap E_1| + |E(W) \cap E_1| + \dots + |\partial W \cap E_k| + |E(W) \cap E_k|} \\ &= \sum_{W_1 \subseteq V_1} x^{|W_1|} z^{|\partial W_1| + |E(W_1)|} \dots \sum_{W_k \subseteq V_k} x^{|W_k|} z^{|\partial W_k| + |E(W_k)|} \\ &= J(G_1; x, z) \dots J(G_k; x, z). \end{aligned}$$

□

**Theorem 6.** *Let  $G = (V, E)$  be a graph and  $e = \{u, v\} \in E$ . Then*

$$J(G; x, z) = z J(G - e; x, z) + (1 - z) J(G \circ e; x, z).$$

*Let  $E_n^*$  be the graph that is obtained from the empty graph of order  $n$  by attaching  $k(v)$ , for  $k(v) \geq 1$ , loops to each vertex  $v$ . Then the extended cut polynomial of the graph is*

$$\prod_{v \in V} (1 + x z^{\deg v}).$$

*These properties uniquely determine the polynomial  $J(G; x, y)$  for any graph  $G$ .*

*Proof.* Let  $H_r$  be a graph consisting of one vertex with  $r$  loops attached. Then we find

$$J(H_r; x, z) = 1 + x z^r \tag{3}$$

according to the definition of the extended cut polynomial.

Lemma 5 gives together with Equation (3)

$$J(E_n^*; x, z) = \prod_{v \in V} (1 + x z^{\deg v}).$$

Now we split the defining sum of the extended cut polynomial as follows:

$$\begin{aligned} J(G; x, z) &= \sum_{W \subseteq V} x^{|W|} z^{|\partial W| + |E(W)|} \\ &= \sum_{\substack{W \subseteq V \\ \{u, v\} \cap W \neq \emptyset}} x^{|W|} z^{|\partial W| + |E(W)|} + \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial W| + |E(W)|}. \end{aligned}$$



If  $\{u, v\} \cap W \neq \emptyset$ , then either  $e \in \partial W$  or  $e \in E(W)$ , which yields

$$\sum_{\substack{W \subseteq V \\ \{u, v\} \cap W \neq \emptyset}} x^{|W|} z^{|\partial W| + |E(W)|} = z \sum_{\substack{W \subseteq V \\ \{u, v\} \cap W \neq \emptyset}} x^{|W|} z^{|\partial_{G-e} W| + |E_{G-e}(W)|}$$

and hence

$$\begin{aligned} J(G; x, z) &= z \sum_{W \subseteq V} x^{|W|} z^{|\partial_{G-e} W| + |E_{G-e}(W)|} - z \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial W| + |E(W)|} \\ &\quad + \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial W| + |E(W)|} \\ &= z J(G - e; x, z) - z \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial W| + |E(W)|} \\ &\quad + \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial W| + |E(W)|}. \end{aligned}$$

For all  $W \subseteq V \setminus \{u, v\}$ , we have  $\partial_G W = \partial_{G \circ e} W$  and  $E_G(W) = E_{G \circ e}(W)$ , which provides

$$\begin{aligned} J(G; x, z) &= z J(G - e; x, z) - z \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial_{G \circ e} W| + |E_{G \circ e}(W)|} \\ &\quad + \sum_{W \subseteq V \setminus \{u, v\}} x^{|W|} z^{|\partial_{G \circ e} W| + |E_{G \circ e}(W)|} \\ &= z J(G - e; x, z) + (1 - z) J(G \circ e; x, z). \end{aligned}$$

□

**Theorem 7.** Let  $G = (V, E)$  be a graph and  $v \in V$ . Then

$$J(G; x, z) = xz^{\deg v} J(G - v; x, z) + J(G \circ v; x, z).$$

*Proof.* Consider the polynomial

$$J(G; x, z) = \sum_{W \subseteq V} x^{|W|} z^{|\partial W| + |E(W)|}.$$

If a vertex  $v \in V$  is included in  $W$ , then all edges that are incident to  $v$  are counted with the exponent of  $z$ , which gives the factor  $xz^{\deg v}$ . Let  $\partial v$  be the set of edges of  $G$  that have  $v$  as an end vertex. The polynomial  $J(G; x, z)$  can be decomposed as follows:

$$\begin{aligned} \sum_{W \subseteq V} x^{|W|} z^{|\partial W| + |E(W)|} &= \sum_{v \in W \subseteq V} x^{|W|} z^{|\partial W| + |E(W)|} + \sum_{W \subseteq V \setminus \{v\}} x^{|W|} z^{|\partial W| + |E(W)|} \\ &= xz^{\deg v} \sum_{v \in W \subseteq V} x^{|W| - 1} z^{|\partial W| + |E(W)| - \deg v} \\ &\quad + \sum_{W \subseteq V \setminus \{v\}} x^{|W|} z^{|\partial W| + |E(W)|} \\ &= xz^{\deg v} \sum_{W \subseteq V \setminus \{v\}} x^{|W|} z^{|\partial W \setminus \partial v| + |E(W)| - \deg v} \\ &\quad + \sum_{W \subseteq V \setminus \{v\}} x^{|W|} z^{|\partial W| + |E(W)|} \\ &= xz^{\deg v} J(G - v; x, z) + J(G \circ v; x, z). \end{aligned}$$

□

Using the previous proved recurrence equation we can obtain equations for the path  $P_n$  and the cycle  $C_n$ .

**Corollary 8.** *Let  $P_n$  be the path with  $n$  vertices and  $p_n = J(P_n; x, z)$ . Then*

$$p_n = \frac{(1 - (x + 2)z + f)g_n + ((x + 2)z - 1 + f)h_n}{2^{n+1}zf},$$

with

$$\begin{aligned} f &= \sqrt{xz((x + 4)z - 2) + 1}, \\ g_n &= (xz - f + 1)^n \text{ and} \\ h_n &= (xz + f + 1)^n. \end{aligned}$$

*Proof.* The path  $P'_n$  can be obtained from the path  $P_n$  by adding a loop to the first vertex of the path. We write short  $p'_n$  for the polynomial  $J(P'_n; x, z)$ . Using Theorem 7 we obtain for the path  $P_n$

$$p_n = xzp_{n-1} + p'_{n-1}$$

and for the path  $P'_n$

$$p'_n = xz^2p_{n-1} + p'_{n-1}.$$

Solving these recurrence equations with the initial conditions

$$p_1 = 1 + x \text{ and } p'_1 = 1 + xz$$

gives the theorem. □

## 1.2 Regular Graphs and Cuts

Let  $G = (V, E)$  be a regular graph of degree  $r$ . Then we have for any subset  $W \subseteq V$  the relation  $r|W| = 2|E(W)| + |\partial W|$ , which yields

$$J(G; x, z) = \sum_{W \subseteq V} (xz^{\frac{r}{2}})^{|W|} z^{\frac{1}{2}|\partial W|}. \quad (4)$$

The substitutions  $x = t^{-r}$  and  $z = t^2$  give

$$\frac{1}{2}J(G; t^{-r}, t^2) = \frac{1}{2} \sum_{W \subseteq V} t^{|\partial W|} = C(G, t).$$

The *cut polynomial*  $C(G, t)$  is the ordinary generating function for the number of cuts of a given size of  $G$ . The division by 2 takes into account that each cut of a graph  $G = (V, E)$  that is generated by a vertex subset  $W$  is also generated by its complement  $V \setminus W$ .

The cut polynomial of the cube graph  $Q_3$  presented in Figure 2 is

$$C(Q_3, t) = t^{12} + 8t^9 + 15t^8 + 24t^7 + 32t^6 + 24t^5 + 15t^4 + 8t^3 + 1.$$

The eight cuts of size 3 correspond to the vertices of  $Q_3$ . A cut of cardinality 4 arises when we choose two end vertices of an edge or four vertices of a square, where in each case the two parallel squares generate the same cut. There is a unique cut of size 12 that

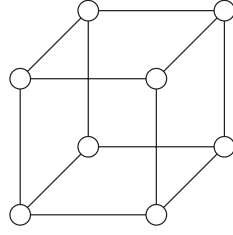


Figure 2: The cube

consists of all edges of the cube, which is generated by selecting a maximum independent set for  $W$ .

Let  $G = (V, E)$  be a graph, then the Ising polynomial  $Z(G; x, y)$  [2] is defined as follows

$$Z(G; x, y) = x^n y^m \sum_{W \subseteq V} x^{-|W|} y^{-|\partial W|}.$$

It follows directly from Equation (4) that the polynomial  $J(G; ut^{-r}, t^2)$  is equivalent to the *Ising polynomial* if  $G$  is  $r$ -regular.

**Definition 9.** [4] Let  $G = (V, E)$  be a graph. Then the bipartition polynomial is defined in the following way

$$B(G; x, y, z) = \sum_{W \subseteq V} x^{|W|} \sum_{F \subseteq \delta W} y^{|N_{(V,F)}(W)|} z^{|F|}.$$

The bipartition polynomial has some nice representations (see [4]). One of these representations is a multiplicative representation, which is useful for many proofs.

**Theorem 10.** [4] The bipartition polynomial has the following multiplicative representation:

$$B(G; x, y, z) = \sum_{W \subseteq V} x^{|W|} \prod_{v \in N_G(W)} [y [(1+z)^{|N_G(v) \cap W|} - 1] + 1]. \quad (5)$$

**Theorem 11.** Let  $G = (V, E)$  be a  $r$ -regular graph,  $B(G; x, y, z)$  its bipartition polynomial and  $J(G; x, y)$  its extended cut polynomial. Then

$$J(G; x, y) = B(G; x\sqrt{y^r}, 1, \sqrt{y} - 1).$$

*Proof.* If  $G$  is a  $r$ -regular graph, then  $r|W| = 2|E(W)| + |\partial W|$  and therefore  $|E(W)| = 1/2(r|W| - |\partial W|)$ . Using Equation (5) we can verify that

$$B(G; xt^r, 1, z - 1) = \sum_{W \subseteq V} x^{|W|} t^{r|W|} z^{|\partial W|}.$$

Now replacing  $t$  with  $\sqrt{y^r}$  and  $z$  with  $\frac{y}{\sqrt{y}}$  gives

$$\begin{aligned} \sum_{W \subseteq V} x^{|W|} \sqrt{y^{r|W|}} \left( \frac{y}{\sqrt{y}} \right)^{|\partial W|} &= \sum_{W \subseteq V} x^{|W|} y^{1/2(r|W| - |\partial W|)} y^{|\partial W|} \\ &= \sum_{W \subseteq V} x^{|W|} y^{|E(W)| + |\partial W|} \end{aligned}$$

which proves the theorem.  $\square$

### 1.3 Independence Polynomial and Computational Complexity

In this section we show the connection from the extended cut polynomial to the independence polynomial. We call a vertex subset  $W \subseteq V$  of a graph  $G$  *independent* if the vertices in  $W$  are pairwise non-adjacent.

**Definition 12.** [6] *Let  $G = (V, E)$  be a graph. Then the independence polynomial  $I(G, x)$  is the ordinary generating function for the number of independent sets of the graph:*

$$I(G, x) = \sum_{\substack{W \subseteq V \\ W \text{ is independent set}}} x^{|W|}.$$

**Theorem 13.** *Let  $G = (V, E)$  be a graph of order  $n$  and size  $m$ . We define the polynomial  $\tilde{J}$  by*

$$\tilde{J}(G, x, z) = x^n z^m J\left(G; \frac{1}{x}, \frac{1}{z}\right).$$

*Then the independence polynomial of  $G$  is given by*

$$I(G, x) = \tilde{J}(G; x, 0).$$

*Proof.* It suffices to observe that the number of vertex covers of cardinality  $k$  of  $G$  is the coefficient of  $x^k z^m$  in  $J(G; x, z)$ . The theorem follows as a set  $W \subseteq V$  is a vertex cover of  $G$  if and only if  $V \setminus W$  is independent in  $G$ .  $\square$

As the problem of deciding whether a given graph has an independent set of size  $k$  is known to be **NP**-complete, the computation of  $J(G; x, z)$  is an computationally intractable problem, too.

### 1.4 Edge Covers

An *edge cover* in a graph  $G = (V, E)$  is an edge set  $F \subseteq E$  such that each vertex of  $G$  is end vertex of at least one edge of  $F$ . Akbari et.al defined the edge-cover polynomial which is the ordinary generating function for the number of edge-covers of a graph.

**Definition 14.** [1] *Let  $G = (V, E)$  be a graph. Then the edge-cover polynomial  $E(G, z)$  of the graph is defined as*

$$E(G, z) = \sum_{\substack{F \subseteq E \\ F \text{ is edge-cover}}} z^{|F|}.$$

**Lemma 15.** [1] *Let  $G = (V, E)$  be a graph. Then*

$$E(G, z) = \sum_{W \subseteq V} (-1)^{|W|} (1+z)^{|E(G-W)|}.$$

The edge-cover polynomial can also be obtained from the extended cut polynomial. This result also shows that the computation of the extended cut polynomial is in  $\#P$ .

**Theorem 16.** *Let  $G = (V, E)$  be a graph with  $m$  edges. Then*

$$E(G, z) = (1+z)^m J\left(G; -1, \frac{1}{1+z}\right).$$

*Proof.* Substitute  $x$  with  $-1$  and  $z$  with  $1/(1+x)$  in Equation (1) gives

$$(1+z)^m J\left(G; -1, \frac{1}{1+z}\right) = \sum_{W \subseteq V} (-1)^{|W|} (1+z)^{m-|\partial W|-|E(W)|}.$$

Together with the Lemma 15 the theorem follows.  $\square$

## 2 Isolated Vertices

**Theorem 17.** *Let  $G = (V, E)$  be a graph with random failing edges. The edges are assumed to fail independently with identical probability  $q$ . Then the probability that  $G$  has no isolated vertex is given by  $J(G; -1, q)$ .*

*Proof.* Let  $A_v$ ,  $v \in V$ , be the random event that vertex  $v$  is isolated in  $G$ . Then the probability that  $G$  has at least one isolated vertex is, according to the inclusion–exclusion principle,

$$\Pr\left(\bigcup_{v \in V} A_v\right) = \sum_{\emptyset \neq W \subseteq V} (-1)^{|W|+1} \Pr\left(\bigcap_{v \in W} A_v\right).$$

Now observe that

$$\Pr\left(\bigcap_{v \in W} A_v\right)$$

is just the probability that all edges in  $E(W) \cup \partial W$  fail, which is  $q^{|\partial W|+|E(W)|}$ , see Figure 3. Taking the complementary event gives the theorem.  $\square$

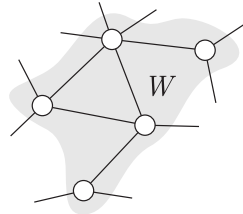


Figure 3: A vertex subset

From Theorem 7 and Theorem 17, we get immediately the following statement.

**Theorem 18.** *Let  $G = (V, E)$  be a graph. Assume all edges of  $G$  fail independently with identical probability  $q$ . Then for each vertex  $v \in V$ :*

$$P_{\text{iso}}(G) = P_{\text{iso}}(G \circ v) - q^{\deg v} P_{\text{iso}}(G - v).$$

This recurrence equation together with the initial value

$$P_{\text{iso}}(H_k) = 1 - q^k$$

for the graph  $H_k$  consisting of one vertex with  $k$  loops provides an algorithm for the calculation of  $P_{\text{iso}}(G)$  for any graph  $G$ . Theorem 18 can be easily generalized to cover also graphs with non-identical edge failure probabilities. Let  $q_e$  be the failure probability

of edge  $e$ , for  $e \in E$ . We denote by  $E(v)$  the set of edges of  $G$  that have  $v$  as one of their end vertices. Then

$$P_{\text{iso}}(G) = P_{\text{iso}}(G \circ v) - \prod_{e \in E(v)} q_e P_{\text{iso}}(G - v), \quad (6)$$

where we have to observe that the attached loops in  $G \circ v$  receive the failure probabilities of the corresponding edges.

Let  $e = \{u, v\}$  be an edge of  $G$  such that  $\deg u = 1$ . Then clearly a failure of  $e$  generates an isolated vertex, which gives the following reduction of degree-1 vertices:

$$P_{\text{iso}}(G) = (1 - q_e) P_{\text{iso}}((G - u) \circ v), \quad \deg u = 1, \quad e = \{u, v\}.$$

The following example shows the application of the reduction and the vertex decomposition according to Theorem 18 and Equation (6):

$$\begin{aligned} P_{\text{iso}} \left( \begin{array}{c} \circ \\ / \quad \backslash \\ a \quad c \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \end{array} \right) &= (1 - q_d) P_{\text{iso}} \left( \begin{array}{c} \circ \\ / \quad \backslash \\ b \quad a \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \end{array} \right) \\ &= (1 - q_d) \left[ P_{\text{iso}} \left( \begin{array}{c} \circ \\ / \quad \backslash \\ b \quad a \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \end{array} \right) - q_a q_c P_{\text{iso}} \left( \begin{array}{c} \circ \\ / \quad \backslash \\ b \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \\ \backslash \quad / \\ \circ \quad \circ \end{array} \right) \right] \\ &= (1 - q_d) [(1 - q_a q_b) - q_a q_c (1 - q_b)] \\ &= 1 - q_d - q_a q_b - q_a q_c + q_a q_b q_c + q_a q_b q_d \\ &\quad + q_a q_c q_d - q_a q_b q_c q_d. \end{aligned}$$

As a somehow surprising observation, we can verify that the paths with 3 and 4 vertices, respectively, satisfy

$$P_{\text{iso}}(P_3) = P_{\text{iso}}(P_4) = 1 - 2q + q^2,$$

presupposed that all edges fail with identical probability  $q$ .

Let  $H = (V, F)$  be a spanning subgraph of  $G = (V, E)$ . The edge set  $F$  of  $H$  is an edge cover of  $G$  if and only if  $H$  has no isolated vertices. Consequently, we can redefine  $P_{\text{iso}}(G)$  as the probability that the set of operating edges of  $G$  forms an edge cover of  $G$ . Hence, we get the next lemma.

**Lemma 19.** *Let  $G = (V, E)$  be a graph with random failing edges. Then*

$$P_{\text{iso}}(G) = (1 - p)^{|E|} E(G, p/(1 - p)).$$

*Proof.* A spanning subgraph  $G\langle F \rangle$ , with  $F \subseteq E$ , has no isolated vertex if and only if  $F$  is an edge-cover in  $G$ . Therefore,

$$(1 - p)^{|E|} E(G, p/(1 - p)) = \sum_{\substack{F \subseteq E \\ F \text{ is edge-cover}}} p^{|F|} (1 - p)^{|E| - |F|}$$

gives the probability that the graph has no isolated vertex.  $\square$

This lemma shows that the computation of  $P_{\text{iso}}(G)$  is in  $\sharp P$ . But it is possible to find a lower bound for  $P_{\text{iso}}(G)$  using the following theorem.

**Theorem 20.** [1] Let  $G = (V, E)$  be a graph with  $m$  edges and no isolated vertex. Furthermore, let  $e_i(G)$  be the number of edge-covers with  $i$  edges in  $G$  and

$$\tilde{e}_i(G) = \binom{m}{i} - \sum_{v \in V} \binom{m - d(v)}{i}.$$

Then

$$\begin{aligned} e_i(G) &\geq \tilde{e}_i(G), & \forall i \in \{1, \dots, m - 2\delta + 1\} \\ e_i(G) &= \tilde{e}_i(G), & \forall i \in \{m - 2\delta + 2, m\}. \end{aligned}$$

We can use this theorem to get a lower bound for  $P_{\text{iso}}(G, p)$ .

**Corollary 21.** Let  $G = (V, E)$  be a graph with  $m$  edges and no isolated vertex. Then for all  $p \in [0, 1]$

$$P_{\text{iso}}(G, p) \geq \sum_{i=1}^m [\tilde{e}_i(G) \geq 0] \tilde{e}_i(G) p^i (1-p)^{m-i}.$$

### 3 Calculation in Special Graph Classes

#### Complete and Complete Bipartite Graphs

First we investigate complete graphs for which we can exploit the symmetry in order to derive explicit expressions for the desired probability. The result can also be obtained from Equation (2) and the usage of the connection between the extended cut polynomial and  $P_{\text{iso}}(G)$ .

**Theorem 22.** If the edges of the complete graph  $K_n$  fail independently with probability  $q$ , then the probability that there is exactly one isolated vertex is

$$nq^{n-1} \left[ 1 - \sum_{k=1}^{n-1} (-1)^{k+1} \binom{n-1}{k} q^{\frac{k}{2}(2n-k-3)} \right].$$

*Proof.* First we compute the probability that there exists at least one isolated vertex in  $K_n$ . The probability that a given vertex  $v$  is isolated is  $q^{n-1}$ . Let  $A_v$  be the random event that vertex  $v$  is isolated. The probability that at least one vertex is isolated is given by the principle of inclusion-exclusion,

$$P \left( \bigcup_{v \in V} A_v \right) = \sum_{\emptyset \neq X \subseteq V} (-1)^{|X|+1} P \left( \bigcap_{w \in X} A_w \right).$$

In order to ensure that all vertices of a given subset  $X \subseteq V$  are isolated, all edges of  $K_n$  with both or one end vertex in  $X$  have to be failed, which happens with probability

$$q^{|X|(n-|X|) + \binom{|X|}{2}}.$$

This probability does not depend on the set  $X$  but only on its cardinality. Hence we obtain

$$\begin{aligned} P \left( \bigcup_{v \in V} A_v \right) &= \sum_{\emptyset \neq X \subseteq V} (-1)^{|X|+1} q^{|X|(n-|X|) + \binom{|X|}{2}} \\ &= \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} q^{k(n-k) + \binom{k}{2}}. \end{aligned}$$

Consequently, the probability that in a complete graph  $K_{n-1}$  no vertex is isolated amounts to

$$1 - \sum_{k=1}^{n-1} (-1)^{k+1} \binom{n-1}{k} q^{k(n-k-1)+\binom{k}{2}} = \sum_{k=1}^n (-1)^{k+1} \binom{n-1}{k} q^{\frac{k}{2}(2n-k-3)}.$$

The random events “vertex  $v$  is the only isolated vertex in  $K_n$ ” are disjoint for all  $v \in V(K_n)$ , which gives the statement of the theorem.  $\square$

The next statement follows directly from the proof.

**Corollary 23.** *Assume the edges of  $K_n$  fail independently with probability  $q$ . The probability that there exists an isolated vertex is*

$$1 - P_{\text{iso}}(K_n) = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} q^{k(n-k)+\binom{k}{2}}.$$

The probability that there are exactly  $r$  isolated vertices in  $K_n$  is

$$\binom{n}{r} q^{\frac{r}{2}(2n-r-1)} \left[ 1 - \sum_{k=1}^{n-r} (-1)^{k+1} \binom{n-r}{k} q^{\frac{k}{2}(2n-2r-k-1)} \right]. \quad (7)$$

Let  $Y_n$  be the random variable giving the number of isolated vertices in  $K_n$  with independently failing edges as given in the last corollary. The expected contribution of a given vertex  $v \in V(K_n)$  to the number of isolated vertices is the probability that  $v$  is isolated, which is  $q^{n-1}$ . Hence, by linearity of expectation, we obtain  $\mathbb{E}Y_n = nq^{n-1}$  for the expected number of isolated vertices in  $K_n$ . Another way to find this expectation is to use Equation (7), which yields

$$\mathbb{E}Y_n = \sum_{r=0}^n r \binom{n}{r} q^{\frac{r}{2}(2n-r-1)} \left[ 1 - \sum_{k=1}^{n-r} (-1)^{k+1} \binom{n-r}{k} q^{\frac{k}{2}(2n-2r-k-1)} \right].$$

However, now it seems to be a non-trivial task to verify that the sum on the right-hand side gives  $nq^{n-1}$ .

Now let  $G = (V, E)$  be an arbitrary graph whose edges fail independently with given probability  $q_e$ , for  $e \in E$ . We denote by  $Y_G$  the number of isolated vertices of  $G$ . Then we obtain (again by linearity of the expectation)

$$\mathbb{E}Y_G = \sum_{v \in V} \prod_{\substack{e \in E \\ v \in e}} q_e.$$

**Theorem 24.** *Let  $K_{s,t}$  be a complete bipartite graph which edges fail independently with probability  $q$ . Moreover, let  $s, t \geq 2$ . Then*

$$P_{\text{iso}}(K_{s,t}) = q^{st} \sum_{k=0}^s (-1)^{s-k} \binom{s}{k} \left( \frac{1-q^k}{q^k} \right)^t.$$

*Proof.* It is already known [1] that the edge-cover polynomial of the complete bipartite graph is

$$\sum_{k=0}^t (-1)^{t-k} \binom{t}{k} ((1+x)^k - 1)^s. \quad (8)$$

Applying Lemma 19 to Equation (8) gives the theorem.  $\square$



### Paths and Cycles

Next we derive the probability  $P_{\text{iso}}(C_n)$  of a cycle  $C_n$ , supposed that all edges fail independently with probability  $q$ . The key observation is that  $C_n$  has no isolated vertices if and only if the set of failing edges forms a *matching* (that is a set of edges such that no two edges share a common end vertex) in  $C_n$ . The number of matchings of size  $k$  ( $k > 0$ ), in the cycle  $C_n$  is

$$\frac{n}{k} \binom{n-k-1}{k-1},$$

see [5]. Consequently, we get

$$P_{\text{iso}}(C_n) = (1-q)^n + \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \frac{n}{k} \binom{n-k-1}{k-1} q^k (1-q)^{n-k}.$$

For the path  $P_n$  we can use the same idea as for the cycle, except of the fact that the first edge of both sides has to be intact. The number of matchings of size  $k$ ,  $k \geq 0$ , in the path  $P_n$  is

$$\binom{n-k}{k}.$$

Thus, we get

$$P_{\text{iso}}(P_n) = (1-q)^2 \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-k-2}{k} q^k (1-q)^{n-k-3}.$$

The star  $S_n$  with  $n$  vertices has  $n-1$  degree-one vertices and therefore, all edges must be intact. This gives immediately  $P_{\text{iso}}(S_n) = (1-q)^{n-1}$ .

## 4 Conclusion and Open Problems

We could show that the extended cut polynomial offers an interesting way in order to calculate the probability of non-existence of isolated vertices in graphs with independently failing edges. It remains to investigate how this measure can be combined with further structural information, e.g. small cuts, to establish improved bounds for the all-terminal reliability.

Interesting open problems with relation to the extended cut polynomial are:

- Which graph classes permit a polynomial-time computation of  $J(G; x, z)$ ?
- Are there other elementary vertex or edge operations in graphs that lead to new recurrence equations for the extended cut polynomial?
- Can we use reduction (transformations) of the graph in order to calculate  $J(G; x, z)$  more efficiently?

Another open problem is the expansion of the given definition in order to cover also digraphs and hypergraphs.

## References

- [1] S. Akbari and M. R. Oboudi. ‘On the edge cover polynomial of a graph’. In: *European Journal of Combinatorics* 34.2 (2013), pp. 297–321.
- [2] D. Andr en and K. Markstr om. ‘The bivariate Ising polynomial of a graph’. In: *Discrete Applied Mathematics* 157.11 (2009), pp. 2515–2524.
- [3] C. Colbourn. *The combinatorics of network reliability*. Oxford University Press, USA, 1987.
- [4] M. Dod, T. Kotek, J. Preen, and P. Tittmann. ‘Bipartition Polynomials, the Ising Model, and Domination in Graphs’. In: *Discussiones Mathematicae Graph Theory* 35.2 (2015), pp. 335–353.
- [5] E. J. Farrell. ‘An introduction to matching polynomials’. In: *Journal of Combinatorial Theory, Series B* 27.1 (1979), pp. 75–86.
- [6] V. E. Levit and E. Mandrescu. ‘The independence polynomial of a graph—a survey’. In: *Proceedings of the 1st International Conference on Algebraic Informatics*. 2005, pp. 233–254.
- [7] E. Moore and C. Shannon. ‘Reliable circuits using less reliable relays’. In: *Journal of the Franklin Institute* 262.3 (1956), pp. 191–208.
- [8] A Satyanarayana and M. K. Chang. ‘Network reliability and the factoring theorem’. In: *Networks* 13.1 (1983), pp. 107–120.
- [9] D. Shier. *Network reliability and algebraic structures*. Oxford University Press, USA, 1991.

# Broadcast in Wireless Ad-hoc Networks

SARA KISCHNICK<sup>1</sup> AND PETER TITTMANN

Network broadcasting offers an efficient and fast method for spreading information that originates in a single node across a network region. Probabilistic broadcast schemes seems to be well-suited in order to deal with the rapid change of topology in car-to-car networks. This chapter provides basic algorithms for random broadcasting in networks together with its mathematical analysis.

## 1 Introduction – Broadcasting in Networks

In near future it will be possible that cars can communicate with each other. The implementation of car-to-car communication will provide a new approach to improving road safety and traffic efficiency. The applications of car-to-car communication systems include

- Driver warning in case of approaching emergency vehicles,
- Forwarding information about hazardous locations,
- Information from road side units about construction zones or dangerous states of road cover,
- Information about traffic density and emergence of traffic jam.

As normal road traffic results in a set of random locations of cars, a car-to-car communication network does not have a predictable structure – it forms a *wireless ad-hoc network*. Hence, every car is both a client node in the network and a router. Because of the high mobility of the communication participating entities, mutual connections changes very fast. Therefore, the architecture of the network is highly instable.

Fast message forwarding within a network of unknown topology cannot rely on classical methods like open shortest path routing. An alternative that guaranties minimal delay is network broadcasting, which is especially appropriate as the relevant information originates usually in a single car and has to be distributed as quickly as possible across its neighborhood.

A huge variety of different broadcasting methods has been developed through the last years. One of the simplest approaches is that each node sends the message to all neighbors within its transmission range. However, this type of broadcasting scheme generates high network load – the *broadcast storm*. The data traffic can be reduced by

---

<sup>1</sup>The research is supported by SMWA/BMVI.

using lifetime counters. A better way to spread the message is flooding [19, p. 427]. The information will be sent to all neighbors except the sender of the message. A further method to reduce the broadcast storm is to equip the messages with a sequence number. Every node will save the number and transmit only messages that have not been sent previously. Tonguz et al. [20] introduced a broadcast procedure which is particular for the car-to-car communication – the Distributed Vehicular Broadcast (DV-CAST) protocol. This method uses the local information of the network topology, especially the one-hop neighborhood of a car <sup>2</sup>.

The topology of a car-to-car communication network can be represented by an undirected graph  $G = (V, E)$  with vertex set  $V$  and edge set  $E$ . A vertex represents a car communication unit whereas an edge stands for a wireless communication channel. The graphs considered here are assumed to be simple (they have neither loops nor parallel edges) and connected.

*Broadcasting* is the transfer of a message that resides initially in exactly one vertex of the network to all other vertices. We distinguish *deterministic* and *random broadcasting algorithms*. Another classification results from *local* versus *global broadcasting*. Local algorithms for broadcasting require only the knowledge of the neighborhood of a vertex whereas global methods presuppose complete information about the network topology. Consequently, localized broadcasting is employed in wireless ad-hoc networks.

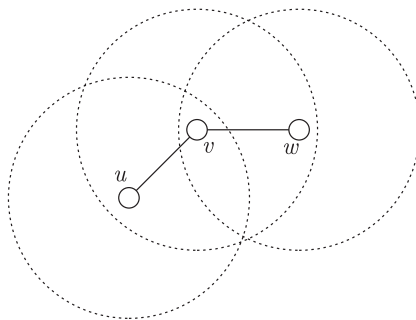


Figure 1: The hidden node problem

In ad-hoc networks, *blind flooding* is the simplest broadcast method. Blind flooding refers to a broadcasting method where each vertex transmits a received packet that has not already been transmitted by this vertex to all its neighbors. However, blind flooding has a lot of disadvantages including redundant transfer of messages implying high network traffic, competition for network resources, and increased frequency of packet collision, see [11, 13]. A *collision* is the simultaneous receiving of a broadcast message from different transmitters in one vertex. Figure 1 illustrates a possible source of *packet collision* in wireless ad-hoc networks. The vertices  $u$  and  $w$  are in the transmission range of  $v$  but  $u$  and  $w$  cannot reach each other – they are “hidden” from each other. This problem is referred to as the *hidden node problem* or as the *hidden terminal problem*. Assume that there is a transmission from  $u$  to  $v$ . As  $w$  is not aware of this situation it might decide to send packets to  $v$  causing a collision at  $v$ . Possible solutions of the hidden node problem and arising difficulties are presented in [7, 14, 15].

There are several methods in order to overcome the problems imposed by blind flooding:

<sup>2</sup>For more information to this broadcast procedure see Tonguz et al.: “Broadcasting in VANET” [20] and Tonguz et al.: “On the Broadcast Storm Problem in Ad hoc Wireless Networks” [21]

- Application of random broadcasting. One way to reduce network traffic in broadcasting processes is to forward a packet with a given probability  $p < 1$  to a neighbor. We can clearly expect that with decreasing values of  $p$  the probability to reach all vertices of the network is decreasing, too. A suitable choice of the parameter  $p$  depends also on the edge density of the network.
- Introduction of a hop limit for packets (corresponding to a “time-to-live”). This parameter also can be used to control random forwarding.
- The vertices of the network can be equipped with a *memory* in order to ensure that only packets that are not sent before will be transmitted.

A *broadcast process* on a graph  $G = (V, E)$  is given by a sequence  $\{X_i\}_{i \in I}$ , where  $I = \mathbb{N}$  or  $I = \{0, 1, \dots, n\}$  and  $X_i \subseteq V$ , for all  $i \in I$ . We call a vertex  $v \in X_t$  *informed* at time  $t$ . A vertex that is not informed is called *ignorant*. We assume that a broadcasting process has the following properties:

- There is at least one informed vertex at time 0, which gives  $X_0 \neq \emptyset$ . This condition ensures that we have a proper broadcasting process, which consists in the distribution of messages.
- A vertex  $v$  that is informed at time  $t$  will be informed also at any time  $t'$  with  $t' > t$ , which means that vertices do not “forget” a message once received. As a consequence of this requirement, the sequence  $\{X_i\}_{i \in I}$  is monotone increasing over time.
- For any  $t \in I$ , we have  $X_{t+1} \subseteq N[X_t]$ . There is no message forwarding over distances greater than 1 within one step (one unit of time) possible.
- There is a given *broadcasting scheme* (a procedure) that assigns a new vertex subset  $X_{t+1} \subseteq V$  with  $X_t \subseteq X_{t+1}$  to any graph  $G = (V, E)$  and any vertex subset  $X_t \subseteq V$ . The broadcasting scheme might be deterministic or random.

In this chapter, we investigate different broadcasting schemes. We focus on mathematical modeling rather than protocol implementation details. The main objective is to distribute a message as fast as possible over the whole network. On the other hand the message distribution should be performed efficiently with respect to network load. Well-established broadcast algorithms for computer networks and public switched telephone networks are inappropriate for ad-hoc networks as they assume complete knowledge about the network topology.

Observe that we consider broadcasting here in a more general setting as introduced in other contexts, see for instance [6]. The main difference to some well-investigated broadcasting schemes is that we allow the simultaneous forwarding from one vertex to all its neighbors.

## 2 Random Broadcasting

A well-known method in order to reduce the network load as well as hidden node problems in broadcasting is the application of randomness. In a *random broadcasting scheme*, each informed vertex forwards the message with a certain probability to one ore

more of its neighbors. There are different problems in social science and epidemiology that are closely related to broadcasting. A first problem related to broadcasting is *rumor spreading*. At time 0 one initial vertex knows a “rumor” that it wants to propagate to all other vertices of the graph. Each vertex that receives the rumor decides randomly to propagate it further or to ignore it. The probability of spreading might depend on the number of neighbors that already know the rumor.

Let us consider the following simple broadcasting scheme. Let  $G = (V, E)$  be a simple undirected graph and  $s \in V$ . At time 0 there is exactly one informed vertex  $s$ . In each time step the message is independently transferred along the edges of the graph with probability  $p$ . Consequently, if the network is a star with  $n$  vertices of degree 1 and the start vertex  $s$  is the central vertex of the star, then the probability  $P_k$  that exactly  $k$  outer vertices of the star are informed at time  $t = 1$  is

$$P_k = \binom{n}{k} p^k (1-p)^{n-k}.$$

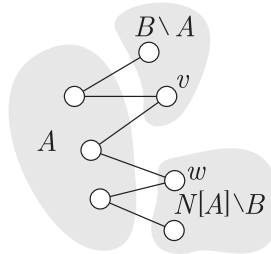


Figure 2: Calculation of the transition probability

More generally, we can describe the broadcasting process  $\{X_t\}$  as a homogeneous Markov chain with state space  $2^V$  (the power set of  $V$ ). The transition probability is given by

$$\begin{aligned} p_{AB} &= \Pr(\{X_t = B\} \mid \{X_{t-1} = A\}) \\ &= \begin{cases} 0 & \text{if } A \not\subseteq B \text{ or } B \not\subseteq N[A], \\ \prod_{v \in B \setminus A} [1 - (1-p)^{|N(v) \cap A|}] \prod_{w \in N[A] \setminus B} (1-p)^{|N(w) \cap A|} & \text{otherwise.} \end{cases} \end{aligned}$$

Figure 2 explains the derivation of the transition probability. A vertex  $v \in B \setminus A$  has to be informed at time  $t$ . Hence we must ensure that at least one of its already informed neighbors in  $A$  forwards the message to  $v$ . There are  $|N(v) \cap A|$  edges linking a vertex in  $A$  to  $v$ , which gives the first product of  $p_{AB}$ . In order to avoid that any vertex not belonging to  $B$  gets the message, all edges leading from a vertex in  $A$  to a vertex in  $N[A] \setminus B$  have to be blocked, which provides the second product. The number of states of the Markov chain grows exponentially with the order of the graph, which renders the analysis of the Markov chain a computational hard problem, see also [9].

Figure 3 shows the result of a simulation of a random broadcasting process in a graph with 70 vertices and 245 edges. In this simulation, the broadcasting process starts in a vertex of eccentricity 6. The presented diagram shows the time  $T$  that is necessary in order to inform all vertices assuming that the message forwarding probability along the edges is  $p$ . This simulation does not consider any blocking effects as caused by the hidden node problem. Observe that for  $p > 0$  the expected cover time is finite for any connected graph.

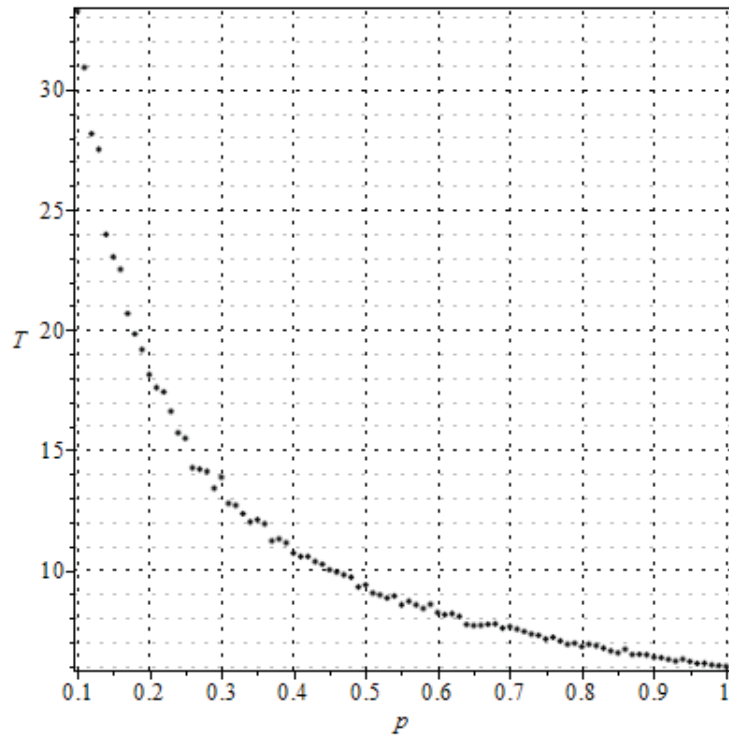


Figure 3: Simulation results for the expected cover time

### A Vertex-oriented Model

We assume now that each vertex independently sends an information that it has received at time  $t$  at each time  $t' > t$  with probability  $p$  to all its neighbors. In contrast to the edge model, if a vertex sends then all of its neighbors get simultaneously informed. In order to include the hidden-node problem, we assume that a vertex is not informed if it simultaneously gets the information from two non-adjacent vertices. Consider, as an example, the graph shown in Figure 4. We assume that at time 0 the only informed

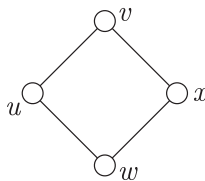


Figure 4: An example for the vertex-oriented model

vertex is  $u$ . Let  $T_A$  be the expected time until all vertices are informed on condition that at time 0 exactly the vertices contained in  $A$  are informed, where  $A \subseteq V$  and  $u \in A$ . Clearly, we have  $T_V = 0$ . The only possible candidate for  $A$  besides  $\{u\}$  and  $V$  is  $B = \{u, v, w\}$ . As we deal with hitting times in a Markov chain we obtain the system of linear equations

$$\begin{aligned} T_{\{u\}} &= pT_B + (1-p)T_{\{u\}} + 1 \\ T_B &= (1-2p+2p^2)T_B + 1 \end{aligned}$$

with the solution

$$T_{\{u\}} = \frac{3-2p}{2p-2p^2}.$$

This function has a local minimum at  $p_0 = \frac{3-\sqrt{3}}{2} \approx 0.634$ . The minimal expected time is  $T_{\{u\}}(p_0) = 2 + \sqrt{3} \approx 3.732$ .

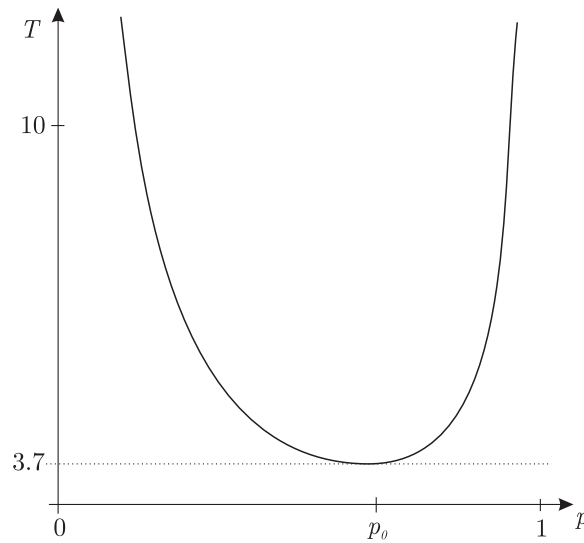


Figure 5: Hitting time

### 3 Rainbow Connection in Graphs

In this section, we will show that edge colorings of graphs with special properties provide suitable mathematical models for random broadcasting processes in networks. In this section, we introduce the necessary notions from graph coloring theory.

Let  $G = (V, E)$  be an undirected graph,  $k$  a positive integer, and  $\phi : E \rightarrow \{1, \dots, k\}$  a map that assigns a *color* to each edge of  $G$ . We call the mapping  $\phi$  an *edge  $k$ -coloring* of  $G$ . A *rainbow path* in  $G$  is a path such that all of its edges have different color. The graph  $G$  is *rainbow connected* if there exists a rainbow path between any two vertices of  $G$ . In this case the edge coloring  $\phi$  is called a *rainbow coloring* of  $G$ . The rainbow coloring is called a *rainbow  $k$ -coloring* if it uses at most  $k$  colors. The *rainbow connection number*  $rc(G)$  of a graph  $G$  is the minimum positive integer  $k$  such that  $G$  has a rainbow  $k$ -coloring, see [4]. Figure 6 shows a rainbow connected graph. In this graph, the

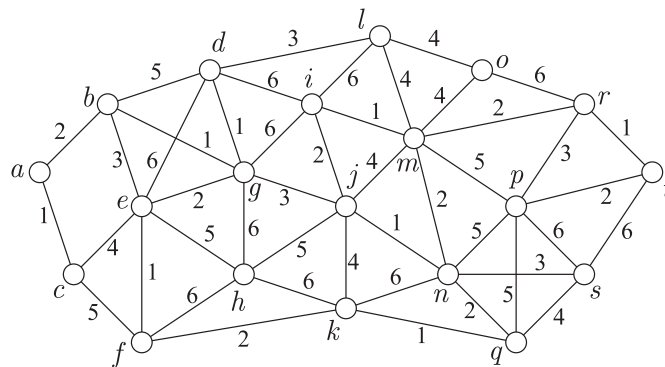


Figure 6: A graph with a rainbow coloring

path  $(a, b, d, l, o, r, t)$  (given as a vertex sequence) is a rainbow path connecting the two vertices  $a$  and  $t$ .



We use in the following the notations  $n$  for the order (number of vertices) and  $m$  for the size (number of edges) of a given graph  $G$ . Whenever  $H = (V, F)$  is a spanning subgraph of  $G = (V, E)$ , we have  $\text{rc}(G) \leq \text{rc}(H)$ . A rainbow colored tree requires as many colors as it has edges, which gives  $\text{rc}(T) = n - 1$  for any tree  $T$  of order  $n$ . The complete graph is the unique simple graph of order  $n$  with rainbow connection number 1. It has been shown in [17] that any connected graph  $G$  with minimum degree  $\delta(G) \geq 3$  satisfies  $\text{rc}(G) < \frac{3n}{4}$ . Let  $G$  be a connected graph of order  $n$  with  $m$  edges such that  $\binom{n-1}{2} + 1 \leq m \leq \binom{n}{2} - 1$ , then  $\text{rc}(G) = 2$ , see [8]. It can be easily verified that a cycle satisfies  $\text{rc}(C_n) = \lceil \frac{n}{2} \rceil$ . Considerably more effort is necessary in order to prove that a complete bipartite graph satisfies

$$\text{rc}(K_{s,t}) = \min\{\lceil \sqrt[3]{t} \rceil, 4\},$$

see [4].

**Theorem 1.** *Let  $G = (V, E)$  be an edge colored graph. Assume that  $e \in E$  is an edge of  $G$  that is colored with color  $c$  and that all other edges of  $G$  have a color different from  $c$ . Then  $G$  is rainbow connected if and only if  $G/e$  is rainbow connected.*

*Proof.* Assume that  $G$  is rainbow connected. Then clearly  $G/e$  is rainbow connected, too, as all rainbow paths of  $G$  that do not use  $e$  are also rainbow paths of  $G/e$ ; rainbow paths of length  $l$  traversing  $e$  are transformed into rainbow paths of  $G/e$  of length  $l - 1$ . Observe that this statement remains true when  $e$  uses a color that is also used by other edges of  $G$ .

Now suppose that  $G/e$  is rainbow connected. Let  $u$  and  $v$  be the end vertices of  $e$  in  $G$  and  $x$  the vertex of  $G/e$  that arises by merging  $u$  and  $v$ . Consider a rainbow path  $P$  connecting two vertices  $s, t$  of  $G/e$  that traverses  $x$ . If  $P$  is also a rainbow path in  $G - e$  then  $P$  does not traverse  $e$ , hence it is a rainbow path in  $G$ , too. In case that  $P$  is not a rainbow path of  $G/e$ , the path  $P$  splits into two rainbow paths, say the first one between  $s$  and  $u$  and the other one between  $v$  and  $t$ . The color sets of these two rainbow paths are disjoint. Consequently, we obtain by inserting  $e$  a rainbow path of  $G$ , which gives the statement.  $\square$

Let  $s$  be a vertex of the graph  $G = (V, E)$  and  $\phi : E \rightarrow \{1, \dots, k\}$  an edge coloring of  $G$ . The graph  $G$  is called *source rainbow connected* with respect to  $s$  if there exist rainbow paths from  $s$  to any other vertex of  $G$ . An edge colored graph  $G = (V, E)$  is called source rainbow connected with respect to  $S$ ,  $S \subseteq V$ , if  $G$  is source rainbow connected with respect to any vertex of  $S$ . We define  $\text{rc}_S(G)$  as the smallest positive integer  $k$  such that there is an edge  $k$ -coloring of  $G$  that makes  $G$  source rainbow connected with respect to  $S$ . Let  $k$  be a positive integer. The  *$k$ -source rainbow number* of  $G$  is defined by

$$\text{sc}_k(G) = \max\{\text{rc}_S(G) \mid S \subseteq V, |S| = k\}.$$

Clearly, we have the relation  $\text{sc}_k(G) \leq \text{rc}(G)$  for any graph  $G$  with equality for  $k = n$ . As  $\text{rc}_{\{v\}}(G) = \text{ecc}(v)$  for any vertex  $v \in V$ , we find  $\text{sc}_1(G) = \text{diam}(G)$ . A star  $S_n$  (a tree with one vertex of degree  $n$  and  $n$  leaves) shows that the difference of  $\text{sc}_1(G)$  and  $\text{sc}_n(G)$  can be arbitrarily large, as we have  $\text{sc}_1(G) = 2$  and  $\text{sc}_n(G) = n$  for all  $n \geq 2$ . More generally, any tree  $T$  with  $m$  edges has  $\text{rc}(T) = m$ .

Let  $s$  and  $t$  be two given vertices of graph  $G = (V, E)$  and  $\phi : E \rightarrow \{1, \dots, k\}$  an edge coloring of  $G$ . It is shown in [2] that the problem of deciding whether there exists

a rainbow path between  $s, t \in V$  is NP-complete. A general introduction to rainbow connection of graphs is given in [3, 10].

A *proper edge coloring* of a graph  $G = (V, E)$  is an edge coloring of  $G$  such that any two edges with a common end vertex are colored differently. Let  $G$  be a properly colored graph of diameter 2. Then there exists a path of length two between any two non-adjacent vertices of  $G$  such that its edges are colored differently. Consequently,  $G$  is rainbow connected. However, the converse is not true – there are improper rainbow colorings, even in graphs of diameter 2.

Let  $G = (V, E)$  be a rainbow connected graph of order  $n$  with an edge coloring  $\phi : E \rightarrow \{1, \dots, k\}$ . What is the minimum size (number of edges) that  $G$  can have? This problem has been investigated in [1, 12, 16]. Some important results are the following statements. The minimum size of a rainbow connected graph of order  $n$  that is edge colored with at most  $k$  colors is denoted by  $t(n, k)$ .

**Theorem 2** ([16]). *Let  $G$  be a connected graph of order  $n$ . Then*

$$\begin{aligned} t(n, 1) &= \binom{n}{2}, \\ t(n, k) &= n \text{ for } \frac{n}{2} \leq k \leq n - 2, \\ t(n, n - 1) &= n - 1. \end{aligned}$$

For  $k = 2$  the following inequality is satisfied:

$$t(n, 2) \leq \lfloor \log_2 n \rfloor (n + 1) - 2^{\lfloor \log_2 n \rfloor + 1} + 2.$$

**Theorem 3** ([12]). *If  $k, n \geq 3$ , then  $t(n, k) = \lceil \frac{k(n-2)}{k-1} \rceil$ .*

In order to describe the connection between network broadcast problems, we introduce a generalization of rainbow colorings of graphs. Let  $G = (V, E)$  be a graph and  $k$  a positive integer. An *edge list coloring* of  $G$  is a map  $\Phi : E \rightarrow 2^{\{1, \dots, k\}}$  that assigns a set of colors from  $\{1, \dots, k\}$  to each edge of  $G$ . A list edge colored graph  $G$  is said to be *list rainbow connected* if there is a rainbow path between any two vertices of  $G$  such that the color of each edge  $e$  of the path is selected from  $\Phi(e)$ .

### 3.1 Rainbow Independent Sets

What is the maximum number of edges that can be colored alike in a rainbow coloring of a graph? This question leads to the following notion. Let  $F \subseteq E$  be a set of edges of a graph  $G = (V, E)$ . The set  $F$  is called *rainbow independent* or short *r-independent* if there exists a rainbow coloring of  $G$  that colors all edges of  $F$  with the same color. An edge set  $F \subseteq E$  that is not r-independent is called *r-dependent*. Obviously there are no r-dependent singletons. A maximum r-independent set might contain only a single edge (in trees) or even all edges of the graph (for complete graphs). Here *maximum* is meant with respect to cardinality. A *maximal* r-independent set  $F$  is a r-independent set that is not a proper subset of an r-independent set. A maximal r-independent set is not necessarily a maximum r-independent set.

**Theorem 4.** *Let  $G = (V, E)$  be a connected graph and  $A = \{e, f\} \subseteq E$ . Then  $A$  is r-independent if and only if at least one of the edges  $e, f$  is not a bridge of  $G$ .*

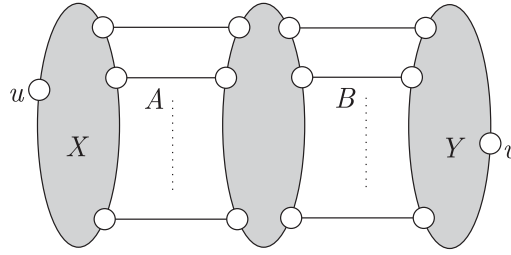


Figure 7: To the proof of Theorem 5

*Proof.* Assume that the edges of  $G$  are colored all differently with the exception of  $e$  and  $f$ , which are colored alike. Let  $e = \{u, v\}$  be an edge that is not a bridge, which is colored with color  $c$ . Then there is a rainbow path between  $u$  and  $v$  that does not use color  $c$ . Consequently, each path in  $G$  that uses both edges  $e$  and  $f$  can be replaced by a path that does not use  $e$ . The graph  $G$  is rainbow connected as this path has no edges of the same color.

Now assume that  $e = \{u, v\}$  and  $f = \{x, y\}$  are bridges of  $G$  such that  $u$  and  $y$  are in different components of  $G - v$  and of  $G - x$ . We allow also the case  $v = x$ . Then there is a unique path from  $u$  to  $y$  in  $G$  that contains the two edges  $e$  and  $f$ . This is not a rainbow path as these edges are colored alike.  $\square$

**Theorem 5.** Let  $G = (V, E)$  be a connected graph and  $A, B \subseteq E$  with  $A \cap B = \emptyset$ . If  $A$  and  $B$  are cuts of  $G$  then  $A \cup B$  is an  $r$ -dependent set of  $G$ .

*Proof.* Let  $X, Y \subseteq V$  be the vertex subsets that generate the cuts  $A, B$ :

$$\begin{aligned} A &= \{\{x, y\} \in E \mid x \in X, y \in V \setminus X\} \\ B &= \{\{x, y\} \in E \mid x \in Y, y \in V \setminus Y\} \end{aligned}$$

see Figure 7. Now assume that all edges of  $A \cup B$  are colored alike. Then there can be no rainbow path between a vertex  $u \in X$  and a vertex  $v \in Y$ .  $\square$

The following properties of  $r$ -independent and  $r$ -dependent sets of connected graphs are easily verified:

- Let  $G = (V, E)$  be a graph and  $(X, F)$  a clique in  $G$ . Then  $F$  is an  $r$ -independent set of  $G$ .
- Let  $(V, F)$  be a spanning tree of a graph  $G = (V, E)$  and  $f \in F$ . Then  $(E \setminus F) \cup \{f\}$  is an  $r$ -independent set of  $G$ .
- Any nonempty set of bridges of a graph is  $r$ -dependent.
- If  $A$  is an  $r$ -independent set of a graph  $G$  and  $B \subseteq A$ , then  $B$  is  $r$ -independent, too.
- Any superset of an  $r$ -dependent set is  $r$ -dependent.
- Let  $G$  be a graph with  $m$  edges. If the largest  $r$ -independent set of  $G$  has  $k$  edges, then  $\text{rc}(G) \leq m - k + 1$ .

In contrast to independent vertex or edge sets in graphs,  $r$ -independent sets of graphs cannot be independently chosen. Consider, as an example, the cycle  $C_6$  with edge set  $E$ . Any edge subset  $A$  of  $E$  with  $|A| = 3$  forms an  $r$ -independent set and so does its complement  $B = E \setminus A$ . However, the cycle  $C_6$  with all vertices of  $A$  colored red and all vertices of  $B$  colored blue is not rainbow connected.

**Theorem 6.** *Let  $F$  be an edge subset of a cycle  $C_n$  of length  $n \geq 3$ . The set  $F$  is  $r$ -independent if and only if  $|F| \leq 3$ .*

*Proof.* Assume  $|F| = k$  and let  $\phi : E(C_n) \rightarrow \{1, \dots, n - k + 1\}$  be an edge coloring of  $C_n$  such that  $\phi(e) = \phi(f)$  if and only if  $e, f \in F$ . Consequently, a color of an edge of  $E \setminus F$  appears exactly once in  $C_n$ . According to Theorem 1,  $C_n$  is rainbow connected if and only if a monochromatic cycle  $C_k$  is rainbow connected. This is true for  $k = 3$  and false for any  $k > 3$ .  $\square$

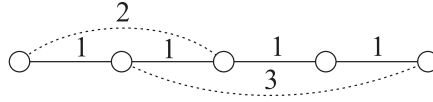


Figure 8: A monochromatically colored  $P_5$  and its edge extension

Let  $F$  be a nonempty edge subset of the complete graph  $K_n$ . Consequently,  $F$  is an  $r$ -independent set. What is the maximum number of edges that we can remove such that  $F$  is still  $r$ -independent in the remaining graph? We can reverse this question. Let  $G = (V, F)$  be a non-complete graph that is monochromatically edge colored. What is the minimum number of edges that have to be inserted in  $G$  in order to render  $F$   $r$ -independent, i.e. to make  $G$  rainbow connected? Consider, as an example, the monochromatic path presented in Figure 8. The four edges of the path are colored with color 1. The two inserted edges colored with colors 2 and 3 yield a rainbow connected graph. The combination of the Theorems 4, 5 and 6 shows that the insertion of only one edge is insufficient in order to make the edge set of the path  $r$ -independent.

### 3.2 Rainbow Partitions

Let  $\Pi(E)$  be the set of all partitions of the edge set of a graph  $G = (V, E)$ . A partition  $\pi$  is called a *refinement* of a partition  $\sigma$  if each block of  $\pi$  is a subset of a block of  $\sigma$ . We order the partitions of  $\Pi(E)$  such that  $\pi \leq \sigma$  is satisfied if and only if  $\pi$  is a refinement of  $\sigma$ .  $\Pi(E)$  together with this ordering relation defines a lattice – the *partition lattice* of  $E$ . The minimal element  $\hat{0}$  of  $\Pi(E)$  is the partition that consists entirely of singletons. The maximal element  $\hat{1}$  has exactly one block, namely  $E$  itself.

A partition  $\pi \in \Pi(E)$  is called a *rainbow partition* if we obtain a rainbow coloring of  $G$  by coloring exactly those edges alike that belong to one block of  $\pi$  but edges of different blocks with different colors. We denote the set of all rainbow partitions of  $G$  by  $\Pi_r(G)$ . As a consequence of the definition of rainbow partitions, we conclude that  $\pi \in \Pi_r(G)$  and  $\sigma \leq \pi$  imply  $\sigma \in \Pi_r(G)$ . The minimal element  $\hat{0}$  is a rainbow partition for any connected graph. All partitions of  $\Pi(E)$  are also rainbow partitions in the complete graph. Figure 9 shows the ordered set of all rainbow partitions of a cycle. This ordered set is (for any given graph) a lower semilattice, i.e. for any two rainbow partitions  $\pi, \sigma$  the *join*, defined by

$$\pi \wedge \sigma = \max\{\tau \in \Pi_r(G) \mid \tau \leq \pi, \tau \leq \sigma\},$$

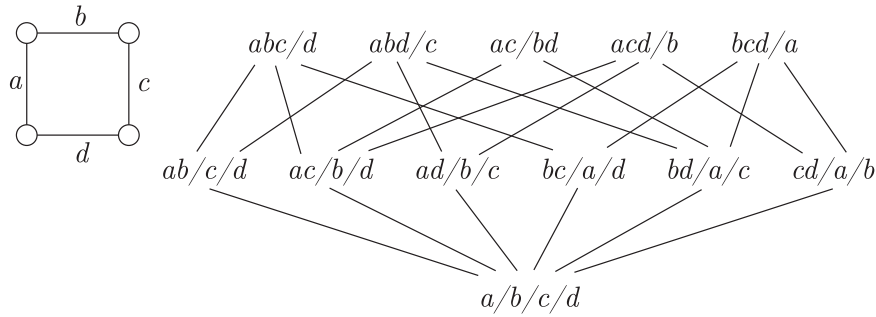


Figure 9: A cycle  $C_4$  together with its semilattice of rainbow partitions

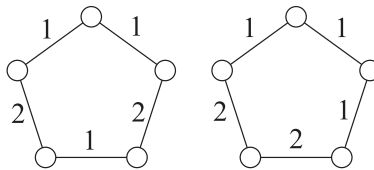


Figure 10: A cycle  $C_5$  has no rainbow coloring of type  $\{\{1, 1, 1\}, \{2, 2\}\}$

is also a rainbow partition.

For any maximum  $r$ -independent set  $F$  of  $G$  there is a maximal element in  $\Pi_r(G)$  that contains  $F$  as a block; the converse is not true in general, see the example given in Figure 9. The rainbow connection number of a graph satisfies

$$rc(G) = \min\{|\pi| \mid \pi \in \Pi_r(G)\}.$$

Let  $\pi$  be a rainbow partition of the graph  $G = (V, E)$  and  $e \notin E$ . Then any partition  $\sigma$  that is obtained from  $\pi$  by inserting  $e$  into one block of  $\pi$  or by appending the block  $\{e\}$  to  $\pi$  is a rainbow partition of the graph  $G + e$ .

**Theorem 7.** *Let  $n \geq 3$  and  $\pi \in \Pi_r(C_n)$  a rainbow partition of the cycle  $C_n$ . Then exactly one of the following alternatives applies to  $\pi$ :*

- *One block of  $\pi$  has three elements and all other blocks are singletons.*
- *There exist exactly  $k$ ,  $0 \leq k \leq n/2$ , two-element blocks in  $\pi$ ; all other blocks are singletons.*

*Proof.* According to Theorem 6, there can be no block with four or more elements in  $\pi$ . Figure 10 shows (up to symmetry) all possible edge colorings of a cycle  $C_5$  with two colors such that one color is three times used and the other one twice. It is easy to check that none of these colorings is a rainbow coloring. Applying Theorem 1, we can conclude that no cycle  $C_n$ ,  $n \geq 5$  can have a rainbow partition with one block of cardinality 3 and one block of cardinality 2.

The only way to generate a rainbow coloring of a cycle  $C_4$  with two colors such that each color is used twice is to color non-adjacent edges of the  $C_4$  alike. Now let  $E = \{e_0, \dots, e_{2n-1}\}$  be the edge set of a cycle  $C_{2n}$  indexed according to the order of traversal along the cycle. We can easily verify that the color assignment  $c(e_i) = c(e_{i+n}) = i + 1$  for  $i \in \{0, \dots, n - 1\}$  results in a rainbow coloring of the  $C_{2n}$ . However, the color classes are unique – any different distribution of colors yields an edge coloring that is not rainbow. In order to see this, assume there are four indices  $i < j < k < l$  (taken mod  $2n$ ) such

that  $c(e_i) = c(e_l)$  and  $c(e_j) = c(e_k)$ . Then by contracting all other edges we obtain a non-rainbow colored  $C_4$ , which shows again by Theorem 1 that the  $C_{2n}$  is not rainbow colored.  $\square$

### 3.3 The Rainbow Polynomial

Let  $x$  be a nonnegative integer. The *rainbow polynomial*  $\rho(G, x)$  of a graph  $G$  is the number of rainbow colorings of  $G$  with at most  $x$  colors. This definition immediately implies:

- If  $G = (\{v\}, \emptyset)$  is a graph with exactly one vertex, then  $\rho(G, x) = 1$ .
- If  $G$  is disconnected, then  $\rho(G, x) = 0$ .
- If  $G$  is a tree with  $n$  vertices, then  $\rho(G, x) = x^{n-1}$ .
- The rainbow polynomial of the complete graph is

$$\rho(K_n, x) = x^{\binom{n}{2}}.$$

The rainbow connection number of a graph  $G$  is the smallest nonnegative integer that is not a root of  $\rho(G, x)$ :

$$\text{rc}(G) = \min\{x \in \mathbb{N} \mid \rho(G, x) > 0\}$$

A rainbow coloring of a graph is invariant with respect to insertion of edges, which yields the next statement. Let  $G = (V, E)$  be a graph with  $m$  edges and  $F \subseteq E$ . Then the relation

$$x^{|F|} \rho(G - F, x) \leq \rho(G, x) \leq x^m$$

is valid for any nonnegative integer  $x \in \mathbb{N}$ .

**Theorem 8.** *Let  $G = (V, E)$ ,  $G^1 = (V^1, E^1)$  and  $G^2 = (V^2, E^2)$  be connected graphs such that  $V^1 \cup V^2 = V$ ,  $E^1 \cup E^2 = E$ ,  $V^1 \cap V^2 = S$  and  $E^1 \cap E^2 = \emptyset$  are satisfied. Then for any  $x \in \mathbb{N}$ ,*

$$\rho(G, x) \geq \max\{\rho(G^1, x)\rho(G^2, x - |E^1|), \rho(G^2, x)\rho(G^1, x - |E^2|)\}.$$

*Proof.* The set  $S$  is a vertex separator of  $G$ . Since the two subgraphs  $G^1$  and  $G^2$  are edge disjoint, we can color them independently. There are  $\rho(G^1, x)$  different rainbow colorings of  $G^1$  with  $x$  colors. A rainbow coloring of  $G^1$  can safely extended to a rainbow coloring of  $G$  if we avoid using colors that have been used in  $G^1$ . There remain  $x - |E^1|$  colors for the edges of  $G^2$  as there are at most  $|E^1|$  different colors already in use. Exchanging the roles of  $G^1$  and  $G^2$  yields the second term of the maximum.  $\square$

**Theorem 9.** *Let  $G = (V, E)$  be a connected graph and  $A \subseteq E$  the set of bridges of  $G$ . Let  $G/A$  be the graph obtained from  $G$  by contraction of all its bridges. Then*

$$\rho(G, x) \geq x^{|A|} \rho(G/A, x - |A|).$$

*Proof.* Any two bridges must have different colors in a rainbow coloring of  $G$ . Thus there are  $x^{|A|}$  possibilities to color the  $|A|$  bridges with  $x$  colors. If the set of colors used in  $G - A$  is disjoint from the set of colors that are used for bridges in  $A$  and  $G/A$  is rainbow connected then clearly  $G$  itself is rainbow connected, too. However, there might be edge colorings making  $G$  rainbow connected even if colors of  $A$  are reused in the remaining graph.  $\square$



**Theorem 10.** *Let  $e$  be an edge of a connected graph  $G$ . Then for any  $x \in \mathbb{N}$ ,*

$$\rho(G, x) \geq x\rho(G/e, x - 1).$$

*Proof.* Let  $e = \{u, v\}$  be the edge to be contracted. Any path of  $G$  that does not traverse the edge  $e$  is still a path in  $G/e$ . A path  $P$  traversing  $e$  is transformed into a shorter path  $P'$  of  $G/e$ . Now assume that there are  $x$  colors available to color the edges of  $G$ . First, we color  $e$  with one of the  $x$  colors, say with color  $c$ . Then we contract  $e$  and construct a rainbow coloring of  $G/e$  avoiding the color  $c$ . There are exactly  $\rho(G/e, x - 1)$  different rainbow colorings of  $G/e$  that do not use the color  $c$ . If the path  $P'$  in  $G/e$  is rainbow colored, then  $P$  is rainbow colored in  $G$  assuming  $e$  is  $c$ -colored. Hence there are at least  $x\rho(G/e, x - 1)$  rainbow colorings for  $G$ .  $\square$

Let  $r_i(G)$  be the number of rainbow partitions of  $G$  with exactly  $i$  blocks. The *rainbow generating function*  $r(G, x)$  is the ordinary generating function for the number of rainbow partitions of  $G$ :

$$r(G, x) = \sum_{\pi \in \Pi_r(G)} x^{|\pi|} = \sum_{i=0}^m r_i(G)x^i.$$

In the following the (unsigned) *Stirling numbers of the first kind* are denoted by  $\left[ \begin{smallmatrix} n \\ k \end{smallmatrix} \right]$ . They give the number of permutations of an  $n$ -set with exactly  $k$  cycles. The rainbow polynomial can be easily derived from the rainbow generating function:

$$\rho(G, x) = \sum_{i=0}^m r_i(G)x^i \tag{1}$$

$$= \sum_{k=0}^m \sum_{j=0}^k \left[ \begin{smallmatrix} k \\ j \end{smallmatrix} \right] (-1)^{k-j} r_k(G)x^j. \tag{2}$$

The complete graph  $K_n$  with  $m = \binom{n}{2}$  edges satisfies

$$r(K_n, x) = \sum_{k=1}^m \left\{ \begin{smallmatrix} m \\ k \end{smallmatrix} \right\} x^k$$

as any partition of the edge set is a rainbow partition. Here  $\left\{ \begin{smallmatrix} m \\ k \end{smallmatrix} \right\}$  are the *Stirling numbers of the second kind*, which give the number of partitions of an  $m$ -set with exactly  $k$  blocks. The following result is an immediate consequence of Theorem 7.

**Theorem 11.** *The rainbow generating function of the cycle  $C_n$ ,  $n \geq 3$ , is*

$$r(C_n, x) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k} x^{n-k} + \binom{n}{3} x^{n-2}.$$

Substituting  $x$  by 1 in this polynomial, we find that the cycle  $C_n$  has  $2^{n-1} + \binom{n}{3}$  different rainbow partitions. Theorem 11 together with Equation (2) yield the rainbow polynomial of the cycle  $C_n$ .

**Theorem 12.** *Let  $G = (V, E)$  be a connected graph of order  $n$  that consists completely of cycles that have at most one vertex in common. Hence,  $G$  is a cactus without any bridges. Figure 11 shows a graph of this kind. Then for any  $k \in \mathbb{N}$*

$$\rho(C_n, k) \leq \rho(G, k).$$

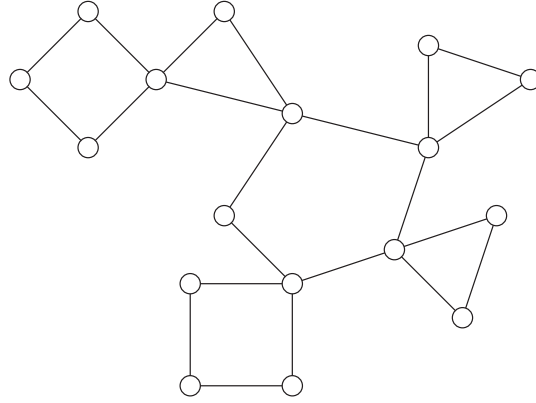


Figure 11: A bridgeless cactus

*Proof.* The following two easy observations suffice to prove the theorem:

- Each bridgeless cactus with  $n$  edges can be obtained from a cycle  $C_n$  by a sequence of merges of non-adjacent vertices.
- Let  $H$  be a graph with two non-adjacent vertices  $u$  and  $v$  and  $H_{uv}$  the graph obtained from  $H$  by merging  $u$  and  $v$ . Then each rainbow path of  $H$  corresponds to a (possibly shorter) rainbow path in  $H_{uv}$ .

□

### 3.4 Random Edge Colorings

Now we consider again the model of random broadcasting. At time  $t = 0$  a single vertex  $s$  of a given graph  $G = (V, E)$  is informed. In each successive time step each informed vertex informs (simultaneously) its neighbors with probability  $p$  along the incident edges. This broadcasting process can be simulated as follows. First we fix a limit time  $t = k, k > 0$ . Then we select randomly  $k$  edge subsets  $F_1, \dots, F_k$  according to the probability distribution

$$\Pr(F_i = F) = p^{|F|}(1-p)^{|E|-|F|}. \quad (3)$$

In case of non-identical broadcast probabilities, the distribution changes to

$$\Pr(F_i = F) = \prod_{e \in F} p_e \prod_{f \in E \setminus F} (1 - p_f).$$

Next we define a list edge coloring  $\Phi : \rightarrow 2^{\{1, \dots, k\}}$  by

$$\Phi(e) = \{i \in \{1, \dots, k\} \mid e \in F_i\}.$$

The interpretation of this list edge coloring is as follows. The colors in a set  $\Phi(e)$  represent time steps where the edge  $e$  is *open*, that is  $e$  can be traversed by the message in order to inform a neighboring vertex. It does not matter whether an end vertex of  $e$  is already informed. The membership  $e \in F_i$  just indicates that an information *could* be transferred along  $e$  at time  $i$ . Now the decision whether the broadcast time is at most  $k$  is easy. We just verify that there is for each vertex  $x \in V$  a path

$$P_x = (s, e_1, v_1, \dots, e_{j-1}, v_{j-1}, e_j, v_j = x)$$



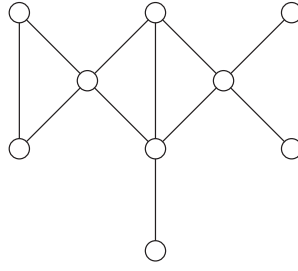


Figure 12: A graph of order 9 with 11 edges

and a sequence  $(c_1, \dots, c_j)$  of colors with  $c_1 \in \Phi(e_1), \dots, c_j \in \Phi(e_j)$  and

$$c_1 < c_2 < \dots < c_j.$$

It is already sufficient to require that the colors  $c_1, \dots, c_j$  are *different*. If there exists a sequence of different colors, then there also exists a monotone increasing sequence of colors with the same probability.

First we consider a random edge coloring where a color for each edge is randomly selected from the color set  $C = \{1, \dots, k\}$ . An edge coloring of this kind might also be considered as a list coloring such that all lists are singletons. Let  $G = (V, E)$  be a graph with  $m$  edges. Then there are  $k^m$  different edge colorings with colors from  $C$  for  $G$ . Let  $P_{rc}(G, k)$  be the probability that a random edge coloring out of a set of  $k$  colors generates a rainbow connected graph. Using the rainbow polynomial, we obtain

$$P_{rc}(G, k) = \frac{\rho(G, k)}{k^m}.$$

This equation yields  $P_{rc}(K_n, k) = 1$  for the complete graph and  $P_{rc}(P_n, k) = \frac{k^{n-1}}{k^{n-1}}$  for a path with  $n$  vertices. The probability of choosing identical colors decreases with increasing number of colors, which results in

$$\lim_{k \rightarrow \infty} P_{rc}(G, k) = 1$$

for any finite connected graph  $G$ .

The graph presented in Figure 12 has the rainbow polynomial

$$\begin{aligned} \rho(G, x) = & x^{11} - 3x^{10} - 23x^9 - 219x^8 - 30x^7 + 39059x^6 - 306558x^5 \\ & + 1045039x^4 - 1831946x^3 + 1604184x^2 - 549504x. \end{aligned}$$

We find  $\rho(G, x) = 0$  for  $x \in \{0, 1, 2, 3, 4\}$  and  $\rho(G, 5) = 535080$  such that  $rc(G) = 5$  follows. Figure 13 shows the probability that the randomly edge colored graph of Figure 12 is rainbow connected as a function of the number of colors.

The following statement is closely related to generalized coupon collector problems, see [18] and [22].

**Theorem 13.** *Let  $G = (V, E)$  be a graph with edge set  $E = \{1, \dots, m\}$  and  $j, k$  positive integers with  $j \leq k$ . We draw independently  $m$  random  $j$ -subsets  $L_1, \dots, L_m$  out of  $\{1, \dots, k\}$ , which define an edge list coloring of  $G$ . Then the probability that each color of  $\{1, \dots, k\}$  appears at least once among the edges of  $G$  is*

$$\Pr \left( \left\{ \left| \bigcup_{e \in E} L_e \right| = k \right\} \right) = \frac{1}{\binom{k}{j}^m} \sum_{i=0}^{k-j} (-1)^i \binom{k}{i} \binom{k-i}{j}^m.$$

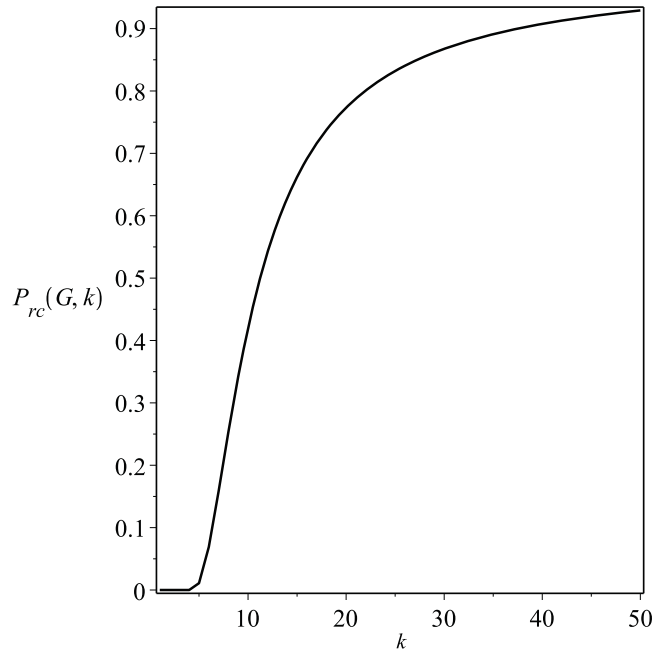


Figure 13: A graph of order 9 with 11 edges

*Proof.* The number of all possible assignments of  $m$  color sets of size  $j$  to the edges of  $G$  is  $\binom{k}{j}^m$ , which gives the denominator on the right-hand side. We define random events  $A_X$  for all  $X \subseteq \{1, \dots, k\}$  by

$$A_X = \left\{ X \not\subseteq \bigcup_{e \in E} L_e \right\}.$$

Notice that  $A_X \cap A_Y = A_{X \cup Y}$  for all  $X, Y \subseteq \{1, \dots, k\}$ . Then the probability requested in the theorem is

$$\begin{aligned} \Pr \left( \left\{ \left| \bigcup_{e \in E} L_e \right| = k \right\} \right) &= 1 - \Pr \left( \left\{ \bigcup_{i=1}^k A_{\{i\}} \right\} \right) \\ &= \sum_{X \subseteq \{1, \dots, k\}} (-1)^{|X|} \Pr(A_X) \\ &= \sum_{X \subseteq \{1, \dots, k\}} (-1)^{|X|} \left[ \frac{\binom{k-|X|}{j}}{\binom{k}{j}} \right]^m \\ &= \frac{1}{\binom{k}{j}^m} \sum_{i=0}^k (-1)^i \binom{k}{i} \binom{k-i}{j}^m, \end{aligned}$$

which coincides with the statement of the theorem as the last binomial coefficient vanishes for  $i > k - j$ .  $\square$

The next statement is a direct consequence of this proof.

**Corollary 14.** *Let  $j, k, m$  be positive integers with  $j \leq m \leq k$ . There are*

$$\sum_{i=0}^{k-j} (-1)^i \binom{k}{i} \binom{k-i}{j}^m$$

different sequences  $(L_1, \dots, L_m)$  of  $m$  (not necessarily different)  $j$ -element subsets of  $\{1, \dots, k\}$  such that

$$\bigcup_{i=1}^m L_i = \{1, \dots, k\}.$$

Assume that we assign an edge list coloring to a path  $P_{m+1}$  such that each edge receives a color list that is a random selection of  $t$  elements out of a color set of cardinality  $k$ . What is the probability that the path has a rainbow coloring that is compatible with the given lists? In order to derive an answer to this question, we need some preliminary results. Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be a family of  $m$  sets. We consider here  $\mathcal{L}$  as a multiset of sets as we allow  $L_i = L_j$ , for  $i \neq j$ . A *transversal* or a *system of distinct representatives* of  $\mathcal{L}$  is a set  $A = \{l_1, \dots, l_m\}$  of  $m$  distinct elements such that  $l_i \in L_i$ , for  $i = 1, \dots, m$ . The following statement is also known as *Hall's marriage theorem*.

**Theorem 15** ([5]). *A set family  $\mathcal{L} = \{L_1, \dots, L_m\}$  has a transversal if and only if the relation*

$$\left| \bigcup_{i \in X} L_i \right| \geq |X|$$

*is satisfied for all subsets  $X \subseteq \{1, \dots, m\}$ .*

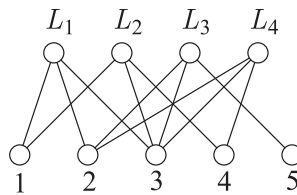


Figure 14: A bipartite graph

A set family  $\mathcal{L} = \{L_1, \dots, L_m\}$  where all sets have cardinality  $t$  and all sets are subsets of  $\{1, \dots, k\}$  can be represented as a bipartite graph  $G = (\mathcal{L} \cup \{1, \dots, k\}, E)$  where  $L_i$  is adjacent to  $j$  if and only if  $j \in L_i$ . Figure 14 shows a bipartite graph representing the set family

$$\mathcal{L} = \{\{1, 2, 3\}, \{1, 3, 4\}, \{2, 3, 5\}, \{2, 3, 4\}\}.$$

A transversal of the set family corresponds to a matching in  $G$  that saturates all vertices of  $\mathcal{L}$ .

## 4 Conclusions and Open Problems

Let  $k, m, t$  be positive integers with  $t \leq m \leq k$ . We draw independently  $m$  random  $t$ -subsets  $L_1, \dots, L_m$  out of  $C = \{1, \dots, k\}$ . Let  $p_{k,m,t}$  be the probability that there exists a transversal of the set family  $\mathcal{L} = \{L_1, \dots, L_m\}$ . Let  $x_{k,m,t}$  be the number of set families  $\{L_1, \dots, L_m\}$  with  $|L_i| = t$ , for  $i = 1, \dots, m$ , that do not have a transversal. Consequently, the probability we are looking for can be expressed by

$$p_{k,m,t} = 1 - \frac{x_{k,m,t}}{\binom{k}{t}^m}. \quad (4)$$

The challenge to compute the numbers  $x_{k,m,t}$  remains an open problem for all but a few special combinations of the parameters  $k, m, t$ . For  $t = 1$  the answer is given by the

rainbow polynomial. However, already the case  $t = 2$  requires much more effort, see the next chapter.

Interesting open problems remain to solve:

- We can easily find the number of different list edge colorings (S-lists) that are compatible with a given rainbow coloring (transversal). However, when we have a set of edge colorings instead, then a complicated version of inclusion–exclusion pattern emerges. Is there a way to describe these pattern in a closed form?
- The smallest set families that can violate the condition of Theorem 15 are families of cardinality  $t + 1$ . In general, we can select a subset  $D$  of  $t + l$  colors ( $l > 0$ ) from  $C = \{1, \dots, k\}$  and apply Corollary 14 in order to determine the number of set systems with  $t + l$  sets of cardinality  $t$  that have not all colors from  $D$ . Here again we need inclusion–exclusion to find the number of set systems without a transversal.
- Can we find the probability that a random set system with a distribution according to Equation (3) has a transversal?

## References

- [1] J.-P. Bode and H. Harborth. ‘The minimum size of  $k$ -rainbow connected graphs of given order’. In: *Discrete Mathematics* 313.19 (2013), pp. 1924–1928.
- [2] S. Chakraborty, E. Fischer, A. Matsliah, and R. Yuster. ‘Hardness and algorithms for rainbow connection’. In: *Journal of Combinatorial Optimization* 21.3 (2011), pp. 330–347.
- [3] G. Chartrand and P. Zhang. *Chromatic graph theory*. CRC press, 2009.
- [4] G. Chartrand, G. L. Johns, K. A. McKeon, and P. Zhang. ‘Rainbow connection in graphs’. In: *Mathematica Bohemica* 133.1 (2008), pp. 85–98.
- [5] P. Hall. ‘On representatives of subsets’. In: *J. London Math. Soc* 10.1 (1935), pp. 26–30.
- [6] J. Hromkovič, R. Klasing, B. Monien, and R. Peine. ‘Dissemination of Information in Interconnection Networks (Broadcasting & Gossiping)’. In: *Combinatorial network theory*. Springer, 1996, pp. 125–212.
- [7] J. Jeong, H. Kim, T. Lee, and J. Shin. ‘An analysis of hidden node problem in IEEE 802.11 multihop networks’. In: *Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on*. IEEE, 2010, pp. 282–285.
- [8] A. Kemnitz and I. Schiermeyer. ‘Graphs with rainbow connection number two.’ In: *Discussiones Mathematicae Graph Theory* 31.2 (2011), pp. 313–320.
- [9] R. Lopus, F. Simon, and P. Tittmann. ‘Random Information Spread in Networks’. In: *arXiv preprint* (2010).
- [10] X. Li, Y. Shi, and Y. Sun. ‘Rainbow connections of graphs: a survey’. In: *Graphs and Combinatorics* 29.1 (2013), pp. 1–38.

- 
- [11] J. Lipman, H. Liu, and I. Stojmenovic. ‘Broadcast in Ad Hoc Networks’. In: *Guide to Wireless Ad Hoc Networks*. Ed. by S. Misra, I. Woungang, and S. C. Misra. 2009.
- [12] A. Lo. ‘A note on the minimum size of k-rainbow-connected graphs’. In: *Discrete Mathematics* 331 (2014), pp. 20–21.
- [13] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu. ‘The broadcast storm problem in a mobile ad hoc network’. In: *MobiCom ’99 Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. 1999, pp. 151–162.
- [14] A. Rahman and P. Gburzynski. ‘Hidden Problems with the Hidden Node Problem’. In: *Communications, 2006 23rd Biennial Symposium on*. 2006, pp. 270–273.
- [15] S. Ray, D. Starobinski, and J. B. Carruthers. ‘Performance of wireless networks with hidden nodes: a queuing-theoretic analysis’. In: *Computer Communications* 28.10 (2005). Performance issues of Wireless LANs, {PANS} and ad hoc networks, pp. 1179–1192. ISSN: 0140-3664.
- [16] I. Schiermeyer. ‘On minimally rainbow k-connected graphs’. In: *Discrete Applied Mathematics* 161.4 (2013), pp. 702–705.
- [17] I. Schiermeyer. ‘Rainbow connection in graphs with minimum degree three’. 20th International Workshop, IWOCOA 2009, Hradec nad Moravici, Czech Republic, June 28–July 2, 2009. In: *Combinatorial Algorithms*. Ed. by J. Fiala, J. Kratochvil, and M. Miller. Springer, 2009, pp. 432–437.
- [18] M. Sharif and B. Hassibi. ‘Delay considerations for opportunistic scheduling in broadcast fading channels’. In: *IEEE Transactions on Wireless Communications* 6.9 (2007), pp. 3353–3363.
- [19] A. S. Tanenbaum and D. J. Wetherall. *Computernetzwerke*. Trans. by K. Pieper and P. Alm. 5th revised edition. Munich: Pearson Deutschland GmbH, 2012.
- [20] O. Tonguz, N. Wisitpongphan, F. Bai, P. Mudalige, and V. Sadekar. ‘Broadcasting in VANET’. In: *2007 Mobile Networking for Vehicular Environments*. 2007, pp. 7–12.
- [21] O. Tonguz, N. Wisitpongphan, J. Parikh, F. Bai, P. Mudalige, and V. Sadekar. ‘On the Broadcast Storm Problem in Ad hoc Wireless Networks’. In: *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*. Oct. 2006, pp. 1–11.
- [22] W. Xu, A. K. Tang, et al. ‘A generalized coupon collector problem’. In: *Journal of Applied Probability* 48.4 (2011), pp. 1081–1094.



# Transversals of Random Set Systems

XIANGYING CHEN AND PETER TITTMANN

In the preceding chapter we started the investigation of random set systems that are related to broadcasting problems in networks. Now we focus on set systems (list colorings) where all sets have cardinality 2. The main techniques employed for the analysis are the exponential formula from enumerative combinatorics and the Tutte polynomial of a graph.

## 1 Introduction

Let  $k$  be an integer greater than 1 and  $S = \{1, \dots, k\}$ . We draw at random two elements from  $S$ . This experiment is independently  $m$  times repeated, which results in a list  $\mathcal{L} = [L_1, \dots, L_m]$  of random two-element sets. We call such a random list of unordered pairs from  $S$  an  $S$ -list. A *transversal* of  $\mathcal{L}$  is a set  $T = \{t_1, \dots, t_m\}$  of  $m$  different elements such that  $t_i \in L_i$  for  $i = 1, \dots, m$ . What is the probability that a given  $S$ -list  $\mathcal{L}$  has a transversal? This probability is denoted by  $p(k, m)$ . In this paper, we present different methods for the computation of  $p(k, m)$ , consider some generalized problems, and show its relation to graph polynomials.

Obviously, there are  $\binom{k}{2}$  possibilities to choose an unordered pair of elements from  $S$ . Hence there exist  $\binom{k}{2}^m$  different  $S$ -lists of length  $m$ . Let  $f_{km}$  be the number of  $S$ -lists of length  $m$  with  $|S| = k$  that possess a transversal. Then the desired probability is given by

$$p(k, m) = \frac{f_{km}}{\binom{k}{2}^m}.$$

Consequently, we might consider the stated problem as an enumeration problem of the following kind. How many lists of  $m$  pairs from  $S$  have a transversal? There is clearly no transversal for  $\mathcal{L}$  if  $|\mathcal{L}| = m > k = |S|$ . Therefore we assume in the following that  $m \leq k$ . We consider first the case  $m = k$ . The  $S$ -list can be represented by a bipartite graph  $G = (\mathcal{L} \cup S, E)$ , where the edge set is defined by

$$E = \{\{L, s\} \mid s \in L \in \mathcal{L}\}.$$

Then any transversal of  $\mathcal{L}$  corresponds to a perfect matching of  $G$ . This implies that  $f_{kk}$  is the number of perfect matchings of  $G$ . We define the biadjacency matrix  $A_G = (a_{ij})$  by

$$a_{ij} = \begin{cases} 1 & \text{if } s_j \in L_i \\ 0 & \text{otherwise} \end{cases}.$$

As a perfect matching of  $G$  is uniquely determined by a selection of ones from the biadjacency matrix that contains from each row and each column exactly one element,

the number of perfect matchings equals the permanent of  $A_G$ :

$$\text{per } A_G = \sum_{\pi \in S_k} \prod_{i=1}^k a_{i, \pi_i}$$

Further interesting presentations and relations to other areas of mathematics can be found in [3, 9, 11]. The asymptotic enumeration of transversals has been considered in [10].

## 2 Pseudoforests

Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be a multiset of sets. We need in the following Hall's marriage theorem [7], which states that the family  $\mathcal{L}$  has a transversal if and only if the relation

$$\left| \bigcup_{i \in X} L_i \right| \geq |X|$$

is satisfied for all subsets  $X \subseteq \{1, \dots, m\}$ . We will use this theorem in order to develop a characterization of  $S$ -lists having transversals.

In the following we consider graphs that might have parallel edges but no loops. A subgraph of a given graph that consists of two vertices connected by two parallel edges is considered as a cycle  $C_2$  of length 2. A *pseudoforest* is a graph in which each component has at most one cycle.

**Theorem 1.** *Let  $G = (S, \mathcal{L})$  be a graph that is defined by a given  $S$ -list  $\mathcal{L}$ . Then  $G$  is a pseudoforest.*

*Proof.* Assume  $\mathcal{L}$  has a transversal but  $G$  has a component  $H = (W, F)$  that has two or more cycles. Then clearly the relation  $|W| < |F|$  is satisfied and hence Hall's condition is violated for the family of pairs that are represented by the edges of  $H$ , which yields a contradiction to the assumed existence of a transversal.

Now assume that  $G = (S, \mathcal{L})$  is a pseudoforest. Then we can construct a transversal of  $\mathcal{L}$  as follows. First we observe that a transversal of  $\mathcal{L}$  can be obtained as the union of transversals for the components of  $G$ . Assume that  $H$  is a component of  $G$  that is a cycle. Then the vertex set of the cycle provides the desired transversal. If  $H$  is not a cycle and has at least one edge then  $H$  has a pendant vertex  $v$  (a vertex of degree 1). We define  $T = \emptyset$  as initialization for the transversal of  $H$ . Now we choose  $v$  as an element of the transversal  $T$  and remove  $v$  and its incident edge from  $H$ . The resulting graph  $H - v$  is either a single vertex or a cycle or it has again a pendant vertex. In the first case the vertices selected so far form a transversal. In the second case the union of the set  $T$  of already selected vertices together with the vertex set of the cycle forms a transversal for  $H$ . In the third case we can recursively proceed with the construction of the transversal by adding the next pendant vertex to  $T$ .  $\square$

Theorem 1 shows that  $f_{km}$  equals the number of pseudoforests with  $k$  vertices and  $m$  edges. First we count connected pseudoforests, i.e. *pseudotrees*. A pseudotree is either a tree or a connected graph with exactly one cycle. The number of trees of order  $n$  is,



according to a famous theorem of Cayley [4], given by  $n^{n-2}$ . As a tree of order  $n$  has exactly  $n - 1$  edges, the number of connected unicyclic graphs with a cycle of length 2 is

$$(n - 1)n^{n-2}, n \geq 2. \quad (1)$$

The calculation of the number of connected unicyclic graphs with a cycle of length  $n$ ,  $n \geq 3$ , requires some more effort. First, we select a subset of  $k$  vertices that induce the cycle. Then we take into account that there exist  $(k - 1)!/2$  different cycles on a given set of  $k$  vertices. Combining this results, we learn that there are

$$c_{n,k} = \binom{n}{k} \frac{(k - 1)!}{2} = \frac{n^k}{2k}$$

different cycles of length  $k$ ,  $k \geq 3$ , on  $n$  vertices.

Assume that  $H$  is a connected unicyclic graph of order  $n$  with a cycle  $C_k$ . Then the graph  $H/C_k$  obtained from  $H$  by contracting the cycle  $C_k$  is a tree  $T$ . If we replace  $C_k$  by another cycle  $\hat{C}_k$  with the same vertex set, then  $H/\hat{C}_k$  is still the same tree  $T$ . The set of different trees that we can obtain by contraction of  $C_k$  is exactly the set of spanning trees of the graph  $K_n/C_k$ . We denote the number of spanning trees of a graph  $G$  by  $t(G)$ . The number of connected unicyclic graphs of order  $n$  with a cycle of length  $k$  ( $k \geq 3$ ) is, consequently,

$$\frac{n^k}{2k} t(K_n/C_k). \quad (2)$$

The graph  $K_n/C_k = (X, E)$  has  $n - k + 1$  vertices. The vertex set  $X$  splits into a subset  $Y$  of cardinality  $n - k$  and  $\{z\}$ , where  $z$  is the vertex obtained by the contraction of the cycle and  $Y$  is a subset of the original vertex set of the complete graph  $K_n$ . Each vertex of  $Y$  is connected with  $z$  by  $k$  parallel edges. The matrix  $L_z$  obtained from the Laplace matrix of  $K_n/C_k$  by deleting its column and row that correspond to the vertex  $z$  has the shape

$$L_z = \begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ -1 & n-1 & -1 & \dots & -1 \\ -1 & -1 & n-1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \dots & n-1 \end{bmatrix}_{n-k, n-k}$$

According to the theorem of Kirchhoff, see [8], the number of spanning trees of  $K_n/C_k$  is

$$t(K_n/C_k) = \det L_z.$$

Multiplying the last row of  $L_z$  with  $-1$  and adding the result to all other rows yields

$$\det L_z = \begin{vmatrix} n & 0 & 0 & \dots & -n \\ 0 & n & 0 & \dots & -n \\ 0 & 0 & n & \dots & -n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \dots & n-1 \end{vmatrix}_{n-k, n-k}$$

Adding all columns to the last column gives

$$\det L_z = k n^{n-k-1}. \quad (3)$$

For  $k = 1$ , we rediscover Cayley's formula for the number of spanning trees of a complete graph of order  $n$ .

The combination of Equations (1), (2), and (3) yields the number of connected unicyclic graphs of order  $n$ :

$$c_n = (n-1)n^{n-2} + \sum_{k=3}^n \frac{n^k}{2k} k n^{n-k-1} = (n-1)n^{n-2} + \frac{n^{n-1}}{2} \sum_{k=3}^n \frac{n^k}{n^k}. \quad (4)$$

Trees and connected unicyclic graphs are the building bricks of pseudoforests. Hence, we can use the exponential formula, see [15], in order to count pseudoforests. First, we define the exponential generating functions

$$T(x) = \sum_{n \geq 1} n^{n-2} \frac{x^n}{n!} \quad (5)$$

$$C(x) = \sum_{n \geq 1} \left( (n-1)n^{n-2} + \frac{n^{n-1}}{2} \sum_{k=3}^n \frac{n^k}{n^k} \right) \frac{x^n}{n!} \quad (6)$$

Using the exponential formula, we obtain

$$P(x) = e^{T(x)+C(x)} \quad (7)$$

as the exponential generating function for the number of pseudoforests.

The introduction of a second variable in the pseudoforest counting function (7) allows to keep track of the number of edges forming the forest:

$$P(x, y) = e^{yT(x)+C(x)} \quad (8)$$

Consider a pseudoforest  $H$  of order  $n$ . Each unicyclic component of  $H$  has the same number of vertices and edges; each tree component of  $H$  has one edge less than its order. Consequently, the variable  $y$  in  $P(x, y)$  counts the edge deficiency of a pseudoforest with respect to its order. Thus we get the following result.

**Theorem 2.** *Let  $m, n$  be nonnegative integers. The number of pseudoforests of order  $n$  with exactly  $m$  edges is  $n!$  times the coefficient of  $x^n y^{n-m}$  in the power series expansion of  $P(x, y)$ .*

Now assume that the  $S$ -list  $\mathcal{L}$  is a set, which implies that no pair of elements from  $S$  appears twice in  $\mathcal{L}$ . In this case, the resulting graph  $G = (S, \mathcal{L})$  is simple. Consequently, a pseudoforest cannot contain any two-cycles. We modify the exponential generating function (6) in order to exclude two-cycles:

$$\tilde{C}(x) = \sum_{n \geq 1} \frac{n^{n-1}}{2} \sum_{k=3}^n \frac{n^k}{n^k} \frac{x^n}{n!} \quad (9)$$

If  $f(x, y)$  is a function of  $x$  and  $y$ , then we denote by  $[x^k y^l]f(x, y)$  the coefficient of  $x^k y^l$  in the formal power series expansion of  $f(x, y)$ , presupposing that the series expansion does exist. Using again the exponential formula, we obtain the following statement.

**Theorem 3.** *Let  $m, n$  be nonnegative integers. The number of pseudoforests of order  $n$  with exactly  $m$  edges and without any two-cycles is*

$$n! [x^n y^{n-m}] e^{yT(x)+\tilde{C}(x)}.$$

We define the exponential generating function for the number of connected unicyclic graphs with a two-cycle by

$$C_2(x) = \sum_{n \geq 2} (n-1)n^{n-2} \frac{x^n}{n!}. \quad (10)$$

**Theorem 4.** *Let  $m$  be a positive integer and  $S = \{1, \dots, m\}$ . The probability that a random  $S$ -list  $\mathcal{L} = \{L_1, \dots, L_m\}$  has a transversal is*

$$\frac{(m!)^2}{\binom{m}{2}^m} [x^m] e^{\tilde{C}(x) + \frac{1}{2}C_2(x)}.$$

*Proof.* The number of all possible  $S$ -lists of length  $m$  is

$$\binom{m}{2}^m. \quad (11)$$

The graph  $G = (\{1, \dots, m\}, \mathcal{L})$  is according to Theorem 1 a pseudoforest. However, as the number of vertices and edges of  $G$  are equal, all components of  $G$  have to be unicyclic. The exponential generating function for the number of pseudoforests without tree components is

$$e^{C(x)} = e^{C_2(x) + \tilde{C}(x)}.$$

As we count in Equation (11) *ordered* lists, we have to distinguish a graph with the edges  $e_k = \{a, b\}$  and  $e_l = \{c, d\}$  from a graph with the edges  $e_l = \{a, b\}$  and  $e_k = \{c, d\}$ . Consequently, the order of the edge set matters, which can be easily taken into account by multiplying the number of pseudoforests with  $m!$ . However, each two-cycle is than counted twice as the permutation of its edges does not change anything. This can be corrected by using the exponential generating function

$$e^{\tilde{C}(x) + \frac{1}{2}C_2(x)}$$

instead. Now the theorem results from the classical definition of probability.  $\square$

### 3 The Tutte Polynomial

The Tutte polynomial has been introduced in [12]. Since its introduction it has found numerous applications in different fields of mathematics and theoretical physics. It might be considered as the most important polynomial invariant of graphs (and matroids). The interested reader can find a variety of results about Tutte polynomials in [5, 13, 14].

The *rank* of a graph  $G = (V, E)$  with  $k(G)$  components is defined by  $r(G) = |V| - k(G)$ . The *nullity* of  $G$  is  $n(G) = |E| - |V| + k(G)$ . Rank and nullity of a graph are the dimensions of its cut and cycle space, respectively, see [1]. The Tutte polynomial is the ordinary generating function of rank and nullity of a graph, however, with respect to the shifted variables  $x - 1$  and  $y - 1$ :

**Definition 5.** *Let  $G = (V, E)$  be a graph (parallel edges and loops are permitted), then the Tutte polynomial of  $G$ ,  $T(G; x, y)$ , is defined as*

$$T(G; x, y) = \sum_{F \subseteq E} (x-1)^{k(G[F]) - k(G)} (y-1)^{|F| - |V| + k(G[F])},$$

where  $G[F]$  denotes the spanning subgraph of  $G$  with edge set  $F$ .

We denote the graph obtained by deleting an edge  $e$ ,  $e \in E$  by  $G - e$ , and the graph obtained by contracting an edge  $e$  by  $G/e$ , then the Tutte polynomial can be represented recursively. The following theorem is essentially William Tutte's definition of this polynomial. For a proof of this statement, see [2].

**Theorem 6.** *If  $G = (V, E)$  is a graph and  $e \in E$  is an arbitrary edge of  $G$ , then*

$$T(G; x, y) = \begin{cases} xT(G-e; x, y) & \text{if } e \text{ is a bridge;} \\ yT(G-e; x, y) & \text{if } e \text{ is a loop;} \\ T(G-e; x, y) + T(G/e; x, y) & \text{otherwise.} \end{cases}$$

Furthermore,  $T(E_n; x, y) = 1$  for any empty graph  $E_n$ ,  $n \geq 1$

The edge deletion–contraction formula provides a powerful tool to prove properties of the Tutte polynomial and to show that it is a generalization of many other well-known graph polynomials, for instance of the chromatic, flow, and reliability polynomial.

Assume there is a fixed but arbitrarily defined linear ordering  $\prec$  on the edge set  $E(G) = \{e_1, \dots, e_m\}$ , where  $e_i \prec e_j$  if  $i < j$ . The set of all spanning trees of  $G$  is denoted by  $\mathcal{T}(G)$ . For any spanning tree  $T \in \mathcal{T}(G)$ , an edge  $e \in T$  is called *internally active* in  $T$ , if  $e \prec f$  for all  $f \in E(G) \setminus E(T)$  such that  $T - e + f \in \mathcal{T}(G)$ . An edge  $e \in E(G) \setminus E(T)$  is *externally active* with respect to  $T$ , if  $e \prec f$  for all  $f \in E(T)$  such that  $T - f + e \in \mathcal{T}(G)$ . A spanning tree  $T$  has *internal activity*  $i = \text{int}(T)$  and *external activity*  $j = \text{ext}(T)$  if there are precisely  $i$  internally active edges and precisely  $j$  externally active edges with respect to  $T$ .

**Theorem 7** ([13]). *Let  $G$  be a graph with a linear order on its edge set, then*

$$T(G; x, y) = \sum_{T \in \mathcal{T}(G)} x^{\text{int}(T)} y^{\text{ext}(T)} = \sum_{i,j} t_{i,j} x^i y^j,$$

where  $t_{i,j}$  is the number of spanning trees with internal activity  $i$  and external activity  $j$ .

A proof for this theorem can be found in [2] or [13]. The most interesting statement contained in Theorem 7 is that the Tutte polynomial is independent of the order of the edge set.

We proceed with calculating the number of unicyclic graphs via application of the spanning trees expansion of the Tutte polynomial.

**Theorem 8.** *The number of unicyclic graphs (without parallel edges) of order  $n$  is*

$$\left. \frac{d}{dy} T(K_n; 1, y) \right|_{y=1}.$$

*Proof.* A connected unicyclic graph can be obtained from a tree by insertion of an edge between two nonadjacent vertices. Unicyclic graphs and trees are considered here as spanning subgraphs of a given complete graph. Consequently, we can uniquely identify each tree or connected unicyclic graph by its edge set. If we add an edge between any two nonadjacent vertices of each tree, then each unicyclic graph with a cycle of length  $k$  is repeatedly generated  $k$  times. In order to avoid multiple generation of the same unicyclic graph, the edges of the complete graph  $K_n$  are linearly ordered. An extension of a given spanning tree of  $K_n$  is only performed with edges that are externally active

with respect to the tree. As we want to form *unicyclic* graphs, we have to select exactly one of the externally active edges for the extension. So there is a bijection between the set of connected unicyclic graphs with  $n$  vertices and the set of spanning trees of  $K_n$  where exactly one of its externally active edges is labeled.

We conclude from Theorem 7 that the generating function for the number of spanning trees of  $K_n$  with  $j$  external edges is given by  $T(K_n; 1, y)$ . The labeling (or *pointing* as it is called in [6]) is realized by the operator  $y \frac{d}{dy}$  applied to  $T(K_n; 1, y)$ , where we save here the multiplication with  $y$  since this variable is finally substituted by 1. Hence the number of connected unicyclic graphs with  $n$  vertices is

$$\sum_{T \in \mathcal{T}(K_n)} \text{ext}(T) = \left. \frac{d}{dy} T(K_n; 1, y) \right|_{y=1}. \quad (12)$$

□

The result given in the last theorem can be generalized in order to count  $S$ -Lists  $\mathcal{L}$  with forbidden two-element sets. Assume that  $\mathcal{F} = \{\{i_1, j_1\}, \dots, \{i_r, j_r\}\}$  is a *family of forbidden pairs*, i.e. pairs that must not appear as sets in  $\mathcal{L}$ . We define an undirected graph

$$G = \left( S, \binom{S}{2} \setminus \mathcal{F} \right),$$

where  $\binom{S}{2}$  denotes the set of all two-element subsets of  $S$ .

**Corollary 9.** *The number of unicyclic graphs without parallel edges which are spanning subgraphs of  $G$  is*

$$\left. \frac{d}{dy} T(G; 1, y) \right|_{y=1}.$$

Interestingly, the number of spanning trees can also be given by  $T(G; 1, 1)$ . Corollary 9 is not the solution for the initially stated problem as we might have identical sets in our set family, which means that we can have parallel edges in  $G$ . However, as we count *unicyclic* graphs, at most one edge can have a second parallel edge. Consequently, the number of families of  $m$  two-element subsets from  $S = \{1, \dots, m\}$  that avoid all pairs from  $\mathcal{F}$  is

$$\left. \frac{d}{dy} T(G; 1, y) \right|_{y=1} + (m-1)T(G; 1, 1).$$

## References

- [1] N. Biggs. *Algebraic Graph Theory*. 2nd ed. Cambridge University Press, 1993.
- [2] B. Bollobás. *Modern Graph Theory*. Springer, 1998.
- [3] P. J. Cameron. *Combinatorics: Topics, Techniques, Algorithms*. Cambridge University Press, 1994.
- [4] A. Cayley. ‘A theorem on trees’. In: *Quart. J. Math* 23.376-378 (1889), p. 69.
- [5] J. A. Ellis-Monaghan and C. Merino. ‘Graph polynomials and their applications I: The Tutte polynomial’. In: *Structural Analysis of Complex Networks*. Springer, 2011, pp. 219–255.

- 
- [6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
  - [7] P. Hall. ‘On representatives of subsets’. In: *J. London Math. Soc* 10.1 (1935), pp. 26–30.
  - [8] G. R. Kirchhoff. ‘Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird’. In: *Annalen der Physik und Chemie*. Band LXXII, 12 (1847).
  - [9] J. H. van Lint and R. M. Wilson. *A course in combinatorics*. 2nd ed. Cambridge University Press, 2001.
  - [10] V. N. Sachkov. ‘Systems of distinct representatives for random sets’. In: *Mathematics of the USSR-Sbornik* 97 (3 1975), pp. 395–402.
  - [11] V. N. Sachkov and V. E. Tarakanov. *Combinatorics of Nonnegative Matrices*. Trans. from the Russian by V. F. Kochin. The American Mathematical Society, 2002.
  - [12] W. T. Tutte. ‘A ring in graph Theory’. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 43 (1 1947), pp. 26–40.
  - [13] W. T. Tutte. *Graph Theory*. Cambridge University Press, 2001.
  - [14] D. J. Welsh and C. Merino. ‘The Potts model and the Tutte polynomial’. In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1127–1152.
  - [15] H. S. Wilf. *generatingfunctionology*. Academic Press, Inc., 1994.

# Two-Edge Connected Reliability and Reductions

MANJA REINWARDT<sup>1</sup>

Reliability measures are a well explored topic in the literature. Given a probabilistic graph with failing edges, we investigate the less-known two-edge connected reliability which is the probability that a two-edge connected graph is realized. The problem is NP-hard and we present general algorithms with exponential complexity. The goal is to reduce the graph so that the computations are faster. The main part of the paper deals with various reductions that fulfil this goal.

## 1 Introduction

The analysis of reliability has a long history in engineering and mathematics. Moore and Shannon wrote the first paper about reliability as it is discussed here in 1956. They studied the reliability of electrical relays and defined the corresponding function [3]. Further measures that are investigated thoroughly in the literature are the  $s$ - $t$ -terminal reliability  $R_{st}(G)$  and the all-terminal reliability  $R(G)$  (see for instance the textbooks [1] and [8]).

We deal with the notion of a probabilistic graph which is a graph  $G = (V, E)$  with a vertex set  $V$  and an edge set  $E$  and each edge  $e \in E$  fails stochastically independently with known probabilities  $q_e = 1 - p_e$ . The two-edge connected reliability was defined by Lucet et al. in [2] and it is the probability that the graph is two-edge connected, i.e. that there are at least two edge-disjoint paths between every two vertices of the graph. The problem of finding the two-edge connected reliability is NP-hard since the number of Hamilton cycles is included as a coefficient of the polynomial and the decision problem of finding a Hamilton cycle is NP-complete.

The paper is structured as follows. First, we give some definitions. Then two algorithms with exponential runtime in the size of the graph are presented. The goal is to reduce the order and size of the graph to speed up the computation. The next section deals with the reductions. In the end, the conclusion gives hints for further reading.

## 2 Definitions

**Definition 1** (Probabilistic Graph). *Let  $G = (V, E)$  be an undirected graph. It is called a probabilistic graph if each edge  $e \in E$  fails stochastically independently with probability  $1 - p_e$ . A state of the graph is characterized by the set of operating edges  $F \subseteq E$ .*

---

<sup>1</sup>The reasearch for this thesis was supported by the grant 100087634 of the European Social Fund.



Let  $G = (V, E)$  be a connected undirected graph. A *bridge* in  $G$  is an edge whose removal from  $G$  renders  $G$  disconnected. The graph  $G$  is called *two-edge connected* if any two vertices of  $V$  are connected by at least two edge-disjoint paths of  $G$ . Consequently,  $G$  is two-edge connected if and only if  $G$  is connected and bridgeless. We denote by  $\lambda_G$  the edge connectivity of  $G$  and by  $\lambda_G(u, v)$  the local edge connectivity between two distinct vertices  $u$  and  $v$  in  $G$ .

**Definition 2** (Two-edge connected reliability). *The two-edge connected reliability of a probabilistic graph  $G = (V, E)$  is the probability that the graph is two-edge connected. It can be computed by listing every possible state  $F$  for which the subgraph  $(V, F)$  is two-edge connected ( $\lambda_{(V, F)} \geq 2$ ) and summing up the corresponding probabilities.*

$$R_{2\text{-ec}}(G) = \sum_{\lambda_{(V, F)} \geq 2} \prod_{e \in F} p_e \prod_{f \in E \setminus F} (1 - p_f)$$

**Definition 3** (Pathsets and essential edges). *A pathset is a subset of edges  $F \subseteq E$  for which the graph  $(V, F)$  is two-edge connected. The pathset  $F$  is called minimal if no proper subset of  $F$  is a pathset. An edge  $e$  is essential if it is an element of every pathset of  $G$ .*

**Remark 4.** *All the edges of a graph  $G = (V, E)$ , which are an element of an edge cut set of cardinality two, are essential. In particular every edge incident to a vertex of degree two is essential.*

## 3 General Algorithms for Two-Edge Connected Reliability

### 3.1 Complete Enumeration

We can compute the two-edge connected reliability using the definition by enumerating all spanning subgraphs and testing each one for two-edge connectivity. Of course, this is impractical for larger graphs since there are  $2^{|E|}$  many of them.

---

#### Algorithm 1 Complete Enumeration

---

**Input:** Graph  $G = (V, E)$  with probability vector  $\mathbf{p}$ .

**Output:** Two-edge connected reliability  $R_{2\text{-ec}}(G)$ .

$R_{2\text{-ec}}(G) := 0$

**if**  $\lambda_G < 2$  **then**

**return** 0

**end if**

**for all**  $F \subseteq E$  with  $|F| \geq |V|$  **do**

**if**  $\lambda_{(V, F)} \geq 2$  **then**

$R_{2\text{-ec}}(G) := R_{2\text{-ec}}(G) + \prod_{e \in F} p_e \prod_{f \in E \setminus F} (1 - p_f)$

**end if**

**end for**

**return**  $R_{2\text{-ec}}(G)$

---



### 3.2 Decomposition

This method first occurred in [4] under the name of factoring theorem and was derived for redundancy networks. The algorithm has an exponential runtime complexity in general since two graphs are generated in each recursion step. Thus, there are  $2^{|E|}$  many graphs that have to be explored in the worst case. The algorithm (see Algorithm 2) follows from the law of total probability and the fact that all edges fail stochastically independently.

**Theorem 5.** *Let  $G = (V, E)$  be a graph,  $e \in E$  an edge of the graph and let  $G|_{p_e=1}$  be the graph  $G$  with  $p_e$  set to one. The following decomposition equation holds for the two-edge connected reliability of  $G$ .*

$$R_{2-ec}(G) = p_e \cdot R_{2-ec}(G|_{p_e=1}) + (1 - p_e) \cdot R_{2-ec}(G - e). \quad (1)$$

---

#### Algorithm 2 Decomposition as a recursive algorithm

---

**Input:** Graph  $G = (V, E)$  with probability vector  $\mathbf{p}$ .

**Output:**  $R_{2-ec}(G)$

**if**  $\lambda_G < 2$  **then**

**return** 0

**end if**

**if**  $p_e = 1 \forall e \in E$  **then**

**return** 1

**end if**

$e :=$  arbitrary edge of  $G$  with  $p_e \neq 1$

**return**  $p_e \cdot R_{2-ec}(G|_{p_e=1}) + (1 - p_e) \cdot R_{2-ec}(G - e)$

---

## 4 Reductions

Reductions for the all-terminal reliability, namely series-parallel reductions, can be found in [6]. Similar methods can be applied for the two-edge connected reliability. Additionally, there is the delta-wye reduction for the all-terminal reliability (see for instance [7]), which is unfortunately not applicable for higher connectivity and thus for  $R_{2-ec}(G)$ .

### 4.1 Degree-Two Reduction

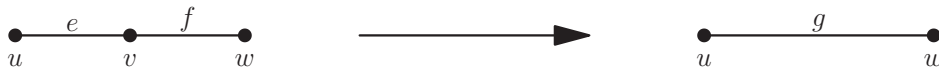


Figure 1: Degree-two reduction.

Every edge adjacent to a degree-two vertex may not be deleted, it is essential. Let  $v$  be a vertex of degree two in a graph  $G = (V, E)$  and let  $u$  and  $w$  be the neighbouring vertices of  $v$ . Since the information that  $u$  and  $w$  are connected by a path must be

conserved, the vertex  $v$  and its incident edges  $e$  and  $f$  can be replaced by a single edge. Because all edges fail stochastically independently, the two-edge connected reliability of  $G$  arises from the product of the probabilities  $p_e$ ,  $p_f$  and the two-edge connected reliability of the reduced graph  $G'|_{p_g=1}$ . This reduced graph has a new edge  $g = \{u, w\}$  with probability one.

$$R_{2\text{-ec}}(G) = p_e \cdot p_f \cdot R_{2\text{-ec}}(G'|_{p_g=1}). \quad (2)$$

## 4.2 Reduction at a Separating Vertex Set of Cardinality Two

The following result is based on ideas from the PhD thesis of Tittmann [9] where a similar reduction is given for the all-terminal reliability.

**Theorem 6.** *Given a graph  $G = (V, E)$  with a separating vertex set  $\{u, v\}$  and two subgraphs  $H$  and  $K$  such that  $K \cap H = (\{u, v\}, \emptyset)$  and  $K \cup H = G$ , the graph  $H$  can be reduced to a pair of parallel edges  $e$  and  $f$  with the operating the probabilities  $p_e$  and  $p_f$ , respectively, in the following way. The reduced graph is  $G'$  (see Figure 2).*

$$R_{2\text{-ec}}(G) = \Omega \cdot R_{2\text{-ec}}(G') \quad (3)$$

with

$$\Omega = A + B + C, \quad (4)$$

$$p_e = \frac{2C}{2C + B + \sqrt{B^2 - 4AC}}, \quad (5)$$

$$p_f = \frac{2C}{2C + B - \sqrt{B^2 - 4AC}}, \quad (6)$$

where  $A = \Pr(H_{u|v})$  is the probability that there is no  $u$ - $v$  path in  $H$ , but all vertices of  $H$  are either in a two-edge connected component containing  $u$  or in a two-edge connected component containing  $v$ ,  $B = \Pr(H_{u-v})$  denotes the probability that the vertices  $u$  and  $v$  lie in different two-edge connected components in  $H$ , but there is exactly one path connecting them. The term  $C = \Pr(H_{uv})$  denotes the probability that  $H$  is two-edge connected.

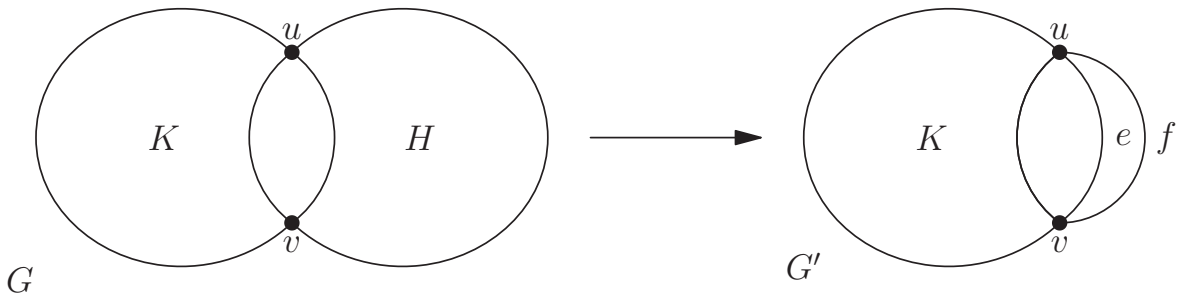


Figure 2: Reduction at a separating vertex set of cardinality two with the original graph  $G$  and the reduced graph  $G'$ .

*Proof.* A case distinction is made for the original graph  $G$  and the reduced graph  $G'$  (see Figure 3) where the local edge connectivity between the vertices  $u$  and  $v$  in  $H$  is considered under the condition that the whole graph is two-edge connected. We

introduce a reduction parameter  $\Omega$  to get an equation system with a unique solution. Case 1 ( $\lambda_H(u, v) = 0$ ) means that there is no  $u$ - $v$  path in  $H$ . In the reduced graph  $G'$ , both edges  $e$  and  $f$  have failed with probability  $(1 - p_e)(1 - p_f)$ . This yields the equation  $\Pr(H_{u|v}) = \Omega(1 - p_e)(1 - p_f)$ . Case 2 ( $\lambda_H(u, v) = 1$ ) stands for exactly one  $u$ - $v$  path in  $H$ . In the reduced graph exactly one of the edges  $e$  and  $f$  has failed and the other is operating, which yields  $\Pr(H_{u-v}) = \Omega(p_e(1 - p_f) + p_f(1 - p_e))$ . In case 3 ( $\lambda_H(u, v) \geq 2$ ), the subgraph  $H$  is two-edge connected. In the reduced graph  $G'$  both edges  $e$  and  $f$  are operating. This leads to  $\Pr(H_{uv}) = \Omega \cdot p_e \cdot p_f$ . We get the following equation system. The non-negativity of the term  $B^2 - 4AC$  cannot be guaranteed, so that  $p_e, p_f$  might not lie in the interval  $[0, 1]$ .

$$A = \Omega(1 - p_e)(1 - p_f) \tag{7}$$

$$B = \Omega(p_e(1 - p_f) + p_f(1 - p_e)) \tag{8}$$

$$C = \Omega \cdot p_e \cdot p_f \tag{9}$$

The solution of this equation system yields the theorem. □

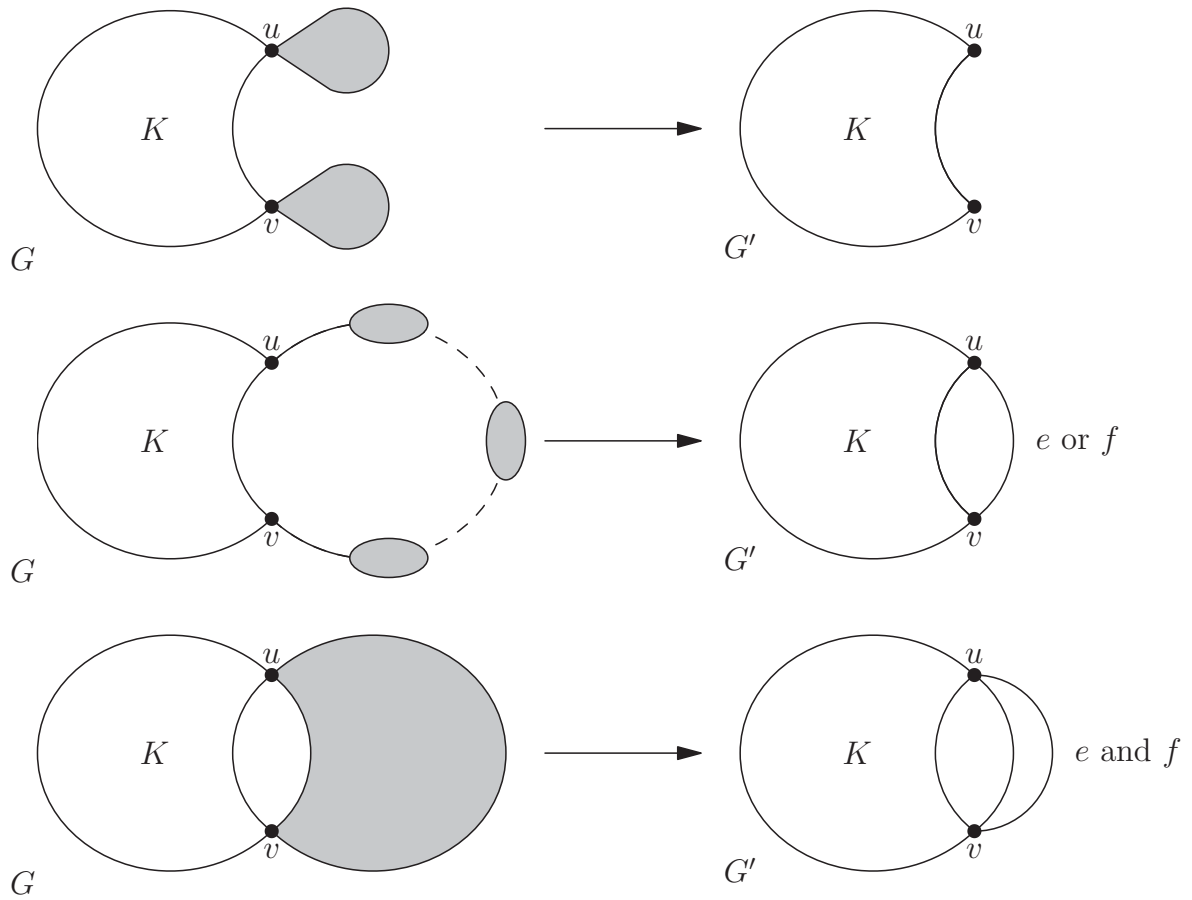


Figure 3: Reduction at a separating vertex set of cardinality two: case 1 (top), case 2 (middle), case 3 (bottom).

### 4.3 Reduction of Multiple Edges

#### Reduction of Multiple Parallel Edges

A bundle of multiple parallel edges  $e_1, \dots, e_k$  with common end vertices can be reduced to a parallel pair of edges  $f$  and  $g$ . The following theorem shows how to compute  $p_f$  and  $p_g$ , which cannot be interpreted as probabilities.

**Theorem 7.** *Let  $G = (V, E)$  be a graph with multiple parallel edges  $e_1, \dots, e_k$  whose end vertices are  $u$  and  $v$ . The probability that all parallel edges fail is*

$$E_k = \prod_{i=1}^k (1 - p_{e_i})$$

and the probability that all but one parallel edges fail is

$$E_{k-1} = \sum_{i=1}^k p_{e_i} \prod_{\substack{j=1 \\ j \neq i}}^k (1 - p_{e_j}).$$

Then we can reduce the multiple edges to a pair of edges  $f$  and  $g$  and compute  $p_f$  as well as  $p_g$  in the following way.

$$p_{f/g} = 1 - E_k - \frac{1}{2}E_{k-1} \pm \sqrt{\frac{1}{4}E_{k-1}^2 - E_k(1 - E_k - E_{k-1})}$$

*Proof.* The probabilities  $E_k$  and  $E_{k-1}$  can be computed as stated in the theorem because all edges are failing stochastically independently. We use the formula from Theorem 6 where the subgraph  $H$  has the vertex set  $\{u, v\}$  and its edge set contains the parallel edges  $e_1, \dots, e_k$ . The reduction factor is  $\Omega = 1$  because all cases are covered and only two equations are needed. In the general formula, we get:

$$\begin{aligned} A &= E_k \\ B &= E_{k-1} \\ C &= 1 - E_k - E_{k-1} \end{aligned}$$

Substituting this in Equation (5) and (6) and simplifying with the property  $A+B+C = 1$  yields the theorem.  $\square$

#### Reduction of Parallel Edge Pairs in a Row

We can reduce multiple pairs of parallel edges if their pairwise common vertices have all degree four, i.e. there are no further incident edges. Let all edges have the same operating probabilities  $p$ . If all edges have different operating probabilities, a formula can be given, too. However, it is long and complicated and can easily be derived with the methods presented here.

**Theorem 8.** *Let  $G = (V, E)$  be a graph and  $e_1, e_2$  a pair of parallel edges with the end vertices  $u$  and  $v$  in  $G$ . Let  $f_1, f_2$  be a pair of parallel edges with the end vertices  $v$  and  $w$ . The common end vertex  $v$  has no other incident edges.*

*Let  $G'$  be the reduced graph where the vertex  $v$  is deleted and the edges  $e_1, e_2$  as well as  $f_1, f_2$  are replaced by edges  $g_1, g_2$  with end vertices  $u$  and  $w$ . For all  $e \in E(G)$ ,  $p_e = p$*

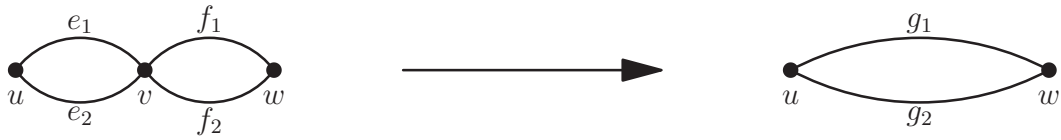


Figure 4: Reduction of parallel edges in a row.

is the probability for an operating edge  $e$ . In the reduced graph, the parallel edge pair  $g_1$  and  $g_2$  has the probability of operating  $p_1$  and  $p_2$ , respectively, and these probabilities can be calculated by the following equation.

$$p_{1/2} = \frac{p^2}{(1-p)^2 + 1 \pm (1-p)\sqrt{2(2-p^2)}}. \quad (10)$$

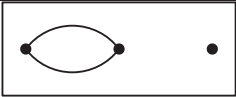
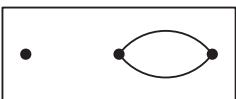

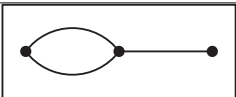
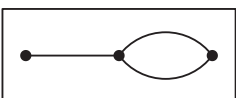
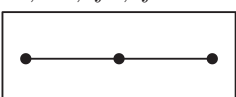

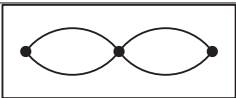

Cases	$G$	$G'$
Case 1: $u w$	 $e_1, e_2, \bar{f}_1, \bar{f}_2$  $\bar{e}_1, \bar{e}_2, f_1, f_2$	 $\bar{g}_1, \bar{g}_2$
Case 2: $u - w$	 $e_1, e_2, f_1, \bar{f}_2$ or $e_1, e_2, \bar{f}_1, f_2$  $e_1, \bar{e}_2, f_1, f_2$ or $\bar{e}_1, e_2, f_1, f_2$  $e_1, \bar{e}_2, f_1, \bar{f}_2$ or $\bar{e}_1, e_2, f_1, \bar{f}_2$ or $e_1, \bar{e}_2, \bar{f}_1, f_2$ or $\bar{e}_1, e_2, \bar{f}_1, f_2$	 $g_1, \bar{g}_2$ or $\bar{g}_1, g_2$
Case 3: $u = w$	 $e_1, e_2, f_1, f_2$	 $g_1, g_2$

Table 1: Cases for the reduction of parallel edge pairs in a row where  $e$  is an operating edge and  $\bar{e}$  a failing edge.

*Proof.* A case distinction is made (Table 1) according to the connectivity between  $u$  and  $w$  and we consider the subgraph induced by  $u$ ,  $v$  and  $w$  in the original graph  $G$  as well as in the reduced graph  $G'$ . In the case 1 ( $u|w$ ), the vertices  $u$  and  $w$  are disconnected, they are connected by one path in the case 2 ( $u - v$ ) and by two edge-disjoint paths in the case 3 ( $u = w$ ).

The vertices  $u$  and  $w$  constitute a separating vertex set of cardinality two. We can replace the probabilities  $A = 2p^2(1 - p)^2$  (case 1 for  $G$ ),  $B = 4p^2(1 - p)$  (case 2 for  $G$ ) and  $C = p^4$  (case 3 for  $G$ ) of Theorem 6 and get the following equations together with  $\Omega = A + B + C = 1$ .

$$\begin{aligned} 2p^2(1 - p)^2 &= (1 - p_1)(1 - p_2) \\ 4p^2(1 - p) &= p_1(1 - p_2) + p_2(1 - p_1) \\ p^4 &= p_1p_2 \end{aligned}$$

The solution of the equation system yields the theorem. □

## 5 Conclusion

In this paper, results for the two-edge connected reliability of graphs with imperfect edges have been presented. Their computation is an NP-hard problem. Two algorithms with exponential runtime have been presented. We have extended some known results of other reliability measures and investigated other new reductions. We have presented the degree-two reduction as well as the reduction at a separating vertex set of cardinality two. For multigraphs, multiple edges can be reduced in different ways. The theorems of this paper are also included in the doctoral thesis [5] where further results for the two-edge connected reliability can be found.

## References

- [1] C. Colbourn. *The combinatorics of network reliability*. Oxford University Press, USA, 1987.
- [2] C. Lucet, J.-F. Manouvrier, and J. Carlier. ‘Evaluating Network Reliability and 2-Edge-Connected Reliability in Linear Time for Bounded Pathwidth Graphs’. In: *Algorithmica* 27 (3 2000), pp. 316–336. ISSN: 0178-4617.
- [3] E. Moore and C. Shannon. ‘Reliable circuits using less reliable relays’. In: *Journal of the Franklin Institute* 262.3 (1956), pp. 191–208.
- [4] F. Moskowitz. ‘The analysis of redundancy networks’. In: *American Institute of Electrical Engineers, Part I: Communication and Electronics, Transactions of the* 77.5 (1958), pp. 627–632.
- [5] M. Reinwardt. ‘Combinatorial and Graph Theoretical Aspects of Two-Edge Connected Reliability’. PhD thesis (submitted). Technische Universität Bergakademie Freiberg, 2015.
- [6] A. Rosenthal. ‘Series-parallel reduction for difficult measures of network reliability’. In: *Networks* 11.4 (1981), pp. 323–334.
- [7] A. Rosenthal and D. Frisque. ‘Transformations for simplifying network reliability calculations’. In: *Networks* 7.2 (1977), pp. 97–111.

- 
- [8] D. Shier. *Network reliability and algebraic structures*. Oxford University Press, USA, 1991.
  - [9] P. Tittmann. ‘Dekompositions- und Reduktionsmethoden zur Zuverlässigkeitsanalyse stochastischer Netzstrukturen’. PhD thesis. Ingenieurhochschule Mittweida, 1990.





# Sharing Data in Vehicular Ad-hoc Networks

CHRISTIAN BAUSCH<sup>1</sup>, STEFAN KAHL<sup>1</sup> AND SARA KISCHNICK<sup>1</sup>

## 1 Abstract

Vehicular connectivity has greatly evolved in the past years, but vehicles mostly do not share any information with other parties, except for immediate accident avoidance, which basically is still future work. Mobile communication is well established and available in today's automobiles. However, sharing information based on social aspects between vehicles is still in its infancy. Electric automobiles, especially lightweight ones, could benefit from data collected and shared by other vehicles even more than conventional, fuel-based automobiles. We propose a concept of socializing vehicles which collect and share information according to the social media blueprint, based on sensor data acquired on-board. We introduce a conceptual workflow of data collection and distribution and present a simulation supporting our concept.

## 2 Introduction

Data sharing and vehicle-to-vehicle communication has become a well-established field of research in recent years. Nonetheless, the communication between multiple vehicles is often associated with short range connections and short term data lifespans with no global exchange. Although mobile connectivity has greatly evolved in the past years, vehicles mostly do not share any information with other parties, except for immediate accident avoidance (for more examples see [7]).

Following that scheme, if a dangerous situation occurs automobiles avoid accidents, inform nearby vehicles and after that, drive on. But, the same source of danger could be relevant to all vehicles passing it in the near future. Persistent information which is globally shared and stored and is available to every vehicle on the road can help to avoid multiple incidents under similar conditions.

Conventional, fuel-based vehicles benefit not only from danger warnings but also from real-time traffic information and adaptive routing through unpleasant environments (e.g. bad road condition, see [8]). Information shared by others hold even more value for vehicles with electric motors, especially lightweight ones. This data, collected and shared by other electric automobiles, could improve range estimation or even extend the range due to fuel efficient routing based on altitude profiles, throttle/break-ratios or even head wind predictions of a route.

---

<sup>1</sup>The research is supported by SMWA/BMVI.

Certain constraints apply in vehicular ad-hoc networks, such as strategy and speed of peer discovery, initializing and maintaining connections, data package size and truncation as well as data distribution using intermediates [11]. Modern communication technology is widespread and highly available, even mobile communication is well established and available in today's vehicles. Communication infrastructures cover cities and countryside, "smart everything" is a common term used to describe concepts of semantic data retrieval and usage. However, sharing information based on social aspects between vehicles is still in its infancy.

Social media can be accessed through almost every modern car, but those built-in services are for human interaction only and are used to share information generated by humans. We propose a concept of socializing vehicles which collect and share information according to the social media blueprint, based on sensor data acquired on-board. Utilizing information socially includes the collection of data, its processing, the extraction of high-level information and sharing it with others, considering whether this information is relevant and which target group it addresses [10].

### 3 Conceptual Workflow

Our proposed workflow consists of four reiterated main components, namely routing, driving, detection and sharing (Figure 1) preceded by the processing of street map material. To prove our concept we implemented a simulation based on OpenStreetMap<sup>2</sup> data and custom web services. It visualizes vehicular networks, the flow of data packages and target specific information relevance evaluation and distribution (see chapter 8).

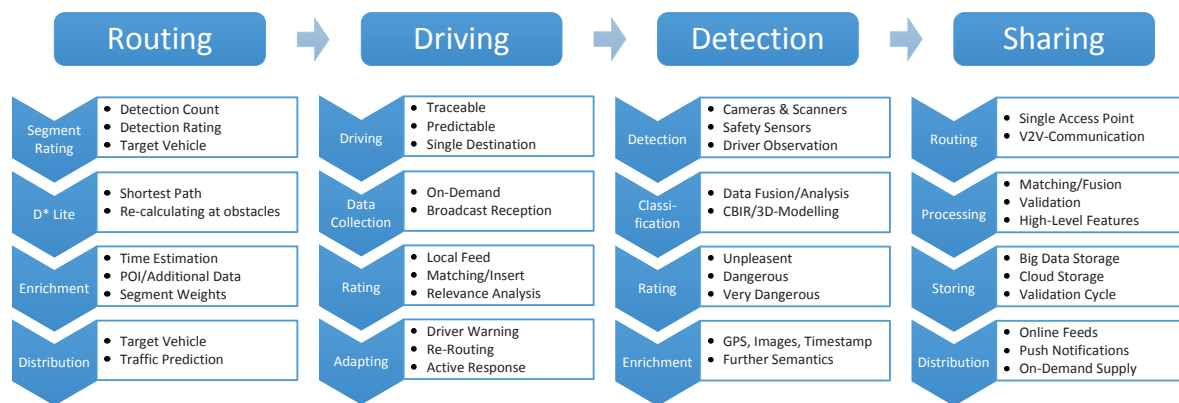


Figure 1: Our proposed workflow consists of four main components. Every step is divided into subcategories and contains concepts for all major parts of social vehicle-to-vehicle communication. Generating routes, driving according to collected high-level information, detecting new environmental or situational conditions and sharing them with others will benefit road safety and driver comfort without any human interaction.

<sup>2</sup><http://www.openstreetmap.de>

## 4 Routing

Finding the best route towards a given destination highly depends on suitable heuristics. In our approach weights of potential candidate segments of road data emerge from shared information. The number of detections and, more importantly, the rating of those reports influence the segment rating and are thereby increasing or decreasing segment weights (see chapter 8). The final route not only consists of plain GPS data, but also contains additional information. This could include time and traffic estimations, as well as points of interest (POI). Other additional data, like high-level semantic information on road conditions, energy efficient driving recommendations or obstacle warnings, could also be incorporated. The distribution of a route depends on vehicle classes and types and, if also submitted, can be used for traffic predictions.

The routing algorithm for the cars bases on the D\* Lite algorithm introduced by Koenig and Likhachev in 2002 [4, 5]. This algorithm works as the A\* algorithm [3] but it has the advantage that there is not necessarily a complete re-routing. If an obstacle occurs, the D\* Lite uses the information from the first routing to re-calculated only the steps which are directly affected by the obstacle. Hence, it is faster than the A\* algorithm if re-routing is required (see also chapter 8.2).

## 5 Driving

When we assume that every drive serves a purpose and has at least one clearly defined destination, we can determine the remaining route ahead of the current vehicle position. Although routes with the same destination may vary depending on certain constraints, every route is traceable and often also predictable, even if no navigation device is used to assist the driver. We therefore assume a controlled environment in which every automobile receives information considering the route ahead, which has been collected (see chapter 6) and shared (see chapter 7) by other vehicles.

In our concept, every automobile holds a list of messages/descriptions of obstacles, dangers or other route-related semantic information. This list is updated constantly with every reception of new data packages. Received data is rated by relevance depending on significance for the vehicle class, safety or driver bias and its correlation with the route ahead. The rating of importance invokes different actions, which could be the request of a new route, driver warnings or an active reaction resulting in braking, suspension regulation, safety system activation or others (Figure 2).

## 6 Detection

The quality of information shared in our social network heavily depends on the quality of the collected sensor data. Moreover, simply recording streams of sensor data does not suffice. It is the smart fusion of data from various on-board sensors that leads to useful information. Those sensors can be related to safety, the engine or passenger comfort, but not every data from every sensor is valuable for other vehicles. Sensors that hold a certain value for other automobiles are cameras and scanners, almost all safety related sensors, battery sensors of electric vehicles and other sensors used for driver observation.

Upgrading low-level data sets to high-level semantic information requires e.g. content-based image retrieval on camera images, 3D-Modelling of objects from scanner point

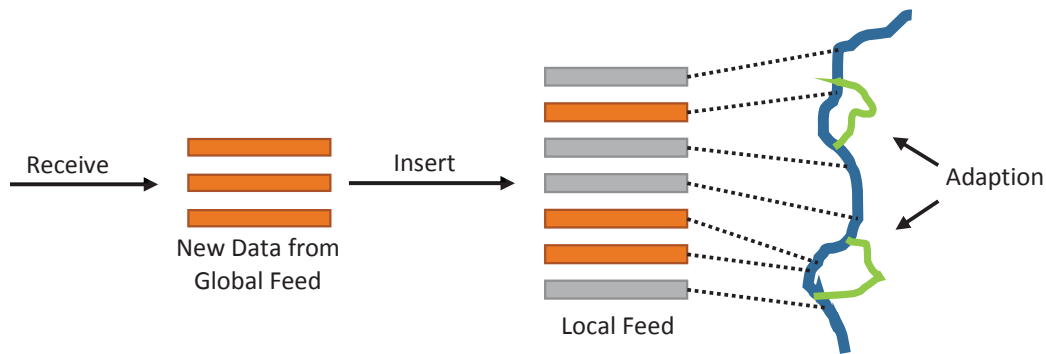


Figure 2: Depending on the vehicles assigned target group, route related data is received and has to be rated and inserted into the local feed of reports. This information is linked to specific route waypoints and an invoked action could be the request of an alternative route (green).

clouds, drawing conclusions from safety sensors (e.g. the activation of the anti-lock braking system could indicate icy road), the rating of roads by fuel efficiency or even the determination of the driver stress level, and by that assuming an unpleasant environment.

After the analysis, the sensor data can be rated depending on its classification. Safety warnings for example can be categorized describing the source of danger as unpleasant, dangerous or very dangerous. The fuel efficiency of roads (or entire cities) could be rated according to the European Union energy label<sup>3</sup>. Other ratings could be based on shape and size of obstacles, traffic density, temporary weather or environmental conditions.

The enrichment of extracted information from sensor data is necessary in order to link the detected features to routes and maps. Therefore the high-level information in our scenario is complemented with the GPS position, timestamps, images or other semantic data, like strategies to resolve or avoid dangerous situations. Those workarounds could evolve from strategies pursued by the vehicle which encountered the dangerous situation in the first place. The enrichment also includes a target group recommendation and finally a submission priority.

## 7 Sharing

There are mainly two strategies for information submission and reception while driving. The first one is a centralistic approach based on a single, central access point (e.g. a webserver), or an on-demand or broadcast distribution from a single, central source (e.g. radio). The more sophisticated strategy however, is using vehicle-to-vehicle communication technologies. It involves highly distributed communication grids due to short range connections to nearby vehicles. This strategy is range dependent and requires fast connection establishment and small data packages. To avoid data overhead and costly connection handshakes, data packages would best be transferred in repeating broadcast cycles.

One simple way of broadcasting is the continuous blind distribution with no specific targets. A better solution is the submission to all discovered peers on a list of nearby receivers (e.g. vehicles in an ad-hoc network), as long as the connection persists [2]. Every data package takes hops using intermediates (e.g. every passing vehicle) and stays

<sup>3</sup><http://eur-lex.europa.eu/legal-content/EN/NOT/?uri=CELEX:32010L0030>

alive until the next access point is reached (e.g. road side units). Its priority raises with the time passing to ensure a submission in a reasonable period of time (Figure 3).

In our scenario, the final receiver is a distributed web service, just like the well-known online social networks, with multiple computing facilities to handle large data occurrences. The three main tasks of this service, aside the reception of data packages, is the processing and analysis of the contained information, the storage of low- and high-level features extracted from the source data and finally the (re-) validated distribution to subscribers (e.g. currently driven vehicles).

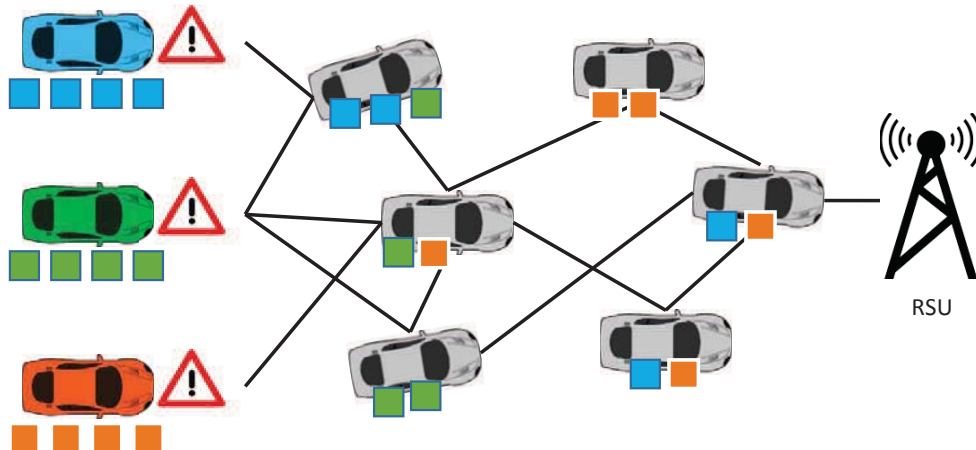


Figure 3: Vehicles which have detected dangerous situations are broadcasting data packages containing relevant information about the source of danger to nearby vehicles, which on their turn repeat the submission of single data packages until the complete message reaches the next access point, in this figure a road side unit.

The processing of incoming data packages includes the fusion with existing information in order to match positions and semantic high-level features, as well as linking multiple reports to one single source. Furthermore, every submitted report has to be validated considering the plausibility of the attached position, high-level description and low-level data. The report is more plausible and more likely to be significant if it is submitted by multiple vehicles in a short period of time. The extraction of semantic high-level features from every data package is part of the classification process. Aside the detection of objects or situations, the classification mainly serves the social aspect of our web service by determining the target group of the contained information.

Providers of social networks for humans use target groups to decide which information is presented to whom. Usually, this is done by analyzing the social background of the subscribers. Relationships between friends or family play a major role for data distribution. Interests, hobbies or news subscriptions also impact the presented contents. Vehicles themselves, as machines, do not have a social background. However, linking certain vehicles to groups of interest, and by that subscribing them to specific, relevant feeds, can be done by vehicle types, driver biases, routes or general profiles and other subscribed feeds.

The most discriminative “social group” is formed by vehicles of the same class, like trucks, motorbikes, mid-size cars, electric automobiles or others. Additional target groups based on driver biases include, among others, vehicles whose drivers like to go fast, vehicles with kids on board or those used for transportation or work related tasks.

The current location, or route ahead, can be used to establish temporary target groups with a common interest in fast and save trips towards individual destinations.

Big data processing, and ultimately storage of huge junks of data, requires mass storage devices, a high availability and highly distributed access capabilities. The so-called cloud storage technologies are evolved and reliable. Suitable infrastructure is affordable and, in order to avoid technical maintenance, often rented from cloud storage specialist. Scalability is very important, as the amount of data stored in our scenario varies from day to day and depends on the specific relevance and timeout of the stored information. Additional processing capacity is needed, as every entry has to be revalidated regularly in order to determine whether an information is still relevant, has exceeded its timeout or has been confirmed/negated by other submissions. The usefulness of an information can also be determined by the rating received from subscribed vehicles, as it is done by the liking or disliking functionality of established social networks.

The way of distribution depends on the intentions and needs of the subscribers. We distinguish three main strategies, all of them present in today's social media. Retrieving information from our proposed social network can be done in form of live feeds, push messages or on-demand data supply either of single packages or the whole collection. Filtering information for specific target groups is vital to exclude irrelevant data and the reduction of submissions. Nonetheless, filtering has to be reliable and accurate in order to preserve important warnings.

## 8 Simulation and Analysis

A lot of effort has been put into the development of traffic simulations. Numerous software tools and frameworks exist that help to emulate realistic traffic situations and, even more important, the simulation of vehicular networks [9]. Those tools have to incorporate a lot of variables depending on the depth of detail that should be provided. Our focus, on the other hand, is on the visualization of the information sharing process and therefore requires only a minimalistic approach of traffic simulation and realistic vehicle behavior.

We created a simulation that implements every aspect from our workflow described above, but emphasizes the process of route generation and information sharing. We did not incorporate a technical layer describing various strategies of connection establishment or range limiting variables, such as buildings. The simulation's UI (see Figure 4) supports the manipulation of the most important parameters, such as number of vehicles, communication range, subscriptions of online feeds, vehicle target groups and obstacle detection. It visualizes vehicular networks and lists shared information, both as a web view, which supports access from mobile devices, and a desktop view for selected vehicles.

### 8.1 Preprocessing Map Data

Despite the non-observance of the technical layer, the simulation should be close to reality. That's why the road map which forms the basis is the street-map of a city. The data for such a map exists as osm-files<sup>2</sup>.

The osm-data is structured in "nodes" and in "ways" whereat the ways are the streets and the nodes are both crossroads and points on a street. With different points on a





Figure 4: Our simulation visualizes vehicular ad-hoc networks in realistic urban environments using OpenStreetMap material. Information about obstacles is shared either globally or locally with every new encounter. Each data package is traceable and our visualization provides an overview of the data currently available to selected vehicles (gray panel on the right).

street it is possible to represent a curve progression by connecting the points with a way. Hence, one way is a sequence of different points. In general, the end points of a way are the crossroads. But there are much more nodes than only crossroads and street points. Most of the maps provide places of interest like restaurants, shops, offices, hydrants, playgrounds and so forth. These nodes must be deleted from the osm-data. Some ways are not suitable for cars, for example side walks, cycle ways and raceways.

If only nodes and ways remain which are important for the network of roads, it is possible to provide the ways and nodes with weights. The weights for the ways are calculated from the length of the separate fractions  $l_{\text{frac}}$  and the maximum speed of the corresponding road  $v_{\text{max}}$ . This results in the formula 2 for calculating the time to pass through the fraction  $t_{\text{frac}}$ . The values  $(lat1, lon1)$  and  $(lat2, lon2)$  are the geographic coordinates for two points of a street and the lengths in the direction of latitudes and longitudes must be multiplied by  $111.3 \text{ km}$  respectively  $111.3 \cdot \cos(lat) \text{ km}$  because these are the distances between two latitudes respectively longitudes. For calculating the time to cross through the full length of the street the separate fractions of time must be added (see equation 3) whereat  $T$  is the number of fractions of the street which are result from the different points.

The remaining nodes and ways are saved in a weighted and directed graph whereat the nodes are the vertices and the complete ways are the weighted edges. If there is a street with two directions, then the graph gets a pair of edges which are directed opposite, otherwise there is only one directed edge.

Adjacent to the edges it is also possible to add weights to the nodes. Those weights result from the road class, for example highways or country roads. If there is a crossover, the node will get the value of the fastest street, that is the least value.

$$l_{\text{frac}} = \sqrt{dx^2 + dy^2} \quad (1)$$

$$\text{mit } dx = 111.3 \cdot \cos(lat) \cdot (lon1 - lon2)$$

$$dy = 111.3 \cdot (lat1 - lat2)$$

$$lat = (lat1 + lat2)/2 \cdot 0.01745$$

$$t_{\text{frac}} = l_{\text{road}}/v_{\text{max}} \quad (2)$$

$$t_{\text{road}} = \sum_{t \in T} t_{\text{frac}} \quad (3)$$

## 8.2 Routing

The routing algorithm used for calculating the shortest path is the D\* Lite algorithm introduced by Koenig and Likhachev in 2002 [4, 5]. This algorithm bases on the Lifelong Planning A\* (LPA\*) from Koenig and Likhachev (2002) [6]. If an obstacle occurs both algorithms can update a shortest path without recalculating the whole way but the LPA\* uses a fixed start point whereas the D\* Lite is able to handle a moving start point.

To improve the algorithm once more Fibonacci heaps are used [1]. With a data structure like this it is possible to find the next neighbour node with the shortest distance faster.

## 8.3 Sharing

Our implementation of the information sharing process is based on vehicle classes. We support three of them (trucks, cars, motorcycles) in our simulation. Every information generated by other vehicles differs in relevance and content depending on these classes. Although every report is shared equally, only vehicles of corresponding classes receive those messages via online feeds. This emulates social target groups and subscriptions.

Every junk of data travels from vehicle to vehicle if they are connected. We implemented a simple connection establishment method, which only depends on range. The submission of data packages follows the “nearest peer” approach, described earlier. The decision which data package has higher sharing priority depends on its lifetime and hops it has taken. Only selected peers can access our web service, which provides storing capabilities and distributed access functionality. This forces the majority of vehicles to wait until relevant data has reached the final receiver, but allows good traceability of data travels and routing.

The global feed of high-level messages can be accessed by mobile devices with online connectivity, just like every real-world road user could. Feeds contain specific information only relevant for certain vehicle classes. Our social online service provides three different feeds. Local feeds of vehicles on the road can be accessed in our desktop view of our simulation. They contain every discovered data package and prediction whether this information is relevant to the route ahead or not. Newly detected obstacles or environmental conditions are also visible in the map view.



## 9 Conclusion

Mobile communication technologies are evolved and well-established, now it is time to make good use of them. Using vehicle-to-vehicle connections for short ranged and temporary communication is important for more secure roads and traffic, but also a bit shortsighted. The benefits of globally shared information are countless for all kinds of vehicle classes and types. We introduced a workflow of vehicular data generation and sharing, adapting methods and habits from the daily use of social media by us humans. Our approach is based on established technologies and our concept is supported by a simulation with focus on adaptive routing and traceable data packages. The information shared in this virtual world is presented in a web view and accessible by mobile devices emulating a “Social Automobile Network”.

## References

- [1] M. L. Fredman and R. Tarjan. ‘Fibonacci Heaps And Their Uses In Improved Network Optimization Algorithms’. In: *Foundations of Computer Science, 1984. 25th Annual Symposium on* (1984), pp. 338–346. ISSN: 0272-5428.
- [2] M. Gerla and L. Kleinrock. ‘Vehicular networks and the future of the mobile internet’. In: *Computer Networks* 55 (2011), pp. 457–467.
- [3] P. Hart, N. Nilsson, and B. Raphael. ‘A Formal Basis for the Heuristic Determination of Minimum Cost Paths’. In: *Systems Science and Cybernetics, IEEE Transactions on* 4.2 (1968), pp. 100–107. ISSN: 0536-1567.
- [4] S. Koenig and M. Likhachev. ‘D\* Lite’. In: *Proceedings of the AAAI Conference of Artificial Intelligence (AAAI)* (2002), pp. 476–483.
- [5] S. Koenig and M. Likhachev. ‘Improved fast replanning for robot navigation in unknown terrain’. In: *Robotics and Automation, 2002. Proceedings. ICRA ’02. IEEE International Conference on* 1 (2002), 968–975 vol.1.
- [6] S. Koenig and M. Likhachev. ‘Incremental A\*’. In: *Advances in Neural Information Processing Systems (NIPS)* (2002).
- [7] R. Popescu-Zeletin, I. Radusch, and M. A. Rigani. ‘Vehicular-2-X Communication: State-of-the-Art and Research in Mobile Vehicular Ad hoc Networks’. In: *Springer Science & Business Media* (2010).
- [8] W. Sha, D. Kwak, B. Nath, and L. Iftode. ‘Social vehicle navigation: integrating shared driving experience into vehicle navigation.’ In: *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications (ACM)* (2013).
- [9] R. Stanica, E. Chaput, and A.-L. Beylot. ‘Simulation of vehicular ad-hoc networks: Challenges, review of tools and recommendations.’ In: *Computer Networks* 55 (14 2011), pp. 3179–3188.
- [10] N. Vastardis and K. Yang. ‘Mobile social networks: architectures, social properties and key research challenges.’ In: *Communications Surveys & Tutorials, IEEE* 15 (2013), pp. 1355–1371.
- [11] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan. ‘Vehicular ad hoc networks (VANETS): status, results and challenges’. English. In: *Telecommunication Systems* 50.4 (2012), pp. 217–241. ISSN: 1018-4864.



# Selective Compensation of Harmonics in Smart Grid with the Active Filter of a Charging Station for Electric Vehicle

JEREMIE FOULQUIER AND LUTZ RAUCHFUSS

This work investigates a smart charging station for the development of smart grid technology in synergy with the electric mobility. In addition to the DC fast charging function, the charging station compensates reactive power and reduces harmonics. This paper presents the inverter control for the mitigation of harmonics from multiple low power Harmonic Generating Loads (HGLs) in a Low Voltage (LV) grid. The inverter acts as adjustable admittance for each individual harmonic frequency in accordance with the harmonic voltage and the grid admittance at its installation point. The aim of the compensation is to maintain the harmonic voltages under the standard threshold and to collect the harmonic currents from the HGLs installed close to the charging station by shaping an optimal compensation current. The harmonic detection method and the controller strategy are described and experimental result on a prototype attests the performances of the harmonics compensation.

## 1 Introduction

The widespread use of electric vehicles will necessitate the installation of many charging stations and will present new challenges for the utilities in particular the peak power management but also new opportunities in which electric vehicles can serve as resources to the grid. Besides, the change from central generation to distributed generation of power encourages the utilities to invest in grid modernization and to develop new decentralized management of Power Quality (PQ). Active power regulation according to the grid frequency, reactive power compensation and harmonics reduction are established tasks that are developed and begin to be integrated in grid converters to provide ancillary PQ ability.

Because power electronics improve the energy efficiency and therefore are used in modern equipment, the part of non-linear low power consumers in the load profile remarkably expands and the summation of all these HGLs increase the level of harmonic distortion of the mains supply. Non-linear loads could be categorized into identified and unidentified loads [2]. Although high-power loads are identified and the utility knows them, small loads are unidentified and are installed in the LV grid. Due to their large number, unidentified loads could not be treated individually for PQ purpose; a global compensation of their harmonics with several Active Power Filters (APFs) is required.

Because a series APF compensates harmonics just in a part of the grid, only a shunt APF can globally reduce the harmonics in the LV grid. Basically, shunt APFs work with current detection and are used to compensate the harmonic currents of identified HGL. Some research on the compensation of harmonics resonances with a shunt APF working with voltage detection has been done [13]. The shunt APF, called Active Resonance Damper (ARD) is installed at the end of a distribution line and presents a resistance equal to the characteristic impedance of the feeder [10]. Operating at similar conductance for all harmonics frequencies, the ARD can magnify another harmonics. To avoid this, a discrete frequency tuning shunt APF has been created to adjust damping conductance of individual harmonics frequencies [6]. The effects of distributed and end-of-line photovoltaic inverters as shunt APF systems on harmonics compensation have been investigated and suggest different compensation strategies depending on the system load characteristics [8]. So far, only the reduction of the harmonic voltages is considered. The new challenge is the compensation of the harmonic currents of all HGLs that are dispersed around the charging station. While the harmonic currents come simultaneously from upstream and downstream sources around the compensation point, the control strategy, as studied in [1], is based on voltage detection.

The project of the smart charging station used as a shunt APF for the decentralized control of PQ aim to develop smart grid in synergy with electric mobility and to prevent grid expansion. For a minimal impact on the grid, the three phase inverter shape a symmetrical and sinusoidal charging current in accordance with the harmonic requirements defined in [7]. This paper focuses on the inverter control for the compensation of the harmonics in the LV grid. The inverter of the charging station can adjust dynamically and separately for each harmonic frequency its admittance with optimally phase and amplitude to comply with the voltage harmonics and the grid admittance at its installation point. The main purpose of this compensation is to maintain the harmonic voltages in LV grid under the standard threshold [12] and restrain the harmonic currents of the HGLs to flow into the medium voltage grid.

The harmonic measurement is presented in section 2. Section 3 presents the proposed algorithm for the harmonics compensation and section 4 explains the control of the inverter of the charging station. The effectiveness of the harmonic compensation is verified experimentally on a prototype in section 5.

## 2 Harmonic Detection

### 2.1 Sliding Discrete Fourier Transform

Most harmonic measurement methods are with Fourier bases and can be classified into batch signal processing and recursive analysis methods [4]. Recursive methods provide real-time spectral analysis that is required in control applications. The Sliding Discrete Fourier Transform (SDFT) is a recursive method that is widely used in APF control for selective harmonic compensation [11]. It is an efficient method for narrowband spectrum analysis of only some frequencies. A good description of the SDFT algorithm can be found in [5]. The harmonic of order  $h$  of a discrete signal  $x(n)$  with sampling frequency  $f_s$  is calculated with the SDFT by a frequency shift  $dq$  and a moving average of  $N$  sample.

$$\underline{x}_{h,dq}(n) = x(n) \cdot e^{-j2\pi \cdot h \cdot dq \cdot n / f_s} \quad (1)$$

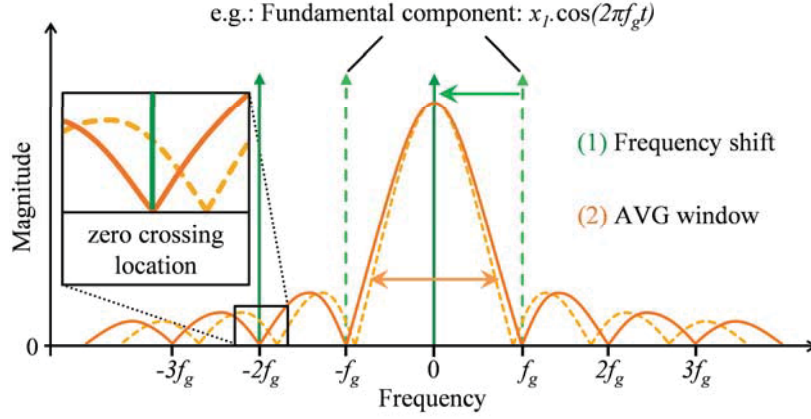


Figure 1: SDFT frequency magnitude response: synchronization adjusts the location of the zero crossings and reduces the spectral leakage

$$\sum_{n=1}^N \underline{x}_{h,dq}(n) = \sum_{n=0}^{N-1} \underline{x}_{h,dq}(n) - \underline{x}_{h,dq}(0) + \underline{x}_{h,dq}(N) \quad (2)$$

To avoid spectral leakage, the rectangular SDFT window of length  $T_{DFT}$  must be synchronized with the mains frequency  $f_g$  as in the IEC 61000-4-7 and shown in Figure 1. Commonly the sampling frequency is synchronized with the mains frequency by means of resampling or interpolation mechanism. Another technique, presented here and similar to interpolation, computes a stepless adaptive SDFT window (e.g.  $T_{DFT} = 2401.35$  samples) by multiplying the fractional part of the window length  $m = 0.35$  with the last sample  $x_{h,dq}(N + 1)$  of the SDFT window as follows

$$\underline{X}_h = \frac{\sqrt{2}}{T_{DFT}} \cdot \left( \sum_{n=1}^N \underline{x}_{h,dq}(n) + \underline{x}_{h,dq}(N + 1) \cdot m \right) \quad (3)$$

$$T_{DFT} = \frac{10 \cdot f_S}{f_g}; \quad N = \lfloor T_{DFT} \rfloor; \quad m = T_{DFT} - N$$

This method synchronizes precisely in the  $\mu\text{Hz}$  range the SDFT with the main frequency while the sampling frequency stays constant as stable time frame. This method is computationally fast and is compatible with oversampling function of Analog-to-Digital Converters (ADC).

## 2.2 Frequency Locked Loop

Passive HGLs like transformers or rectifiers generate harmonic frequencies that are an exact multiple of the mains frequency and cause the most distortions. An accurate synchronization with the mains frequency is required for their compensation and is released here with a frequency locked loop (FLL) based on the SDFT as shown in Figure 2.

In the SDFT, the frequency shifting is equal to the oscillator and the phase measurement of the fundamental component  $\Delta\varphi_1$  of the grid voltage  $u_a$  is equal to the phase detector. The derivative of  $\Delta\varphi_1$  provides the frequency deviation between the SDFT frequency and the mains frequency. This deviation is corrected by the integrator with a

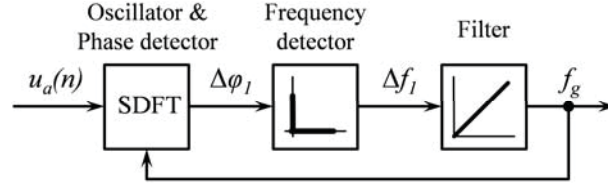
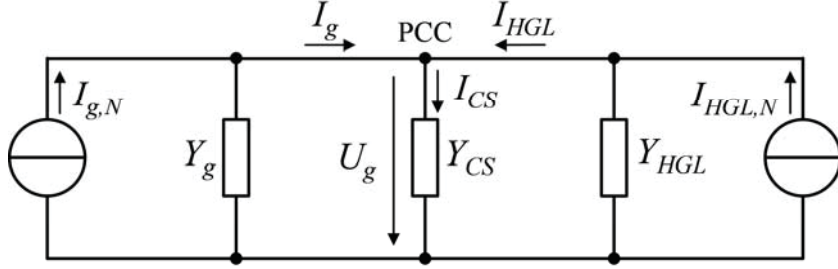


Figure 2: Frequency locked loop for the synchronization with the mains frequency

Figure 3: Norton equivalent circuit of the grid  $g$ , the charging station  $CS$  and the harmonic generating loads  $HGL$ 

gain  $K_{i,FLL}$  to adjust the loop dynamics. The phase coherence between the measurement of the  $h^{th}$  harmonic and the voltage fundamental is maintained in the SDFT by

$$\theta_{h,DFT} = h \cdot (2\pi \cdot f_g \cdot n / f_S + \Delta\varphi_1) \quad (4)$$

### 2.3 Fundamental Component Filter

The measurement of the harmonic voltages is improved by the pre-filtering of the fundamental component. The fundamental is measured by means of the SDFT and subtracted from the grid voltage.

## 3 Compensation Algorithm

Thanks to the SDFT, the harmonic compensation algorithm operates for single harmonic frequencies and it will be studied with the Norton equivalent circuit presented in Figure 3. In this figure, the HGLs in the vicinity of PCC and the LV grid are the harmonic sources  $I_{HGL,N}$  and  $I_{g,N}$ , respectively as well as  $Y_{HGL}$  and  $Y_g$  are the respective harmonic admittances.  $Y_{CS}$  is the harmonic admittance of the shunt APF.

The  $h^{th}$  harmonic voltage at PCC is calculated with the principle of superposition as follows

$$\underline{U}_g = \frac{\underline{I}_{g,N} + \underline{I}_{HGL,N}}{\underline{Y}_g + \underline{Y}_{CS} + \underline{Y}_{HGL}} \quad (5)$$

and the result of the compensation can be written as

$$\lambda = \left| \frac{\underline{U}_g(\underline{Y}_{CS} \neq 0)}{\underline{U}_g(\underline{Y}_{CS} = 0)} \right| = \left| \frac{\underline{Y}_g + \underline{Y}_{HGL}}{\underline{Y}_g + \underline{Y}_{CS} + \underline{Y}_{HGL}} \right| \quad (6)$$

where  $\lambda$  denote the harmonic voltage compensation ratio before ( $Y_{CS} = 0$ ) and after ( $Y_{CS} \neq 0$ ) the compensation with the shunt APF. The higher the absolute value of the denominator in equation (5), the lower the harmonic voltage and the more efficiently

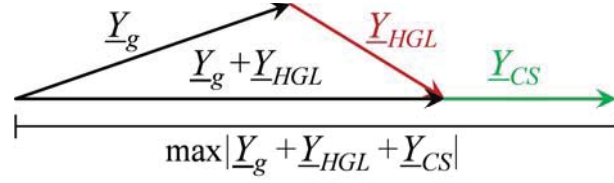


Figure 4: Vector diagram of the admittances for the optimal compensation

the compensation ( $\lambda < 1$ ). The summation of the three admittances is illustrated in Figure 4.

It shows that best compensation is obtained if the admittance of the shunt APF is proportional to the sum of the admittances of the grid and the HGLs together,

$$\underline{Y}_{CS} = K_{CS} \cdot \underline{Y}_{g+HGL} \quad (7a)$$

$$\phi_{Y_{CS}} = \phi_{Y_{g+HGL}} \quad (7b)$$

where the proportional factor  $K_{CS}$  is a positive real number. Consequently, the phase  $\phi_{Y_{CS}}$  of the admittances of the shunt APF must be equal to the phase  $\phi_{Y_{g+HGL}}$  of the admittances at PCC before compensation. To achieve this optimal compensation, the shunt APF generates a harmonic current  $\underline{I}_{CS}$  that is controlled in magnitude and phase according to the harmonic voltage  $\underline{U}_G$  at PCC as explained in section 4. In addition to the harmonic voltage mitigation, the second target is to reduce, or at least not to amplify the harmonic current  $\underline{I}_g$  from the grid. According to Figure 3 this current can be calculated as

$$\underline{I}_g = U_g \cdot (\underline{Y}_{CS} + \underline{Y}_{HGL}) - \underline{I}_{HGL,N} \quad (8)$$

Introducing a harmonic source ratio as follows

$$\underline{K}_h = \frac{\underline{I}_{g,N}}{\underline{I}_{HGL,N}} \quad (9)$$

and including (5), this current can be written as

$$\underline{I}_g = \frac{\underline{K}_h \cdot (\underline{Y}_{CS} + \underline{Y}_{HGL}) - \underline{Y}_g}{\underline{Y}_g + \underline{Y}_{CS} + \underline{Y}_{HGL}} \quad (10)$$

The result of the compensation can be written as

$$\gamma = \left| \frac{\underline{I}_g(\underline{Y}_{CS} \neq 0)}{\underline{I}_g(\underline{Y}_{CS} = 0)} \right| = \lambda \cdot \left| \frac{\underline{K}_h \cdot (\underline{Y}_{CS} + \underline{Y}_{HGL}) - \underline{Y}_g}{\underline{K}_h \cdot \underline{Y}_{HGL} - \underline{Y}_g} \right| \quad (11)$$

where  $\gamma$  denotes the harmonic current compensation ratio. Once more, all quantities in this section refer to a single harmonic frequency. The result of the compensation from equations (11) and (6) is illustrated in Figure 5 for  $\underline{Y}_g = 10S/0^\circ$  and  $\underline{Y}_{HGL} = 2S/0^\circ$  and can be divided into three cases:

- 1) The grid generates no harmonic,  $\underline{I}_{g,N} = 0 \Leftrightarrow \underline{K}_h = 0$ . It results before the compensation that, the amount of harmonic current between grid and HGLs is medium, and the voltage distortion at PCC is medium too, compared to the two other cases. No grid harmonic infers in equation (11) that, voltage and current compensation ratio are the same  $\gamma = \lambda$  and, as the black curve shows in Figure 5, the shunt APF reduces the harmonic voltage  $\underline{U}_g$  and current  $\underline{I}_g$  at the same time:  $\gamma = \lambda < 1$ . The compensation has only positive effects for the grid.



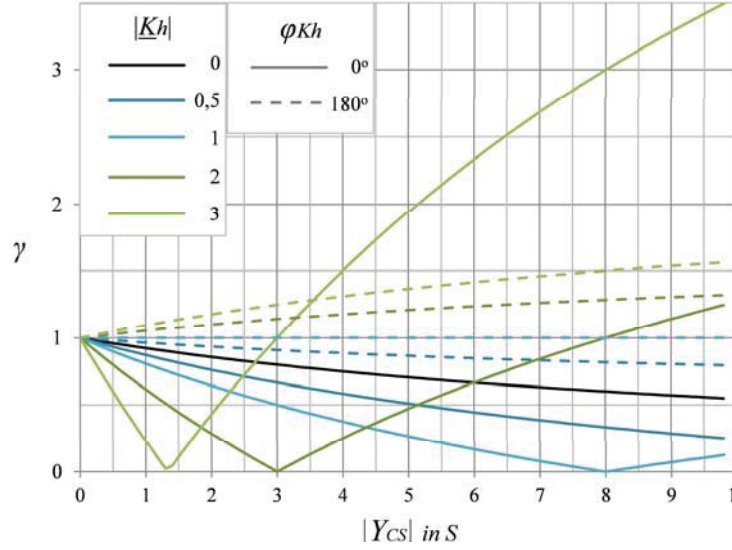


Figure 5: Harmonic current compensation ratio  $\gamma$  according to the shunt APF admittance  $|Y_{CS}|$  and the harmonic current source ratio  $\underline{K}_h$  for  $\underline{Y}_g = 10S/0^\circ$  and  $\underline{Y}_{HGL} = 2S/0^\circ$

- 2) The grid is a harmonic source in opposite phase to the HGLs harmonic source, i.e.  $|\underline{K}_h| > 0$  and  $\phi_{Kh} = 180^\circ$ . Consequently, these two sources have a generator-consumer relation where the direction of the harmonic current at PCC is only a function of the strongest current source i.e if  $|\underline{K}_h| > 1$  or  $|\underline{K}_h| < 1$ . This opposite phase results in a high amount of harmonic current between grid and HGLs and a low voltage distortion at PCC before the compensation starts. Once the shunt APF compensates, the harmonic voltage  $\underline{U}_g$  is reduced (black curve). In the case that the strongest harmonic source are the HGLs, i.e.  $|\underline{K}_h| < 1$ , the grid current  $\underline{I}_g$  is reduced (blue dashed line). In the other case,  $|\underline{K}_h| > 1$ , (green dashed line) the compensation amplifies the grid current so that a high compensation rate, i.e.  $K_{CS} > 1$ , is unfavorable. For a moderate compensation rate  $K_{CS} < 1$ , this amplification is not critical for the grid.
- 3) The grid is a harmonic source in phase with the HGLs harmonic source, i.e.  $|\underline{K}_h| > 0$  and  $\phi_{Kh} = 0^\circ$ . These two sources have a generator-generator relation and the direction of the harmonic current at PCC is a function of the strongest current source divided by its respective admittance, i.e. the strongest voltage source. It results in a low amount of harmonic current between grid and HGLs but a high voltage distortion at PCC before the compensation. While the APF compensates, the harmonic voltage  $\underline{U}_g$  is reduced (black curve) and, in the case that the HGLs is the strongest voltage source, the grid current  $\underline{I}_g$  is reduced (blue line). Because the harmonic current at PCC was low before the compensation, the harmonic current compensation ratio increase rapidly (green curve) and high compensation rate are unfavorable for the grid current. In the other case, the grid is the strongest voltage source and the compensation always amplifies the grid harmonic current.

Summary: The shunt APF admittance defined in equation (7) is always profitable for the harmonic voltage mitigation at PCC. Because the harmonic current flows from the strongest to the lowest source, the APF admittance reduces the harmonic current of the lowest source. Thus, the compensation is most effective when the HGLs are the strongest source and are located in the vicinity of the shunt APF. The moderate



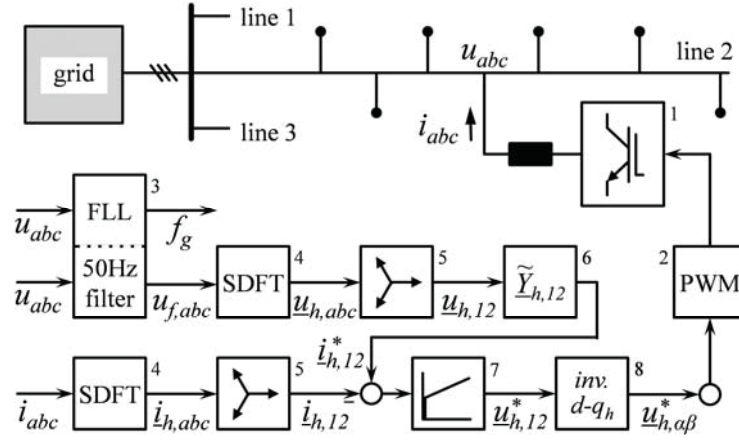


Figure 6: Block diagram of the inverter control for the compensation of one harmonic

compensation rate  $K_{CS} < 1$  for partial compensation ensures that harmonic current amplification of the strongest source is not excessive.

## 4 Inverter Control

A block diagram of the inverter control for the compensation of one harmonic is presented in Figure 6.

The charging station is equipped with a 3-phase 2-level inverter (block 1) controlled by space vector modulation [9, p. 40] (block 2). The synchronization with the voltage fundamental component and its filtering (block 3) and the harmonics measurement (blocks 4) has been summarized in section 2. The measured phasors of the voltage and current harmonics of the 3 phases  $a, b, c$  are computed into positive and negative sequence 1, 2 (block 5) for the compensation of each harmonic in the appropriated sequence. The control of the harmonic admittance  $Y_{h,12}$  of the shunt APF contains two loops. The inner control loop adjusts the compensation current  $i_{h,12}$  with PI controllers (block 7). The outer control loop adjust the optimal admittance  $Y_{h,12}$  with extremum seeking controls [3] (block 6, Figure 7): the magnitude and phase of the compensation current  $i_{h,12}^*$  are alternatively adjusted according to the rising or falling magnitude of the compensated harmonic voltage  $u_{h,12}$ . As defined in equation (7), the optimal current phase is obtained when the harmonic voltage  $u_{h,12}$  has the lowest magnitude i.e. when the compensation has best efficiency. In a similar way, the optimal current magnitude is obtained when the harmonic voltage is reduced to the fixed threshold and no amplification occurs. To increase the correlation between the compensation current and its effect on the magnitude of the harmonic voltage, a series of measurements for two different current adjustments are applied.

Figure 7 outlines the control of the phase of the compensation current for one harmonic. Two different phases  $\phi_{i,h}$  and  $\phi_{i,h} + \Delta\phi_{i,h}$  are  $n$  times alternatively adjusted while the magnitude of the harmonic voltage  $U_h$  is measured and respectively added or subtracted to  $\sum U_h$ . This differential sum is the comparison of the compensation effect between the two different phases. Then, the best phase is held and the sign of the new phase step  $\Delta\phi_{i,h}$  is selected. The phase step is adjusted for best precision while maintaining a distinct compensation effect. The principle to control the magnitude of the compensation current is the same, with the distinction that the harmonic voltage is adjusted to the

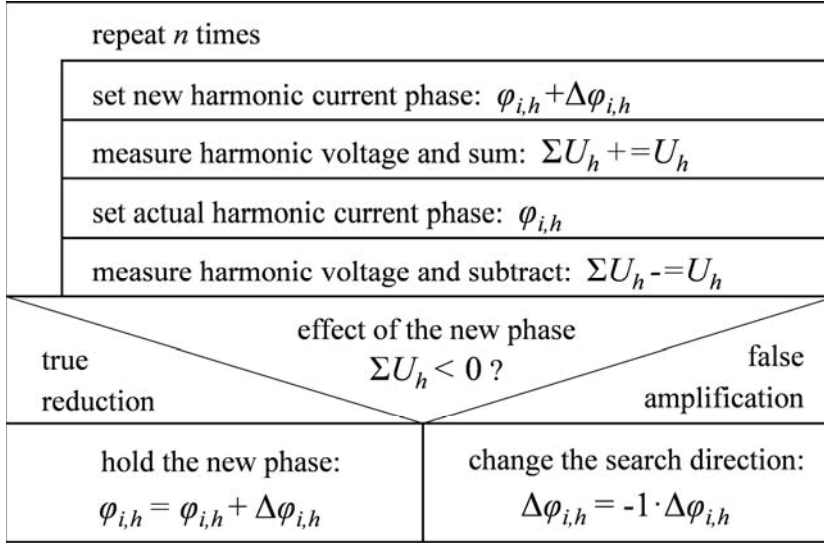
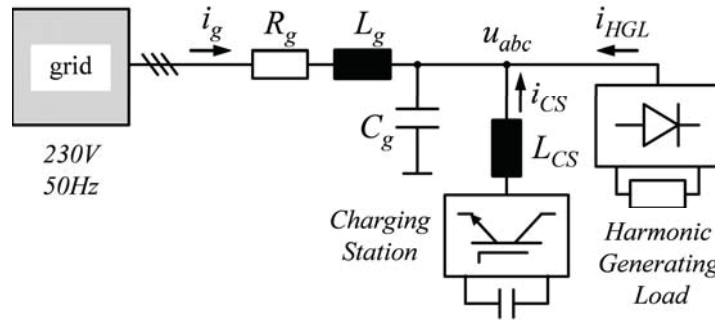


Figure 7: Control of the phase for the optimal compensation current

Figure 8: Laboratory test circuit:  $R_g = 2,5\Omega$ ,  $L_g = 0,8\text{mH}$ ,  $C_g = 4\mu\text{F}$ ,  $L_{CS} = 12\text{mH}$ 

threshold value and not to the lowest value. The compensation magnitude is constantly adapted to avoid harmonic amplification while the grid distortion decreases.

## 5 Laboratory Test Results

The operation of a 10-kVA prototype of the charging station with the proposed control method is demonstrated in laboratory with the test circuit illustrated in Figure 8.

The inverter control is implemented on the digital signal processor ADSP-BF537 EZ-KIT Lite. The ADC, the SFFT, the digital control and the switching frequency of the inverter run synchronously at 12 kHz. The passive elements  $R_g$ ,  $L_g$  and  $C_g$  increase the grid impedance and thus amplify the harmonic voltage distortion caused by the rectifier. Only steady-state active filtering performance is studied while fast transient response is not required for the compensation of grid harmonics produced by multiple low power HGLs. The 5<sup>th</sup>, 11<sup>th</sup>, 13<sup>th</sup>, 17<sup>th</sup>, 19<sup>th</sup>, 23<sup>th</sup> and 25<sup>th</sup> harmonics are compensated with independent control loops while the 7<sup>th</sup> is not compensated to verify the selectivity of the compensation. Figure 9 shows the uncompensated grid current  $i_{g,a}$  and grid voltage  $u_a$  at charging station connection point. The current waveform displays high harmonic levels with a THD = 28% which in turn leads to noticeable voltage harmonics. Figure 10 shows the grid voltage and current while the compensation

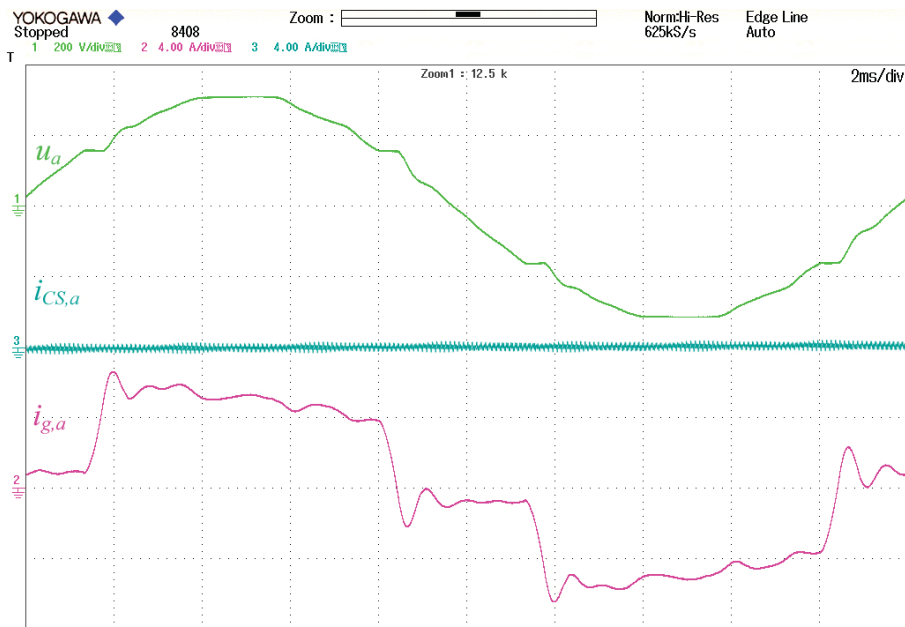


Figure 9: Grid voltage and current without compensation



Figure 10: Selective harmonic compensation with voltage controlled APF

operates and reaches the steady state. The voltage controlled compensation of the harmonics is activated in the appropriated positive or negative sequence if the threshold of 1V is overstepped. As intended, the compensation current  $i_{CS,a}$  has a high harmonic level and virtually no fundamental component. The effectiveness of the shunt APF is obvious, as the grid current waveform with a THD = 10% and the grid voltage waveform show. Harmonic spectrums of currents and voltages with/without active filtering are compared on Figure 11 with respectively filled/outlined bars. The partial mitigation of the grid harmonic voltages under the threshold and the selectivity of the compensation are evident. The voltage controlled compensation algorithm adapts the admittance of the charging station for each harmonic and thus absorb the harmonic currents of the HGL installed in the vicinity. It results in a significant reduction of the harmonic current of the grid and lower voltage distortion.

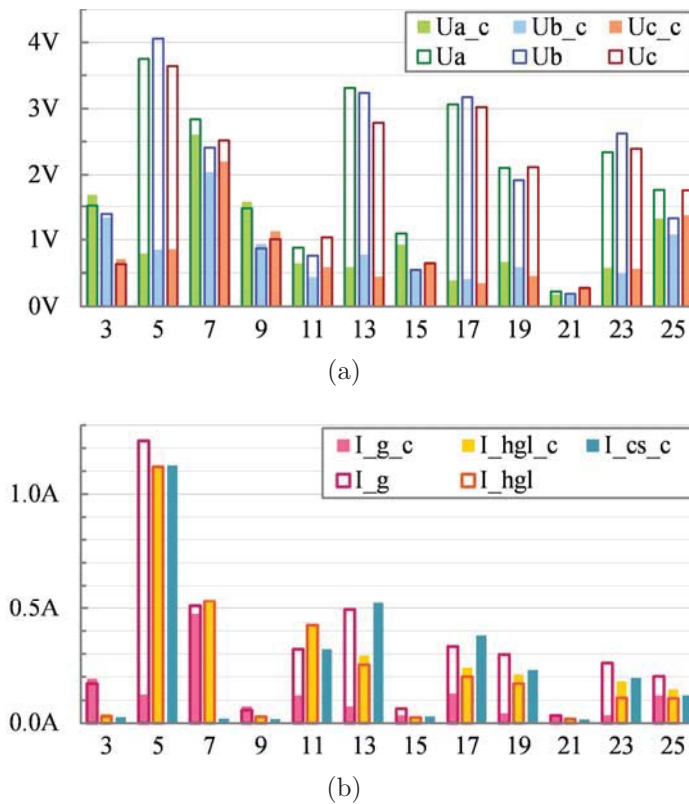


Figure 11: Harmonic spectrums with/without compensation: filled/outlined bars, (a) grid voltages  $U_{abc}$ , (b) currents phase  $a$

## 6 Conclusion

To develop smart grid in synergy with electric mobility, a smart charging station used as a shunt APF in LV grid is proposed. In response to the rising number of non-linear modern equipment this paper has described the selective harmonic compensation as an ancillary function of the charging station. The sliding discrete Fourier transform synchronized with a frequency locked loop is adopted for the harmonics measurement. Then, the adjustment of the shunt APF as a controlled admittance is studied. At anytime the shunt APF can mitigate the harmonic voltages of the grid at its installations point. The reduction at the same time, of the harmonic current has been examined and is

achieved if the shunt APF is installed in the vicinity of the HGLs. The objective is a moderate compensation rate to maintain the harmonic voltages under the threshold value and restrain a part of the harmonic currents of the HGLs from flowing into the medium voltage grid. The structure of the harmonic control loops is described and the adjustment of the compensation current in magnitude and phase with extremum seeking control is explained. Experimental results demonstrate the effectiveness of the voltage controlled selective harmonic compensation implemented in the prototype of the charging station.

## References

- [1] H. Akagi. ‘Control strategy and site selection of a shunt active filter for damping of harmonic propagation in power distribution systems’. In: *Power Delivery, IEEE Transactions on* 12.1 (1997).
- [2] H. Akagi. ‘New trends in active filters for power conditioning’. In: *Industry Applications, IEEE Transactions on* 32.6 (1996).
- [3] B. Calli, W. Caarls, P. Jonker, and M. Wisse. ‘Comparison of extremum seeking control algorithms for robotic applications’. In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on* (2012).
- [4] G. Chang and C.-I. Chen. ‘Measurement techniques for stationary and time-varying harmonics’. In: *Power and Energy Society General Meeting, 2010 IEEE* (2010).
- [5] E. Jacobsen and R. Lyons. ‘The sliding DFT’. In: *Signal Processing Magazine, IEEE* 20.2 (2003).
- [6] T.-L. Lee, J.-C. Li, and P.-T. Cheng. ‘Discrete Frequency Tuning Active Filter for Power System Harmonics’. In: *Power Electronics, IEEE Transactions on* 24.5 (2009).
- [7] *Limits for harmonic currents produced by equipment connected to public low-voltage systems with input current > 16 A and < 75 A per phase*. IEC 61000-3-12. Electromagnetic compatibility (EMC), 2011.
- [8] S. Munir and Y. W. Li. ‘Residential Distribution System Harmonic Compensation Using PV Interfacing Inverter’. In: *Smart Grid, IEEE Transactions on* 4.2 (2013).
- [9] N. P. Quang and J. Dittrich. *Vector Control of Three-Phase AC Machines*. Ed. by Springer. 2008.
- [10] W. Santana, K. Al-Haddad, and L. da Silva. ‘Modeling and active damping of harmonic propagation on electric distribution systems’. In: *Electrical Power & Energy Conference (EPEC), 2009 IEEE* (2009).
- [11] K. Sozanski. ‘Selective harmonics compensator’. In: *Nonsinusoidal Currents and Compensation (ISNCC), 2013 International School on* (2013).
- [12] *Voltage characteristics of electricity supplied by public distribution networks*. DIN EN 50160. 2011.
- [13] K. Wada, H. Fujita, and H. Akagi. ‘Considerations of a shunt active filter based on voltage detection for installation on a long distribution feeder’. In: *Industry Applications Conference, 2001. Thirty-Sixth IAS Annual Meeting. Conference Record of the 2001 IEEE* 1 (2001).



## **SCIENTIFIC REPORTS**

## **WISSENSCHAFTLICHE BERICHTE**

The main aspect of the Scientific Reports is to promote the discussion of modern developments in research and production and to stimulate the interdisciplinary cooperation by information about conferences, workshops, promotion of partnerships and statistical information on annual work of the Hochschule Mittweida (FH) University of Applied Sciences. This issue will be published sporadically. Contributors are requested to present results of current research, transfer activities in the field of technology and applied modern techniques to support the discussion among engineers, mathematicians, experts in material science and technology, business and economy and social work.

Journal of the University of Applied Sciences Mittweida  
Wissenschaftliche Zeitschrift der Hochschule Mittweida

Hochschule Mittweida  
University of Applied Sciences  
Referat Forschung  
PF 1457  
D-09644 Mittweida

Tel: +49 (0) 3727 / 58 1264

Fax +49 (0) 3727 / 58 1178

© Hochschule Mittweida  
University of Applied Sciences  
Technikumplatz 17  
D-09648 Mittweida