Asymptotic Properties of Probabilistic Models

Thesis for the Degree of Doctor of Philosophy (PhD)

László Fórián
Supervisor: Prof. Dr. István Fazekas

UNIVERSITY OF DEBRECEN
Doctoral Council of Natural Sciences and Engineering
Doctoral School of Informatics
Debrecen, 2026

Hereby I declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Engineering, Doctoral School of Informatics, University of Debrecen in order to obtain a PhD Degree in Informatics at Debrecen University.

The results published in the thesis are not reported in any other PhD theses.

Debrecen, 2026.................

. . . .
László Fórián
candidate

Hereby I confirm that László Fórián candidate conducted his studies with my supervision within the Theoretical foundation and applications of information technology and stochastic systems Doctoral Programme of the Doctoral School of Informatics between 2020 and 2024. The independent studies and research work of the candidate significantly contributed to the results published in the thesis.
I also declare that the results published in the thesis are not reported in any other theses.
I support the acceptance of the thesis.

Debrecen, 2026.................

. . . .
Dr. István Fazekas
supervisor

Asymptotic Properties of Probabilistic Models

Dissertation submitted in partial fulfilment of the requirements for the doctoral (PhD) degree in the science of informatics.

Written by László Fórián certified mathematician.

Prepared in the framework of the Doctoral School of Informatics of University of Debrecen (Theoretical foundation and applications of information technology and stochastic systems programme)

Dissertation supervisor: Dr. István Fazekas

The official opponents of the dissertation:

Dr.	....	
Dr.	....	

The evaluation committee:

chairperson:	Dr.	....
members:	Dr.	....
	Dr.	....
	Dr.	....
	Dr.	....

The date of the dissertation defence: 20... ....

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor, Dr. István Fazekas, for his invaluable guidance, professional insight and support throughout my journey. Without his inspiration, knowledge and mentorship, this accomplishment would not have beeen possible.

I also would like to thank my family and friends for their unwavering support, patience and encouragement.

Contents

List of Tables

List of Figures

Introduction

In the Preface of their fundamental book [37], Gnedenko and Kolmogorov wrote

> "In the formal construction of a course in the theory of probability, limit theorems appear as a kind of superstructure over elementary chapters, in which all problems have finite, purely arithmetical character. In reality, however, the epistemological value of the theory of probability is revealed only by limit theorems. Moreover, without limit theorems it is impossible to understand the real content of the primary concept of all our sciences — the concept of probability. In fact, all epistemologic value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity. The very concept of mathematical probability would be fruitless if it did not find its realization in the frequency of occurrence of events under large-scale repetition of uniform conditions (a realization which is always approximate and not wholly reliable, but which becomes, in principle, arbitrarily precise and reliable as the number of repetitions increases)."

These principles govern most of the studies in probability theory including the mathematical research, the statistical applications and simulation experiments. The basic asymptotic theorems of probability theory concern independent and identically distributed random variables. Their main versions are the strong law of large numbers and the central limit theorem. The first one contains almost sure convergence, the second one convergence in distribution. Additionally, there exist more sophisticated limit theorems, e.g. functional limit theorems. Donsker's theorem and the limit theorem for the empirical process are the best-known functional limit theorems, see [9].

If there is no convergence in distribution, we might be able to find accompanying distributions, these kind of theorems are also called merging theorems. The merging theorems can have functional versions, e.g. in [22] a functional merge theorem for laws being in the domain of geometric partial attraction of a semistable law is presented.

In this thesis, we shall obtain the following types of limit theorems: theorems containing almost sure convergence, convergence in distribution, functional limit theorems and we shall also find accompanying distributions.

In a usual limit theorem the basic structure is a sequence of random variables or a real valued continuous time stochastic process. In the first two chapters of this thesis, the basic structures are more complex, they are networks. However, our limit theorems concern certain real valued characteristics of the network, so our limit theorems are usual limit theorems. We do not have a theorem for the possible limit structure. Here we just mention that nowadays a popular research goal is to find the limit structure of sequences of certain mathematical models. As an example we refer to the permuton, i.e. a probability measure which serves as the limiting structure of a convergent permutation sequence, see [47].

The dissertation contains asymptotic results related to the limiting behaviour of network evolution models and contaminated runs of heads in the coin tossing experiment. In Chapter 1, we study a continuous-time network evolution model that is based on cliques of nodes and is governed by a branching process. Chapter 2 studies a parametrized family of discrete-time network evolution models, where the evolution is based on constructions and deletions of cliques. In Chapter 3, the length of the longest at most T-contaminated head runs is considered. We call a run at most T-contaminated if it contains at most T tails. Here we provide a brief overview of the relevant literature, explain the motivation of our work and outline the structure of the thesis.

Nowadays, network theory is a popular and important research field. It studies general properties of networks, provides models and methods to understand their evolution. Well-known large networks are e.g. the World Wide Web, the Internet, metabolic networks, and social networks. It has applications in logistics, electrical engineering, biology, economics, ecology, public health, sociology and many other fields.

A key mathematical tool for describing an evolving network is the theory of random graphs. The vertices represent the nodes of the network and the edges correspond to the connections between the nodes. In real life, a connection can represent a collaboration, an interaction, or some other form of relationship.

Modeling evolving networks as random graphs started with the paper of Erdős and Rényi [20]. In the Erdős-Rényi model, the number of vertices in the graph is fixed and in each time step, a new pair of nodes is connected, which is chosen uniformly at random. Another important step in the history of random graphs and network evolution models was the groundbreaking work of Barabási and Albert [8] with the preferential attachment method. In that evolving graph, a new vertex is added to the network in each time step in a way that the probability that the new vertex is connected to an existing node is proportional to the degree of that already existing vertex. Barabási and Albert also presented empirical studies for several real-life networks to show that they have the power law degree distribution. The famous book by Barabási [7] also contains several general facts on network theory.

However, in [8], Barabási and Albert did not give precise mathematical definitions and precise mathematical proofs. In the fundamental paper [12] Bollobás, Riordan, Spencer and Tusnády formulated a precise mathematical definition for the preferential attachment model and mathematically precise proof for the asymptotic power law degree distribution.

In the paper [57] by Móri, a model for the evolution of random trees was studied and a strong law of large numbers and a central limit theorem were proved for the number of vertices. Rigorous proofs based on martingale theory were presented. An interesting model was studied by Cooper and Frieze [15]. In their paper both preferential attachment and uniform choice of the nodes were allowed. The first publications analysed evolving networks, but there were papers where deletions were also allowed. In [14] vertices and edges are added to the graph using preferential attachment and vertices and edges are deleted randomly.

There are papers that propose general methods to analyse several different networks using the same tool. E.g. in [65] a common framework is given to analyse a wide class of preferential attachment models. A further generalization of the method is presented in [31].

We also mention that after the publication of the paper [8] by Barabási and Albert it was a common belief that most of the large networks are scale-free, that is they have power law degree distribution. It means for the probability that a node has k connections is $P(k) \sim k^{-\gamma}$ for large values of k, where the value of the parameter γ is typically between 2 and 3. However, nowadays there are several publications asserting that it is not the case and there are large networks following other degree distributions. In [13] the authors study real life networks and state that a large portion of biological, social and technological networks are not scale free. In certain cases they offer other approximations for the degree distribution.

Concerning the recent development of network theory, we should mention some directions, researchers and research teams. Remco van der Hofstad and his co-workers study e.g. the configuration model, generalized random graphs, preferential attachment models, and spatial random graphs. His two fundamental books are [44] and [45]. There is another famous research team around László Lovász. His book on this topic is [53]. Their well-known invention is the graphon, i.e. the limit structure of a sequence of dense graphs, see [54] by László Lovász and Balázs Szegedy.

The general aspects of network theory are presented in several textbooks see e.g. the book of Bollobás [10], the book by Newman, Barabási, and Watts [63] and the book by Janson, Luczak and Rucinsky [51].

In Chapter 1 of this thesis, we use branching processes to describe a network evolution, so now we focus on papers using branching processes in the vast literature of network theory. Bollobás and Riordan in [11] considered several problems of random graphs and applied Galton-Watson branching processes to prove some of the theorems. Rudas, Tóth, and Valkó [68] as well as Rudas and Tóth [69] applied continuous-time branching processes to study random tree growth models. In [3], a preferential attachment random graph is embedded in a continuous-time branching process to find the asymptotic behaviour of the graph. In [36], continuous-time branching processes were applied to study an estimator in sublinear preferential attachment trees. In their survey paper [46], Holmgren and Janson studied the asymptotics of certain trees using Crump-Mode-Jagers branching processes. In [67], multi-type preferential attachment trees were studied using the theory of multi-type continuous-time branching processes. Banerjee and Huang [6] studied growing networks via embedding into a continuous-time branching process. In [48], Iksanov et al. underlined that one of the main applications of the theory of branching processes is the study of evolving networks. Usually, the advanced theory of branching processes is applied to prove new results for growing networks. However, sometimes network theory inspires new research in the field of branching processes, e.g., [59].

Most studies in network theory concentrate on connections of two nodes. However, cooperation of more than two units is also important. There are models describing such cooperation. Backhausz and Móri studied three-interactions in [4], while Fazekas and Porvázsnyik in [33] considered N-interactions.

In the first two chapters of our dissertation, we consider network evolution models describing interactions of several units. Our study is motivated by the paper of Móri and Rokob [58]. In that paper, pairwise collaborations were modelled by edges of a graph. A collaboration attracts a newcomer, that is a new vertex that starts collaborating with one or both participants. However, the connections can also disappear. The random graph process is studied by the tools of general time-dependent branching processes.

In the paper [25] the following version of the model of Móri and Rokob [58] was studied. The basic units of the model were not edges, but triangles describing 3-interactions and the evolution of the triangles was governed by a continuous-time branching process. The asymptotic behaviour of the model was studied by the methods of branching processes and it was proved that the number of triangles, edges and vertices have the magnitude $e^{\alpha t}$, where α is the Malthusian parameter.

A further step was the paper [23] of Fazekas and Barta. In that paper, the evolution of the network was based on 2- and 3-interactions, the 2-interactions were described by edges, and 3-interactions were described by triangles. The evolution of the edges and triangles were governed by a two-type continuous-time branching process. The theory of multi-type branching processes was applied to prove the results. The behaviour of the model was described by explicit formulae. It was proved that the number of triangles and edges have the magnitude $e^{\alpha t}$, where α is the Malthusian parameter.

In Chapter 1, our model was motivated by the activities of teams of people. We focused on cooperation among friends, recruitment of party members, recruitment of volunteers, collaboration among scientists, and teams in social networks. In the above cases, a person can be a member of several teams at the same time, new teams can emerge and disappear, and newcomers can join an existing team. A k-clique is a sub-graph containing k vertices and any two different vertices are connected by 1 edge. The cliques will be the individuals of a branching process. The evolution is governed by a multi-type branching process. In Section 1.1 we define our model precisely. In Section 1.2, we present our general results. We calculate the survival function of a clique (Theorem 1.2.1), the mean offspring number of a clique (Corollary 1.2.1), the Laplace transforms (Proposition 1.2.1), and then we turn to the Perron root and the Malthusian parameter. We do not obtain an explicit expression for the Malthusian parameter, so we can numerically compute it. For the asymptotic analysis, we use some known results of the theory of continuous-time branching processes. We mention that the single type Crump-

Mode-Jagers branching processes are discussed e.g. in [50], [39] and [61], and the general multi-type branching processes are described e.g. in [55], [60] and [49]. In Section 1.3 we give a brief summary of the general multi-type branching processes based on [49]. In Section 1.4, we prove asymptotic theorems for the number of cliques of a given size (Theorem 1.4.1). Each has magnitude $e^{\alpha t}$ in the event of non-extinction, where α is the Malthusian parameter. The proof of Theorem 1.4.1 is based on the asymptotic theorems presented in Section 1.3. One of the most interesting questions in the theory of evolving networks is the degree process of a fixed vertex. Therefore, in Section 1.5, we prove asymptotic theorems for the degree of a fixed vertex. For the proof, we introduce a new branching process, called the "good children" process, and then we can apply the asymptotic theorems of Section 1.3. We obtain that the magnitude of the degree of a fixed vertex is $e^{\hat{\alpha} t}$, where $\hat{\alpha}$ is the Malthusian parameter of the "good children" process. In Section 1.6, we obtain the generating functions. We apply the generating functions to find the probability of extinction of the network, see Theorem 1.6.1. In Section 1.7, we show simulation results that support our theorems.

Discrete-time network evolution models are also intensively studied, see [18], and widely used to describe the evolution of real-life networks. Backhausz and Móri [5] introduced a random graph evolution model with moderate edge density. Their model is the following. The starting graph is an empty one of size 2. At each step, two vertices are chosen uniformly at random. If the two vertices chosen are not connected, we connect them with 1 edge. If the two vertices are connected, we delete the connecting edge, add a new vertex to the graph, and connect the new vertex to both of the selected vertices. The result of this procedure is an evolving graph containing n edges after the n^{th} step. In [5], several asymptotic theorems are proved for the number of vertices and the degree of a fixed vertex. Also in [5], a short overview is presented about the edge densities of some well-known random graph models.

In our paper [29], we study the following extension of the model of [5]. Instead of connections of two vertices, we consider connections of k vertices, where $k \geq 2$ is a fixed integer. So the main ingredients of our model are the k-cliques. The evolution of our graph is based on constructions and deletions of k-cliques. When we form a k-clique, then we draw $\binom{k}{2}$ new edges among k vertices, and we add this new clique to the list of k-cliques.

Chapter 2 is based on our paper [29], so we consider the generalization of the model of Backhausz and Móri [5] to k-cliques. The detailed description of the evolution is given in Section 2.2. In Section 2.3, using martingale theory, we prove an almost sure limit theorem for the number of vertices, then we show its asymptotic normality, see Theorem 2.3.1. In Section 2.4, we obtain an almost sure limit theorem for the degree of a fixed vertex, see Theorem 2.4.1. Then, we present an asymptotic normality result for degree of a fixed vertex. In sections 2.3 and 2.4, our results are extensions of the results of [5]. However, the results of Section 2.5 are new for any value of k, including the particular case of $k = 2$ studied in [5]. These new results are functional limit theorems. Theorem 2.5.1 is a functional limit theorem for the number of vertices. Theorem 2.5.2 and Proposition 2.5.1 are multidimensional functional limit results for the joint distribution of the degrees of several fixed vertices. For the proof, we apply general functional limit theorems for martingales. In Section 2.6, we present simulation results.

Chapter 3 is based on our paper [26], where we study a problem related to the usual coin tossing experiment. Let p be the probability of heads and $q = 1 - p$ be the probability of tails. Here p is a fixed number with $0 < p < 1$. We toss a coin N times independently. We write 1 for heads and 0 for tails. Therefore we consider independent identically distributed random variables $X_1, X_2, \ldots, X_N$ with distribution $P(X_i = 1) = p$ and $P(X_i = 0) = q = 1 - p$, $i = 1, 2, \ldots, N$.

Let T be a fixed non-negative integer. We study the length of at most T-contaminated (in other words at most T-interrupted) runs of heads. It means that there are at most T zeros in an m length sequence of ones and zeros.

We show a 2-contaminated run of heads of length 10:

$$0, 1, \cdots, 1, 0, 0, \underbrace{1, 1, 1, 0, 1, 1, 0, 1, 1, 1}\, 0, 0, 1, 1, \cdots$$

Concerning the length of the pure head run in the case of a fair coin, we know that it is about $\log n$, where we use logarithm to base 2. It is mentioned by Pál Révész in the booklet [66] that Tamás Varga used this fact to distinguish real experiments and artificial experiments.

The length of the pure head runs was considered in several papers. The case of the fair coin was studied in the classical paper of Erdős and Rényi [19]. Let log denote the logarithm to base 2 and let [.] denote the integer part, let us use the notation $\mu(N)$ for the longest pure head run in the first N trials, and let $0 < C_1 < 1 < C_2 < \infty$. Then for almost all elementary event ω there exists a finite $N_0 = N_0(\omega, C_1, C_2)$ such that

$$[C_1 \log N] \le \mu(N) \le [C_2 \log N]$$

if $N \ge N_0$.

Besides asymptotic theorems, the length of the longest head run can be studied by recursive formulae [7, 11] and by computer simulations, see [30] and [52].

An early paper obtaining almost sure limit results for the length of the longest runs containing at most T tails is [21]. They considered the usual coin tossing experiment with a fair coin. Let $\mu_T(N)$ denote the longest head run containing at most T tails in the first N trials. Let log denote the logarithm to base 2 and let [.] denote the integer part,

$$h(N) = \log N + T \log\log N - \log\log\log N - \log T! + \log\log e,$$

and let ε be an arbitrary positive number. Then for almost all $\omega \in \Omega$ there exists a finite $N_0 = N_0(\omega)$ such that

$$\mu_T(N) \ge [h(N) - 2 - \varepsilon] \quad \text{if} \quad N \ge N_0,$$

moreover, there exists an infinite sequence $N_i = N_i(\omega)$ of integers such that

$$\mu_T(N_i) < [h(N_i) - 1 + \varepsilon].$$

The details of the proof are not presented in [21].

Földes [35] presented asymptotic results for the distribution of the number of T-contaminated head runs, the first hitting time of a T-contaminated head run having a fixed length, and the length of the longest T-contaminated head run. Móri [56] proved an almost sure limit theorem for the longest T-contaminated head run.

Gordon, Schilling, and Waterman [38] applied extreme value theory to obtain the asymptotic behaviour of the expectation and the variance of the length of the longest T-contaminated head run. Then accompanying distributions were obtained for the length of the longest T-contaminated head run. Novak [64] proved results on the accuracy of the approximation to the distribution of the length of the longest head run in a Markov chain.

In Chapter 3, we approximate the distribution of the length of the longest at most T-contaminated runs of heads. It is based on our paper [26]. We follow the lines of Arratia, Gordon, and Waterman [2], where Poisson approximation was used to find the asymptotic behaviour of the length of the longest at most T-contaminated head run. We use the basic results presented in [2], and give a new approximation for the distribution of the length of the longest at most T-contaminated head run. We show that for $T > 0$ the rate of the approximation in our new result is $\mathcal{O}\left(1/(\log(n))^2\right)$, where log denotes the logarithm to base $1/p$. Here and in what follows, $f(n) = \mathcal{O}(h(n))$ means that $f(n)/h(n)$ is bounded as $n \to \infty$. We see that for $T > 0$ the rate of the approximation offered by [2] is $\mathcal{O}(\log(\log(n))/\log(n))$, so our result considerably improves the former result. In our opinion the much better rate $\mathcal{O}(\log(n)/n)$ presented without detailed proof in [2] is just a misprint, that is true only for $T = 0$. The main result is Theorem 3.3.1. For completeness, we also give a proof of the former result, see Proposition 3.2.1. In Section 3.4, we present some simulation results as well, supporting our theorem.

Another approach to the topic is offered by a powerful lemma by Csáki, Komlós and Földes [16]. That approach was applied in [27] for the case of $T = 1$ and $T = 2$. In these particular cases, the results were the same as in [35], but for larger values of T the calculations became chaotic.

It is another interesting problem if there are two types of contaminations. This problem can be studied also by the lemma of Csáki, Komlós and Földes [16], see [28]. A similar topic is the problem of consecutive switches. In the novel papers [41] and [42] the length of the longest consecutive switches was considered. Their results are similar to the ones of [21] with $T = 0$.

Chapter 1

A continuous-time network evolution model describing N-interactions

Introduction

We study a continuous-time network evolution model based on cooperations of several units. The mathematical tool to describe an evolving network is the theory of random graphs. The graph's vertices are the nodes of the network and the edges of the graph are the connections among the nodes. In real life, a connection can be a collaboration or an interaction.

We intend to apply branching processes to describe a network evolution. There are several branching process-based random graph models in the literature of network theory. We present a short overview of the related works. Bollobás and Riordan in [11] considered several problems of random graphs and applied Galton-Watson branching processes to prove some of the theorems. Rudas, Tóth, and Valkó [68] as well as Rudas and Tóth [69] applied continuous-time branching processes to study random tree growth models. In [3], a preferential attachment random graph is embedded in a continuous-time branching process to find the asymptotic behaviour of the graph. In [36], continuous-time branching processes were applied to study an estimator in sublinear preferential attachment trees. In their survey paper [46], Holmgren and Janson studied the asymptotics of certain trees using Crump-Mode-Jagers branching processes. In [67], multi-type preferential attachment trees were studied using the theory of multi-type continuous-time branching processes. Banerjee and Huang [6] studied growing networks via embedding into a continuous-time branching process. In [48], Iksanov et al. underlined that one of the main applications of the theory of branching processes is the study of evolving networks. Usually, the advanced theory of branching processes is applied to prove new results for growing networks. However, sometimes network theory inspires new research in the field of branching processes, e.g., [59].

Most studies in network theory concentrate on connections of two nodes. However, cooperation of more than two units is also important. There are models describing such cooperation. Backhausz and Móri studied three-interactions in [4], while Fazekas and Porvázsnyik in [33] considered N-interactions.

In this chapter, we studied a continuous-time network evolution model based on the cooperation of several units. Our model was motivated by the activities of teams of people. We focused on cooperation among friends, recruitment of party members, recruitment of volunteers, collaboration among scientists, and teams in social networks. In the above cases, a person can be a member of several teams at the same time, new teams can emerge and disappear, and newcomers can join an existing team.

We applied similar mathematical tools as in [58], [23], where continuous-time branching processes were used (see also [34]). In [58], the starting network is a single edge. Any edge produces several new vertices during evolution, and then the edge disappears. Any new vertex produced by an edge is joined to the endpoints of the reproducing edge with one or two new edges. In [58], the reproduction process of the edges is a continuous-time branching process. In [23], 2- and 3-cliques are considered and some results of [58] are extended.

We study networks containing cliques (i.e., groups) of sizes $1, 2, \ldots, N$, where N is fixed. We shall apply multi-type branching processes, where the type is the size of a clique. The researches [17,67] also applied multi-type branching processes, but there the vertices had different types. We also mention, that the evolution process considered here is similar to the one studied in [23], but we emphasize that the process in [23] is not a particular case of the process in this work.

During the evolution of the network, when a newcomer joins the network, it joins directly with certain members of an existing clique in the network. Then they will form a new clique which we shall consider as a child of the old clique. Any clique produces an offspring clique each time its driving Poisson process jumps. Then these offspring start their reproduction processes. The length of life of a clique depends on the number of offspring it has.

This chapter is organized as follows. In Section 1.1, we define our model precisely. In Section 1.2, we present our general results. We calculate the survival function of a clique (Theorem 1.2.1), the mean offspring number of a clique (Corollary 1.2.1), the Laplace transforms (Proposition 1.2.1), and then we turn to the Perron root and the Malthusian parameter. We do not obtain an explicit expression for the Malthusian parameter, so we can numerically compute it. In Section 1.3 we give a brief summary of the general multi-type branching processes based on [49]. In Section 1.4, we prove asymptotic theorems for the number of cliques of a given size (Theorem 1.4.1). Each has magnitude $e^{\alpha t}$ in the event of non-extinction, where α is the Malthusian parameter. The proof of Theorem 1.4.1 is based on the asymptotic theorems of [49]. A most interesting question in the theory of evolving networks is the degree process of a fixed vertex. Therefore, in Section 1.5, we prove asymptotic theorems for the degree of a fixed vertex.

For the proof, we introduce a new branching process, called the "good children" process, and then we can apply the asymptotic theorems of [49]. We obtain that the magnitude of the degree of a fixed vertex is $e^{\hat{\alpha}t}$, where $\hat{\alpha}$ is the Malthusian parameter of the "good children" process. In Section 1.6, we obtain the generating functions. We apply the generating functions to find the probability of extinction of the network, see Theorem 1.6.1. In Section 1.7, we show simulation results. We shall see that the simulation results support our theorems.

We mention that basic results on branching processes are contained e.g., in [50], [39], [55], [60], and [61]. In Section1.3 we also give a brief summary of the general multi-type branching processes based on [49].

1.1 The model

We consider the following model to describe certain evolving networks. The main ingredients of the network are teams. Any team is considered as a clique, i.e., a graph having n vertices so that any two vertices are connected with one edge. The size of a clique is the number of its vertices. In our model, there are cliques of sizes $1, 2, \ldots, N$, where N is an arbitrarily large but fixed number. We shall describe the evolution of the network by a multi-type branching process. In terms of multi-type branching processes, see, e.g., [55], an n-clique would be a type-n individual.

At the time $t = 0$ there is one team, and the size of this team can be $1, 2, \ldots, N$. In terms of branching processes, we call this team the ancestor. This ancestor team produces offspring teams which can be cliques of sizes $1, 2, \ldots, N$. Then these children teams also produce their own children teams, and so on. Any team has its rate of 1 Poisson process. The jumping times of this Poisson process are the reproduction times of the team. We shall suppose that the reproduction processes of different teams are independent. The reproduction processes of the teams of size n are independent copies of the reproduction process of the generic team of size n.

Now, we identify any team with the clique representing it. The mathematical description of the evolution of the generic n-clique is the following. Let $\Pi_n(t)$ denote the Poisson process with parameter 1 corresponding to the generic n-clique. The jumping times of $\Pi_n(t)$ are the reproduction times. When $\Pi_n(t)$ jumps, then a new vertex appears and we connect it to certain vertices of the generic n-clique. The new vertex will be connected to j vertices of the generic n-clique with probability $q_{n,j}$, where $0 \leq q_{n,j} \leq 1$, $j = 0, 1, \ldots, n$, and $\sum_{j=0}^{n} q_{n,j} = 1$. (We assume that $q_{N,N} = 0$ because the largest team is of size N.) When j is chosen, then together with the new vertex, j new edges appear. One endpoint of any new edge is the new vertex. The other j endpoints of the j new edges are chosen randomly from the vertices of the generic n-clique. The number of subsets having j elements is $\binom{n}{j}$. We choose one of these $\binom{n}{j}$ subsets uniformly at random. The vertices of this subset are already connected, and now we connect each of them to the new vertex with one edge. So the j old connected vertices chosen, the new vertex, and the j new edges form a $(j + 1)$-clique. This new $(j + 1)$-clique is a child of the generic n-clique and it is the only child at this step. We should emphasize that the result of the above step is precisely one child because we shall count only this one child. The reason for this point of view is that we are interested in the number of teams, and the new sub-cliques of the new $(j+1)$-clique are not considered as teams.

We see that the generic n-clique produces precisely one child clique at any birth time. If $j = 0$, this child is just one vertex, i.e., the new vertex joining the network with 0 edges. If $j = 1$, then this child is an edge, which consists of the new vertex and the old vertex connected to the new one. If $j = 2$, this child is a triangle consisting of the new vertex and two vertices of the generic n-clique. If $j = n < N$, this child is an $(n+1)$-clique consisting of the whole n-clique ancestor and the new vertex. We emphasize that when the parent is the largest possible clique, i.e., an N-clique, we do not allow the birth of an $(N+1)$-clique. We also see that the probability that an i-type ancestor produces a j-type child is $p_{i,j} = q_{i,j-1}$, $j = 1, 2, \ldots, i+1$. The ancestor clique, the children cliques of the ancestor, and the grandchildren cliques, etc., will form an evolving network.

Example 1.1.1. *We visualize four simple reproduction steps in Figures 1.1–1.4.*

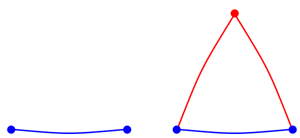


Figure 1.1: The parent is a 2-clique, and the child is a 3-clique.

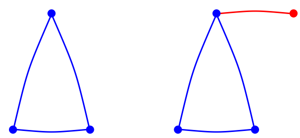


Figure 1.2: The parent is a 3-clique, and the child is a 2-clique.

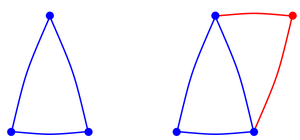


Figure 1.3: The parent is a 3-clique, and the child is a 3-clique.

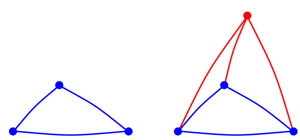


Figure 1.4: The parent is a 3-clique, and the child is a 4-clique.

In our model, cliques can die. When a clique dies, it will be an inactive clique not producing children. But we do not delete its vertices and edges, because any vertex or edge can belong to several cliques. So we shall study both the number of all j-cliques born during the evolution process and the number of active j-cliques. We underline again that we are interested in the number of teams. The teams (i.e., cliques) are born during the above evolution process, i.e., the offspring of the ancestor. So we will not count those cliques that are not offspring of the ancestor clique.

Let us denote by $\xi_{i,j}(t)$ the number of type-j children cliques of the type-i generic clique up to time t $(i, j = 1, 2, \ldots, N)$. The processes $\xi_{i,j}$ are point processes. Then

$$\xi_i(t) = \sum_{j=1}^{i+1} \xi_{i,j}(t) \tag{1.1.1}$$

is the number of all children of the generic i-clique up to time t.

Let $\tau_i(1), \tau_i(2), \ldots$ be the birth times of the generic i-clique and denote by $\varepsilon_i(1), \varepsilon_i(2), \ldots$ the corresponding total litter sizes. In our model, $\varepsilon_i(k) \equiv 1$.

Let λ_i be the life length of the generic i-clique. λ_i is a random variable with $\mathbb{P}(0 \leq \lambda_i < \infty) = 1$. After its death, a clique does not produce offspring, so $\xi_i(t) = \xi_i(\lambda_i)$, when $t > \lambda_i$. So

$$\xi_i(t) = \sum_{\tau_i(k) \leq t \wedge \lambda_i} \varepsilon_i(k) = \Pi_i(t \wedge \lambda_i), \tag{1.1.2}$$

where $\Pi_i(t)$ is the Poisson process, and $x \wedge y$ is the minimum of $\{x, y\}$.

Now we turn to the survival function of the life length of an i-clique. $L_i(t)$ will denote the distribution function of λ_i. Then the survival function of λ_i is

$$1 - L_i(t) = \mathbb{P}(\lambda_i > t) = \exp\left(-\int_0^t l_i(u)\, du\right). \tag{1.1.3}$$

Here $l_i(t)$ denotes the hazard rate function of λ_i.

A major assumption of this chapter is that the hazard rate is determined by the total number of offspring in the following way:

$$l_i(t) = b + c\xi_i(t), \tag{1.1.4}$$

where $b \geq 0$ and $c > 0$ are fixed constants. It means that an individual having a lot of children dies with high probability.

We emphasize that the reproduction processes of the i-cliques are independent copies of the reproduction process of the generic i-clique.

1.2 Preliminary results

First, we calculate the survival function of an i-clique.

Theorem 1.2.1. *For any i, the survival function of an i-clique is*

$$1 - L_i(t) = \mathbb{P}(\lambda_i > t) = e^{-(b+1)t} e^{\frac{1-e^{-ct}}{c}}. \tag{1.2.1}$$

Proof. Using the general calculation of Theorem 1 in [23], we have

$$\mathbb{P}\left(\lambda_i > t\right) = e^{-(b+1)t} e^{\sum_{j=1}^{\infty} s_j \frac{1-e^{-ctj}}{cj}},$$

where s_j denotes the distribution of ε_i. As now $\varepsilon_i \equiv 1$, so $s_1 = 1$, and therefore the survival function for an i-clique is

$$\mathbb{P}\left(\lambda_i > t\right) = e^{-(b+1)t} e^{\frac{1-e^{-ct}}{c}}.$$

□

Now, we turn to the mean offspring number. It is $m_{i,j}(t) = \mathbb{E}\xi_{i,j}(t)$, which is the expectation of the number of type-j offspring of a type-i parent up to time t.

Corollary 1.2.1. *For any $t \geq 0$, we have*

$$m_{i,j}\left(t\right) = p_{i,j} F(t), \tag{1.2.2}$$

where

$$F(t) = \int_0^t \left(1 - L_i\left(s\right)\right) ds = \int_0^t e^{-(b+1)s} e^{\frac{1-e^{-cs}}{c}} ds = \frac{1}{c}\int_0^{1-e^{-ct}} \left(1-u\right)^{\frac{b+1}{c}-1} e^{\frac{u}{c}} du,$$

$0 < F(\infty) < \infty$.

$$\mathbb{E}\lambda_i = \frac{1}{c}\int_0^1 \left(1-u\right)^{\frac{b+1}{c}-1} e^{\frac{u}{c}} du. \tag{1.2.3}$$

Proof. Similar to the proof of Corollary 1 in [23], we have

$$m_{i,j}\left(t\right) = \mathbb{E}\xi_{i,j}\left(t\right) = \mathbb{E}\left(\varepsilon_{i,j}(1) + \varepsilon_{i,j}(2) + \cdots + \varepsilon_{i,j}\left(\Pi\left(t \wedge \lambda_i\right)\right)\right),$$

where $\varepsilon_{i,j}(k)$ is the number of type-j children of a type-i mother at her k^{th} birth event. Applying Wald's identity, we get

$$m_{i,j}\left(t\right) = \mathbb{E}\left(\varepsilon_{i,j}(1)\right) \mathbb{E}\left(\Pi\left(t \wedge \lambda_i\right)\right). \tag{1.2.4}$$

Π is a rate-1 Poisson process, and $t \wedge \lambda$ is bounded, so (1.2.4) implies that

$$m_{i,j}\left(t\right) = \mathbb{E}\left(\varepsilon_{i,j}(1)\right) \mathbb{E}\left(t \wedge \lambda_i\right) = \mathbb{E}\left(\varepsilon_{i,j}(1)\right) \int_0^t \left(1 - L_i\left(x\right)\right) dx. \tag{1.2.5}$$

In our case $\mathbb{P}\left(\varepsilon_{i,j}(k) = 1\right) = p_{i,j}$, so from (1.2.1) and inserting $u = 1 - e^{-cs}$, we get

$$m_{i,j}\left(t\right) = p_{i,j}\int_0^t e^{-(b+1)s} e^{\frac{1-e^{-cs}}{c}} ds = \frac{p_{i,j}}{c}\int_0^{1-e^{-ct}} \left(1-u\right)^{\frac{b+1}{c}-1} e^{\frac{u}{c}} du. \tag{1.2.6}$$

□

We need the Laplace transform of $m_{i,j}$,

$$m^*_{i,j}(\kappa) = \int_0^\infty e^{-\kappa s} m_{i,j}(ds), \qquad i,j = 1,2,\ldots,N.$$

Proposition 1.2.1. *We have*

$$m^*_{i,j}(\kappa) = p_{i,j} A(\kappa), \quad \kappa \geq 0, \tag{1.2.7}$$

where

$$A(\kappa) = \int_0^\infty e^{-\kappa t} e^{-(b+1)t} e^{\frac{1-e^{-ct}}{c}} dt = \frac{1}{c} \int_0^1 (1-u)^{\frac{\kappa+b+1}{c}-1} e^{\frac{u}{c}} du. \tag{1.2.8}$$

$m_{i,j}(\infty) = p_{i,j} A(0)$. *If* $A(0) > 1$, *then* $A(\alpha) = 1$ *for some* $\alpha > 0$.

Proof. For the proof, one can use Corollary 1.2.1. □

Now, we shall consider the Perron root. The Laplace transform matrix is

$$M(\kappa) = \left(m^*_{i,j}(\kappa) \right)_{i,j=1}^N. \tag{1.2.9}$$

The characteristic roots of $M(\kappa)$ are denoted by $\varrho_l(\kappa)$, $l = 1,\ldots,N$. The Perron root is the greatest of the values $\varrho_l(\kappa)$. We denote it by $\varrho(\kappa)$.

We suppose that

$$\varrho(0) > 1, \tag{1.2.10}$$

that is, we assume that our process is supercritical. A supercritical process with positive probability will not go extinct.

We shall need the Malthusian parameter. α is called the Malthusian parameter if $\varrho(\alpha) = 1$.

If α is the Malthusian parameter, then we denote by $\mathbf{v} = (v_1,\ldots,v_N)^\top$ the right eigenvector of $M(\alpha)$ corresponding to eigenvalue 1 and normalized as $v_1 + \cdots + v_N = 1$. Let $\mathbf{u} = (u_1,\ldots,u_N)^\top$ be the left eigenvector of $M(\alpha)$ satisfying $u_1 v_1 + \cdots + u_N v_N = 1$.

We shall suppose that the stochastic matrix $(p_{i,j})_{i,j=1}^N$ is irreducible and acyclic. Then its Perron root is 1, so $\varrho(\kappa) = A(\kappa)$. Therefore, our branching process is supercritical if and only if $A(0) > 1$. If $A(0) > 1$, then there exists a Malthusian parameter, that is, an $\alpha > 0$ such that $A(\alpha) = 1$. We also see that $\mathbf{v} = (1/N,\ldots,1/N)^\top$.

1.3 A brief summary of the general multi-type branching processes

To prove our theorems about the limiting behaviour our model, we shall use some known results of the theory of continuous-time branching processes. The single type general Crump-Mode-Jagers branching processes are discussed e.g. in [50], [39] and [61]. The general multi-type branching processes are described e.g. in [55], [60] and [49].

In this section we give a brief summary of the general multi-type branching processes based on [49]. In such a process, the individuals can be of p different types, that are denoted by $1, 2, \ldots, p$. Any individual x is described by the quantities $\lambda_x, \xi_x, \Phi_x, \Psi_x, \ldots$. The quantities $\lambda_x, \xi_x, \Phi_x, \Psi_x, \ldots$ belonging to the indivuidual x are independent copies of the quantities $\lambda, \xi, \Phi, \Psi, \ldots$. Hence we should define $\lambda, \xi, \Phi, \Psi, \ldots$ that we consider as the quantities corresponding to the generic individual.

The random variable λ represents the lifetime. It is non-negative and not necessarily independent from the reproduction. The lifetime distribution is $L(t) = \mathbb{P}(\lambda \leq t)$. The reproduction process is $\xi_i(t) = (\xi_{i,1}(t), \ldots, \xi_{i,p}(t))$, $t \geq 0$. The random point process $\xi_{i,j}$ describes the births of type j offspring of a type i mother. $\xi_{i,j}(t)$ gives the number of type j offspring of a type i mother up to time t. $\xi_{i,j}$ is determined by the birth events and the numbers of offspring. The process starts at time $t = 0$ with one individual that is referred to as the ancestor and is denoted by x_0. When a child is born, then it also starts its own reproduction process, and so on. We use the notation σ_x for the birth time of the individual x.

Let $\Phi(t)$ be a non-negative random function which describes a certain aspect of the life history of the individual. We usually assume that $\Phi(t) = 0$ for $t \leq 0$ and then $\Phi(t)$ is called a random characteristic. Let $\Psi(t)$ denote another random characteristic. Thus, the behaviour of the individual x is described by the quantitites $\xi_x, \lambda_x, \Phi_x, \Psi_x, \ldots$.

We define the branching process ${}_{x_0}Z^{\Phi}(t)$ counted by the characteristic Φ as

$$ {}_{x_0}Z^{\Phi}(t) = \sum_x \Phi_x\left(t - {}_{x_0}\sigma_x\right), $$

where the sum is taken over all individuals x. The left subscript x_0 of Z and of the birth time σ_x is important, because it indicates that the process starts with ancestor x_0 and the type of x_0 influences the evolution of the population.

Let us use the notation $m_{i,j}(t)$ for the reproduction function which is defined as the expected reproduction number $m_{i,j}(t) = \mathbb{E}\xi_{i,j}(t)$.

The following statements are well-known and can be found in [49] or [60].

Throughout the chapter, we suppose that the following basic conditions hold:

(a) Not all of the measures $m_{i,j}$ are concentrated on a lattice. Let

$$ m^*_{i,j}(\kappa) = \int_0^\infty e^{-\kappa t} m_{i,j}(dt), \qquad i, j = 1, \ldots, p, $$

be the Laplace transform of $m_{i,j}$. Let $M(\kappa)$ be the matrix

$$ M(\kappa) = \left(m^*_{i,j}(\kappa)\right)_{i,j=1}^{p}. $$

(b1) There exists a positive Malthusian parameter α, which is a finite positive value such that $M(\alpha)$ has finite entries only and the Perron-Frobenius root of

$M(\alpha)$ is equal to 1. Here the Perron-Frobenius root is the largest eigenvalue of the matrix. Let $(v_1, \ldots, v_p)^\top$ be the right positive eigenvector and $(u_1, \ldots, u_p)^\top$ the left positive eigenvector of $M(\alpha)$ corresponding to the Perron-Frobenius root. These vectors are normalized so that $\sum_{i=1}^p v_i = 1$ and $\sum_{i=1}^p u_i v_i = 1$.

(b2) The matrix $(m_{i,j}(\infty))_{i,j=1}^p$ has an infinite entry, or all of its entries are finite and its Perron-Frobenius root is greater than 1.

(c) The first moment of $e^{-\alpha t} m_{i,j}(dt)$ is finite and positive, that is

$$0 < \int_0^\infty t e^{-\alpha t} m_{i,j}(dt) < \infty, \qquad i, j = 1, \ldots, p.$$

(d) There exists a finite positive integer K such that all elements of the Kth power of the matrix $(m_{i,j}(\infty))_{i,j=1}^p$ are positive. Let

$$_\alpha\xi_{i,j}(\infty) = \int_0^\infty e^{-\alpha t} \xi_{i,j}(dt). \tag{1.3.1}$$

Proposition 1.3.1. *Let α be the Malthusian parameter. We suppose that the random characteristic Φ satisfies the following conditions:*

(i) $\Phi(t) \geq 0$,

(ii) the trajectories of Φ belong to the Skorohod space D, i.e. they do not have discontinuities of the second kind,

(iii) $\mathbb{E}(\sup_t \Phi(t)) < \infty$.

Assume also

(iv) for some $\varepsilon > 0$

$$\int_0^\infty t(\log(1+t))^{1+\varepsilon} e^{-\alpha t} m_{i,j}(dt) < \infty, \qquad i, j = 1, \ldots, p$$

and

(v) for some $\varepsilon > 0$

$$\mathbb{E} \sup_{t \geq 0} \left\{ \max \left\{ t(\log(1+t))^{1+\varepsilon}, 1 \right\} e^{-\alpha t} \Phi(t) \right\} < \infty$$

for any ancestor.

Then

$$\lim_{t \to \infty} e^{-\alpha t} {}_{x_0}Z^\Phi(t) = {}_{x_0}Y_\infty v_i m_\infty^\Phi \tag{1.3.2}$$

almost surely, where i is the type of x_0,

$$m_{\infty}^{\Phi} = \frac{\sum_{j=1}^{p} u_j \int_0^{\infty} e^{-\alpha t}\mathbb{E}\Phi_j\left(t\right) dt}{\sum_{l,j=1}^{p} u_l v_j \int_0^{\infty} t e^{-\alpha t} m_{l,j}\left(dt\right)}, \tag{1.3.3}$$

${}_{x_0}Y_{\infty}$ is an a.s. non-negative random variable that depends on the type of the ancestor x_0 but does not depend on the choice of Φ.

If we also suppose that

(vi)

$$\mathbb{E}\left[{}_{\alpha}\xi_{i,j}(\infty)\log^{+}{}_{\alpha}\xi_{i,j}(\infty)\right] < \infty, \qquad i,j = 1,\ldots,p, \tag{1.3.4}$$

then $\mathbb{E}\left({}_{x_0}Y_{\infty}\right) = 1$, ${}_{x_0}Y_{\infty}$ is positive with positive probability, and ${}_{x_0}Y_{\infty}$ is a.s. positive on the survival set.

This proposition is a simple consequence of Theorem 2.4 and Proposition 4.1 of [49].

1.4 The limiting behaviour of the number of cliques

For any evolving network, a basic question is the growth of the number of ingredients of the network. So, for our network, we should find the number of cliques. For the proof, we shall use powerful results on multi-type branching processes that were summarized in Proposition 1.3.1 of Section 1.3. So here we shall check conditions (a), (b1), (b2), (c), (d), and (*i*)–(*vi*) of Section 1.3.

We shall assume that the matrix $(p_{i,j})$ is irreducible and acyclic. Therefore, Theorem 1.4 of [70] gives that there exists a positive integer K, such that each element of the K^{th} power of the matrix $(p_{i,j})$ is positive. As $m_{i,j}(\infty) = p_{i,j}A(0)$ and $A(0) > 0$, condition (d) will be satisfied.

For condition (a), we have to show that not all measures $m_{i,j}$ are concentrated on a lattice. By Corollary 1.2.1, these measures are absolutely continuous, so condition (a) is fulfilled.

For (b2), we shall assume that $A(0) > 1$. As $m_{i,j}(\infty) = p_{i,j}A(0)$, it will imply (b2). Concerning condition (b1), we mention that $A(0) > 1$ implies that there exists a Malthusian parameter, that is, an $\alpha > 0$ such that $A(\alpha) = 1$. We can numerically calculate the value of α.

We shall check condition (c) during the proof of Theorem 1.4.1.

Now, we shall find the denominator in the limit theorem. We can see that the denominator of m_{∞}^{Φ} in the asymptotic expression does not depend on Φ and it has the form

$$D(\alpha) = \sum_{l,j=1}^{N} u_l v_j \int_0^{\infty} s e^{-\alpha s} m_{l,j}\left(ds\right).$$

It is the same as

$$D(\alpha) = \sum_{l,j=1}^{N} u_l v_j \left(-m_{l,j}^{*}(\alpha)\right)'. \tag{1.4.1}$$

Here u_i and v_i are the coordinates of the eigenvectors and we know that $\mathbf{v} = (1/N, \ldots, 1/N)^\top$. Moreover, by Proposition 1.2.1, we can see that

$$\left(-m^*_{l,j}(\alpha)\right)' = p_{l,j}\left(-A'(\alpha)\right). \tag{1.4.2}$$

So

$$D(\alpha) = -A'(\alpha) \sum_{l,j=1}^{N} u_l v_j p_{l,j} = -A'(\alpha). \tag{1.4.3}$$

Here

$$-A'(\alpha) = \int_0^\infty t e^{-\alpha t} e^{-(b+1)t} e^{\frac{1-e^{-ct}}{c}} dt = -\frac{1}{c^2}\int_0^1 \ln(1-x)\,(1-x)^{\frac{\alpha+b+1}{c}-1} e^{\frac{x}{c}} dx. \tag{1.4.4}$$

Now, we shall consider the number of the n-cliques.

Theorem 1.4.1. *Assume that the matrix $(p_{i,j})_{i,j=1}^N$ is irreducible and acyclic. Assume that $A(0) > 1$. Let α be the Malthusian parameter, i.e., a finite positive solution of equation $A(\alpha) = 1$. Let n be fixed, $1 \le n \le N$. We denote by ${}_kT(t)$ the number of all n-cliques being born up to time t if the ancestor of the network was a k-clique, $k = 1, 2, \ldots, N$. Then*

$$\lim_{t\to\infty} e^{-\alpha t}{}_kT(t) = {}_kW \frac{v_k u_n}{\alpha\left(-A'(\alpha)\right)} \tag{1.4.5}$$

almost surely for $k = 1, 2, \ldots, N$.

We denote by ${}_k\hat{T}(t)$ the number of all n-cliques alive at time t if the ancestor of the network was a k-clique, $k = 1, 2, \ldots, N$. Then

$$\lim_{t\to\infty} e^{-\alpha t}{}_k\hat{T}(t) = {}_kW \frac{v_k u_n A(\alpha)}{\left(-A'(\alpha)\right)} \tag{1.4.6}$$

almost surely for $k = 1, 2, \ldots, N$.

The quantity ${}_kW$ is a.s. non-negative, $\mathbb{E}({}_kW) = 1$, and ${}_kW$ is a.s. positive on the event of survival.

Proof. We shall use Proposition 1.3.1 for the proof. To prove condition (vi), it is enough to show that

$$\mathbb{E}\left[{}_\alpha\xi_i(\infty)\log^+ {}_\alpha\xi_i(\infty)\right] < \infty, \qquad i = 1, 2, \ldots, N, \tag{1.4.7}$$

where

$${}_\alpha\xi_i(\infty) = \int_0^\infty e^{-\alpha t}\xi_i\,(dt)\,, \qquad i = 1, 2, \ldots, N, \tag{1.4.8}$$

and

$$\xi_i(t) = \xi_{i,1}(t) + \xi_{i,2}(t) + \cdots + \xi_{i,i+1}(t), \qquad i = 1, 2, \ldots, N. \tag{1.4.9}$$

At any birth, there is precisely 1 child, so

$$_\alpha\xi_i(\infty) = \int_0^\infty e^{-\alpha t}\xi_i\,(dt) = \sum_{\tau(j)\le\lambda_i} 1e^{-\alpha\tau(j)} \le \sum_{j=1}^{\infty} 1e^{-\alpha\tau(j)} = M,$$

where $\tau(1), \tau(2), \ldots$ are the jumping times of the Poisson process Π_i. The interarrival time $(\tau(j) - \tau(j-1))$ is exponentially distributed with rate 1, so $\tau(j)$ has Γ-distribution $\Gamma\,(j, 1)$. It implies that

$$\mathbb{E}(M) = \sum_{j=1}^{\infty} \mathbb{E}\left(e^{-\alpha\tau(j)}\right) = \sum_{j=1}^{\infty} \frac{1}{(1+\alpha)^j} = \frac{1}{\alpha}. \tag{1.4.10}$$

Let $\eta_j = \tau(j) - \tau(j-1)$. Let η_0 be a random variable having an exponential distribution with parameter 1 and assume that η_0 and M are independent. Then

$$e^{-\alpha\eta_0}\,(1+M) = e^{-\alpha\eta_0} + e^{-\alpha\eta_0}\sum_{j=1}^{\infty} e^{-\alpha(\eta_1+\cdots+\eta_j)} = \sum_{j=0}^{\infty} e^{-\alpha(\eta_0+\eta_1+\cdots+\eta_j)}.$$

So $e^{-\alpha\eta_0}\,(1+M)$ has the same distribution as that of M. From this and Eq (1.4.10), we get

$$\mathbb{E}M^2 = \mathbb{E}\left(e^{-\alpha\eta_0}\,(1+M)\right)^2 = \frac{1}{1+2\alpha}\left(1 + \frac{2}{\alpha} + \mathbb{E}M^2\right).$$

So we obtain

$$\mathbb{E}M^2 = \frac{\alpha+2}{2\alpha^2} < \infty.$$

Therefore (1.4.7) is true for any i.

To prove conditions (c) and (iv), it is enough to show that $\int_0^\infty t^2 e^{-\alpha t} m_{i,j}(dt) < \infty$, for $i, j = 1, 2, \ldots, N$. Now, from Corollary 1.2.1, we get that

$$\int_0^\infty s^2 e^{-\alpha s} m_{i,j}(ds) \le \int_0^\infty s^2 e^{-\alpha s} e^{-s(b+1)} e^{\frac{1-e^{-cs}}{c}}\,ds \le \int_0^\infty s^2 e^{-s(\alpha+b+1-1)} ds < \infty$$

as $\alpha + b$ is positive. Therefore conditions (c) and (iv) are true.

Now, consider the number of n-cliques. To show (1.4.5), let $\Phi_x(t) = 1$ if x is an n-clique, and $\Phi_x(t) = 0$ otherwise. Therefore $\mathbb{E}\Phi_n(t) = 1$ and $\mathbb{E}\Phi_j(t) = 0$ for $j \ne n$. So assumptions $(i$–$iii)$ and (v) are true. Therefore, Proposition 1.3.1 implies (1.4.5).

To prove (1.4.6), let $\Phi_x(t) = 1$ if x is an n-clique and it is alive at time t, and let $\Phi_x(t) = 0$ otherwise. Therefore $\mathbb{E}\Phi_n(t) = 1 - L_n(t)$ and $\mathbb{E}\Phi_j(t) = 0$ for $j \ne n$. So $(i$–$iii)$ and (v) are fulfilled. We see that

$$\int_0^\infty e^{-\alpha s}\mathbb{E}\Phi_n\,(s)\,ds = \int_0^\infty e^{-\alpha s}(1 - L_n(s))ds = A(\alpha).$$

Using Proposition 1.3.1, we get (1.4.6). □

Remark 1.4.1. *The process is supercritical if*

$$1 < \varrho(0) = A(0) = \frac{1}{c} \int_0^1 (1-u)^{\frac{b+1}{c}-1} e^{\frac{u}{c}} du.$$

If the process is supercritical, then there exists a Malthusian parameter.

For any pair b, c, one can show whether the process is supercritical. Here we just list a few cases.

If $1 \le c < 1/\ln 2$ and $0 \le b \le c-1$, then the process corresponding to b, c is supercritical. Moreover the case $b=0$, $c=1/2$ is also supercritical.

If $b \ge 1$, $c=1$, then the process is not supercritical. The case $b=1$, $c=2$ is also not supercritical.

To consider the other values, we have made a numerical investigation on a rectangle: We have considered the cases when $0 \le b \le 2$ and $0.01 \le c \le 2$, because the parameter c is in the denominator. Figure 1.5 shows the set of parameters (b, c) for which the process is supercritical.

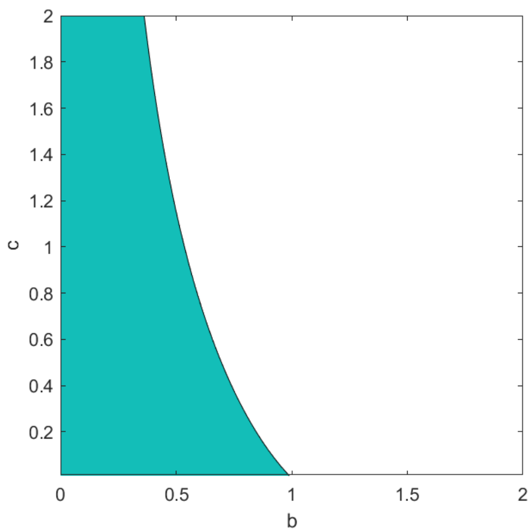


Figure 1.5: The region where the process is supercritical on the set $[0, 2] \times [0.01, 2]$.

Example 1.4.1. *Consider the Leslie model. For $n = 1, 2, \ldots, N-1$, when a new vertex joins to an n-clique, then either it joins to all vertices of the n-clique, or the new vertex alone creates a new clique. So $p_{n,n+1} = p_n$, $p_{n,1} = 1 - p_n$ for $n = 1, 2, \ldots, N-1$. But for $n = N$, when a new vertex appears, then either it creates alone a new clique or the new vertex and $N-1$ old vertices create an N-clique. So $p_{N,N} = p_N$ and $p_{N,1} = 1 - p_N$.*

Now, the matrix of the Laplace transforms is

$$M(\kappa) = A(\kappa) \begin{pmatrix} 1-p_1 & p_1 & 0 & \dots & 0 \\ 1-p_2 & 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 1-p_{N-1} & 0 & 0 & \dots & p_{N-1} \\ 1-p_N & 0 & 0 & \dots & p_N \end{pmatrix}.$$

We assume that $p_1 > 0, \dots, p_{N-1} > 0$, *and* $1 - p_N > 0$. *We also assume that the greatest common divisor of the set* $\{i = 1, 2, \dots, N : 1 - p_i > 0\}$ *is equal to* 1. *So the matrix* $(p_{i,j})$ *is irreducible and acyclic. Therefore its Perron root is* 1. *So the Perron root of* $M(\kappa)$ *is* $A(\kappa)$, *and the corresponding right eigenvector is*

$$\mathbf{v} = \left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right)^\top.$$

Direct calculations show that the coordinates of the left eigenvector are

$$u_N = \frac{N p_1 p_2 \cdots p_{N-1}}{(1-p_N)(1 + p_1 + p_1 p_2 + \dots + p_1 p_2 \cdots p_{N-2}) + p_1 p_2 \cdots p_{N-1}},$$

and

$$u_{N-1} = \frac{1-p_N}{p_{N-1}} u_N, \quad u_{N-2} = \frac{1-p_N}{p_{N-1} p_{N-2}} u_N, \ \dots, \ u_1 = \frac{1-p_N}{p_{N-1} p_{N-2} \cdots p_1} u_N.$$

Now, Theorem 1.4.1 gives the asymptotic number of cliques.

Consider the particular case of the Leslie model when $p_i = a$ *for any* i, *where* $0 < a < 1$. *Then* $u_N = N a^{N-1}$, *and* $u_n = N(1-a)a^{n-1}$, $n = 1, 2, \dots, N-1$. *Inserting these values into the formulae* (1.4.5) *and* (1.4.6), *we obtain the asymptotic number of cliques. For* n*-cliques with* $n = 1, 2, \dots, N-1$,

$$\lim_{t\to\infty} e^{-\alpha t} {}_kT(t) = {}_kW(1-a)a^{n-1} \cdot \frac{1}{\alpha(-A'(\alpha))}, \tag{1.4.11}$$

but for the N*-cliques,*

$$\lim_{t\to\infty} e^{-\alpha t} {}_kT(t) = {}_kW a^{N-1} \cdot \frac{1}{\alpha(-A'(\alpha))} \tag{1.4.12}$$

almost surely for $k = 1, 2, \dots, N$.

Similarly, for n*-cliques with* $n = 1, 2, \dots, N-1$,

$$\lim_{t\to\infty} e^{-\alpha t} {}_k\hat{T}(t) = {}_kW(1-a)a^{n-1} \cdot \frac{A(\alpha)}{(-A'(\alpha))} \tag{1.4.13}$$

and for the N*-cliques,*

$$\lim_{t\to\infty} e^{-\alpha t} {}_k\hat{T}(t) = {}_kW a^{N-1} \cdot \frac{A(\alpha)}{(-A'(\alpha))}. \tag{1.4.14}$$

We see that as $a \to 1$, *then* N*-cliques dominate. It is a plausible consequence of the definition of the model.*

Another particular case of the Leslie model is $p_1 = p_N = 1/2$ *and* $p_i = 1$ *for* $i = 2, \dots, N-1$. *Then* $u_1 = u_N = (2N)/(N+2)$, *and* $u_i = u_N/2$ *for* $i = 2, \dots, N-1$.

1.5 The connected network and the degree of a fixed vertex

A basic question for any evolving network is the degree process of a fixed vertex. We study the topics of the connected network and the degree of a fixed vertex in the same chapter because both of them are studied by multi-type branching processes having $N-1$ types.

Remark 1.5.1. *When the network is connected then it is not possible that a separated vertex is born. In that case, we can consider just the* $2, 3, \dots, N$ *cliques. So the possible types of our branching process will be* $2, 3, \dots, N$*, where the type is* 2 *if the clique is an edge, it is* 3 *if the clique is a triangle,* $\dots$*, and it is* N *in the case of an* N*-clique. The probability that the new vertex will be connected to* j *vertices of the generic* n*-clique is again* $q_{n,j}$*, where* $0 \le q_{n,j} \le 1$*, but now* $j = 1, \dots, n$ *and* $\sum_{j=1}^{n} q_{n,j} = 1$*. So we should consider a multi-type branching process having* $N-1$ *types. Now, the probability that a type-*i *parent gives birth to a type-*j *child is* $p_{i,j} = q_{i,j-1}$ *for* $i, j = 2, 3, \dots, N$*.*

The results of the previous two sections remain valid with obvious modifications. So $m_{i,j}$*,* $m^*_{i,j}$ *are the same, but* $i, j = 2, 3, \dots, N$*.*

We shall suppose that the stochastic matrix $(p_{i,j})_{i,j=2}^{N}$ *is irreducible and acyclic. Then its Perron root is* 1*, so* $\varrho(\kappa) = A(\kappa)$*. Therefore, our branching process is supercritical if and only if* $A(0) > 1$*. If* $A(0) > 1$*, then there exists a Malthusian parameter, that is, an* $\alpha > 0$ *such that* $A(\alpha) = 1$*.* $\mathbf{v} = (1/(N-1), \dots, 1/(N-1))^\top$ *is the right eigenvector of* $(p_{i,j})_{i,j=2}^{N}$ *corresponding to eigenvalue* 1*.* $\mathbf{u} = (u_2, \dots, u_N)^\top$ *is the left eigenvector of* $(p_{i,j})_{i,j=2}^{N}$ *satisfying the condition* $u_2 v_2 + \dots + u_N v_N = 1$*.*

We can modify Theorem 1.4.1 to find the number of the n*-cliques in the case of the connected network. Assume that the matrix* $(p_{i,j})_{i,j=2}^{N}$ *is irreducible and acyclic. Assume that* $A(0) > 1$*. Let* α *be the Malthusian parameter. Let* n *be fixed,* $2 \le n \le N$*. Denote by* ${}_kT(t)$ *the number of all* n*-cliques being born up to time* t *if the ancestor of the population was a* k*-clique,* $k = 2, 3, \dots, N$*. Then*

$$\lim_{t\to\infty} e^{-\alpha t}\, {}_kT(t) = {}_kW \frac{v_k u_n}{\alpha\left(-A'(\alpha)\right)} \tag{1.5.1}$$

almost surely for $k = 2, 3, \dots, N$*.*

This is similar for the number of n*-cliques alive at time* t*.*

Now, we turn to the degree process of a fixed vertex. Let V be a fixed vertex. Assume that at the beginning, V is a member of an i-clique. We assume that $i \ge 2$, so the initial degree of V is $i - 1 \ge 1$. We call a child a "good child" if it contains V. So the degree of V increases by 1, when a good child is born. More precisely, the degree of V increases when good children of the initial i-clique are born, and then good children of the good children are born, etc. Then the degree of V at time t is the total number of good children at that time.

Let $\hat{p}_{i,j}$ be the probability that a type-i parent gives birth to a type-j good child. Then

$$\hat{p}_{i,j} = p_{i,j}\frac{j-1}{i}.$$

We see that the good children process is a branching process having types $2, 3, \dots, N$. The length of the life of an i-clique is the same λ_i as in the case of our original process.

Let $\hat{\xi}_{i,j}$ be the number of type-j good children of a type-i good child. Then the average number of good children is

$$\hat{m}_{i,j}(t) = \mathbb{E}\hat{\xi}_{i,j}(t) = \frac{j-1}{i}m_{i,j}(t) = \frac{j-1}{i}p_{i,j}F(t).$$

The Laplace transform of $\hat{m}_{i,j}(t)$ is

$$\hat{m}^*_{i,j}(\kappa) = \frac{j-1}{i}m^*_{i,j}(\kappa) = \frac{j-1}{i}p_{i,j}A(\kappa).$$

Let

$$\hat{M}(\kappa) = \left(\hat{m}^*_{i,j}(\kappa)\right)_{i,j=2}^{N}$$

be the matrix of the Laplace transforms. Denote by $\hat{\varrho}(\kappa)$ the Perron root of $\hat{M}(\kappa)$. We shall suppose that the good children process is supercritical, i.e., $\hat{\varrho}(0) > 1$. If the process is supercritical, then in our case there exists a positive Malthusian parameter $\hat{\alpha}$ such that $\hat{\varrho}(\hat{\alpha}) = 1$. Let $\hat{\mathbf{v}} = (\hat{v}_2, \dots, \hat{v}_N)^\top$ be the right eigenvector of $\hat{M}(\hat{\alpha})$ corresponding to eigenvalue 1 and normed according to $\hat{v}_2 + \dots + \hat{v}_N = 1$. Let $\hat{\mathbf{u}} = (\hat{u}_2, \dots, \hat{u}_N)^\top$ be the left eigenvector of $\hat{M}(\hat{\alpha})$ corresponding to eigenvalue 1 and normed according to $\hat{v}_2\hat{u}_2 + \dots + \hat{v}_N\hat{u}_N = 1$.

Now, we can present our result on the asymptotic behaviour of the degree of a fixed vertex.

Theorem 1.5.1. *Assume that the matrix $(p_{i,j})_{i,j=2}^N$ is irreducible and acyclic. Assume that the good children process is supercritical, i.e., $\hat{\varrho}(0) > 1$. Let $\hat{\alpha}$ be the positive Malthusian parameter, so $\hat{\varrho}(\hat{\alpha}) = 1$. Let V be a fixed vertex which is initially a member of an i-clique, $2 \le i \le N$.*

Let ${}_iV(t)$ denote the degree of V, more precisely, the number of all edges being connected to V up to time t. Then

$$\lim_{t\to\infty} e^{-\hat{\alpha}t}\,{}_iV(t) = {}_i\hat{W}\frac{\hat{v}_i\sum_{j=2}^N \hat{u}_j}{-\hat{\alpha}A'(\hat{\alpha})} \tag{1.5.2}$$

almost surely.

Let ${}_i\bar{V}(t)$ denote the number of those edges, for which one of the endpoints is V and belonging to a clique alive at time t. Then

$$\liminf_{t\to\infty} e^{-\hat{\alpha}t}\,{}_i\bar{V}(t) \ge {}_i\hat{W}\frac{A(\hat{\alpha})\hat{v}_i\sum_{j=2}^N \hat{u}_j}{-A'(\hat{\alpha})} \tag{1.5.3}$$

almost surely.

The quantity $_i\hat{W}$ *is a.s. non-negative,* $\mathbb{E}(_i\hat{W}) = 1$, *and* $_i\hat{W}$ *is a.s. positive in the event of survival of the good children process.*

Proof. Here we shall use some results obtained during the proof of Theorem 1.4.1. We have

$$_iV(t) = (i-1) + \sum_{j=2}^{N} V_{i,j}(t),$$

where $V_{i,j}(t)$ denotes the number of type-j good offspring of a type-i mother. We shall find the limit of $V_{i,j}(t)$.

As in the proof of Theorem 1.4.1, we shall check the conditions of Proposition 1.3.1. Conditions (a), (c), (iv), and (vi) are true because they are true for the original process (see Theorem 1.4.1). (b2) and (b1) are true because we supposed that $\hat{\varrho}(0) > 1$. Condition (d) is true because we supposed that the matrix $(p_{i,j})_{i,j=2}^{N}$ is irreducible and acyclic.

Now, let $\Phi_x(t) = 1$, if the clique x is a good j-clique, and $\Phi_x(t) = 0$ otherwise. Then conditions (i), (ii), (iii), and (v) are satisfied. Moreover, $\mathbb{E}\Phi_j(t) = 1$ and $\mathbb{E}\Phi_l(t) = 0$ for $l \neq j$.

So Proposition 1.3.1 implies

$$\lim_{t\to\infty} e^{-\hat{\alpha}t} V_{i,j}(t) = {}_i\hat{W} \frac{\hat{v}_i \hat{u}_j}{-\hat{\alpha} A'(\hat{\alpha})}$$

almost surely. So we obtain (1.5.2).

To obtain (1.5.3), let $\Phi_x(t) = 1$ if x is a good j-clique and it is alive at time t, and let $\Phi_x(t) = 0$ otherwise. Therefore $\mathbb{E}\Phi_j(t) = 1 - L_j(t)$ and $\mathbb{E}\Phi_l(t) = 0$ for $l \neq j$. Conditions $(i–iii)$ and (v) are true. Now

$$\int_0^\infty e^{-\alpha s} \mathbb{E}\Phi_j(s)\, ds = \int_0^\infty e^{-\alpha s}(1 - L_j(s)) ds = A(\alpha).$$

Now, we can apply Proposition 1.3.1. However, in (1.5.3) we cannot offer equality, because an edge can belong to a clique being alive and at the same time it can belong to a clique being dead. □

1.6 The extinction probability

1.6.1 The joint generating function

To find the probability of the extinction of our network, we need the generating function. We calculate the joint generating function of the variables $\Pi_n(\lambda_n)$ and $\xi_{n,j}(\lambda_n)$, $j = 1, \ldots, (n+1) \wedge N$, $n = 1, \ldots, N$. Consider the generic n-clique, where n is fixed with $1 \leq n \leq N$. Consider the sequence

$$w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = \mathbb{P}\left(\Pi_n(\lambda_n) = i, \xi_{n,j}(\lambda_n) = k_j, j = 1, \ldots, (n+1) \wedge N\right), \quad (1.6.1)$$

where $i = 0, 1, 2, \ldots,\ k_j = 0, 1, 2, \ldots$.

Then (1.6.1) describes the joint distribution of the last reproduction time and the offspring size of the generic n-clique during its whole lifetime. Here Π_n is the Poisson process that describes the birth times of the generic n-clique and λ_n is its life length. In other words, we can say that $w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$ gives the probability of the event that the last birth time of the generic n-clique before its death is τ_i and the total number of type-j offspring up to its death is k_j. That is, we can write $w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$ in the following form:

$$w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = \mathbb{P}\left(\tau_i \leq \lambda_n < \tau_{i+1}, \xi_{n,j}\left(\tau_i\right) = k_j, j = 1, \ldots, (n+1) \wedge N\right). \quad (1.6.2)$$

To determine the desired joint generating function, first we consider the following sequence:

$$u_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = \mathbb{P}\left(\tau_i \leq \lambda_n, \xi_{n,j}\left(\tau_i\right) = k_j, j = 1, \ldots, (n+1) \wedge N\right). \quad (1.6.3)$$

We use the notation $\tau_0 = 0$, so $u_{0,\{k_j\}_{j=1}^{(n+1)\wedge N}} = 1$ if each k_j is zero, but it is 0 if any k_j is positive. Moreover, $u_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = 0$, if any subscript is negative. Now, for a while, assume that τ_i and τ_{i-1} are fixed and $\xi_{n,j}\left(\tau_{i-1}\right) = m$ is known. Then using the definition of the survival function of the life length given in (1.1.3) and by Assumption (1.1.4), we have

$$\mathbb{P}\left(\lambda_n \geq \tau_i | \lambda_n \geq \tau_{i-1}\right) = \exp\left(-\left(\tau_i - \tau_{i-1}\right)(b + cm)\right). \quad (1.6.4)$$

But τ_i and τ_{i-1} are random. So, a simple calculation that uses the fact that the increments of a Poisson process with intensity 1 are exponential with parameter 1 can lead us to obtain that

$$\mathbb{P}\left(\lambda_n \geq \tau_i | \lambda_n \geq \tau_{i-1}\right) = \frac{1}{1 + b + cm}. \quad (1.6.5)$$

That is, (1.6.5) is the probability that the object will not die before the i^{th} birth event. Using the above calculations and the law of total probability, we can give the following recursion for the sequence $u_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$, $i = 1, 2, \ldots$

$$\begin{aligned} u_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = & \quad (1.6.6) \\ = \sum_{l=1}^{(n+1)\wedge N} & \mathbb{P}\left(\tau_{i-1} \leq \lambda_n, \xi_{n,l}\left(\tau_{i-1}\right) = k_l - 1, \xi_{n,j}\left(\tau_{i-1}\right) = k_j,\right. \\ & \left. j = 1, \ldots, (n+1) \wedge N, j \neq l\right) p_{n,l-1} \frac{1}{1 + b + c\left(\left(\sum_{j=1}^{(n+1)\wedge N} k_j\right) - 1\right)} = \\ = \sum_{l=1}^{(n+1)\wedge N} & u_{i-1,k_l-1,\{k_j\}_{j=1,j\neq l}^{(n+1)\wedge N}} p_{n,l-1} \frac{1}{1 + b + c\left(\left(\sum_{j=1}^{(n+1)\wedge N} k_j\right) - 1\right)}. \end{aligned}$$

Here $p_{n,j}$ is the probability that the new vertex is born with j new edges. To obtain the above recursion, we also used the following. Considering the generic n-clique, the definition of the evolution process implies that at each birth step, exactly 1 offspring is born, the smallest possible offspring size is 1, and the maximal offspring size is $(n+1)\wedge N$. Moreover, the offspring sizes of the generic n-clique at any two consecutive birth steps are independent. Using (1.6.3) and (1.6.5), we can also see that

$$\begin{aligned}
w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} &\mathbb{P}\left(\tau_i \le \lambda_n < \tau_{i+1}, \xi_{n,j}\left(\tau_i\right) = k_j, j=1,\dots,(n+1)\wedge N\right) \\
&= \mathbb{P}\left(\lambda_n < \tau_{i+1} | \tau_i \le \lambda_n, \xi_{n,j}\left(\tau_i\right) = k_j, j=1,\dots,(n+1)\wedge N\right) \\
&\times \mathbb{P}\left(\tau_i \le \lambda_n,\ \xi_{n,j}\left(\tau_i\right) = k_j, j=1,\dots,(n+1)\wedge N\right) \\
&= \frac{b + c\sum_{j=1}^{(n+1)\wedge N} k_j}{1 + b + c\sum_{j=1}^{(n+1)\wedge N} k_j} u_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}. \qquad (1.6.7)
\end{aligned}$$

Let us consider the following sequence $v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$, where $w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$ is defined in (1.6.1):

$$v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = \frac{w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}}{b + c\sum_{j=1}^{(n+1)\wedge N} k_j} = \frac{u_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}}{1 + b + c\sum_{j=1}^{(n+1)\wedge N} k_j}. \qquad (1.6.8)$$

Moreover, using the recursion (1.6.6), we see that the sequence $v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$ satisfies the following recurrence relation:

$$\left(1 + b + c\sum_{j=1}^{(n+1)\wedge N} k_j\right) v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} = \sum_{l=1}^{(n+1)\wedge N} p_{n,l-1} v_{i-1,k_l-1,\{k_j\}_{j=1,j\neq l}^{(n+1)\wedge N}}, \qquad (1.6.9)$$

where the initial values are

$$v_{0,\{0\}_{j=1}^{(n+1)\wedge N}} = \frac{1}{1+b} \quad \text{and} \quad v_{0,\{k_j\}_{j=1}^{(n+1)\wedge N}} = 0 \text{ if } \exists j : k_j \neq 0. \qquad (1.6.10)$$

Let us denote by $G\left(x, \{x_j\}_{j=1}^{(n+1)\wedge N}\right)$ the generating function of the sequence $v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$. We have

$$G\left(x, \{x_j\}_{j=1}^{(n+1)\wedge N}\right) = \sum_{i=0}^{\infty} \sum_{j=1}^{(n+1)\wedge N} \sum_{k_j=0}^{\infty} v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}} x^i \prod_{j=1}^{(n+1)\wedge N} x_j^{k_j}. \qquad (1.6.11)$$

To determine the generating function $G\left(x, \{x_j\}_{j=1}^{(n+1)\wedge N}\right)$, we multiply with $x^i \prod_{j=1}^{(n+1)\wedge N} x_j^{k_j}$ and then take the sum of both sides of (1.6.9). In this way, we obtain that

$$\sum_{i=0}^{\infty}\sum_{j=1}^{(n+1)\wedge N}\sum_{k_j=0}^{\infty}\left(1+b+c\sum_{j=1}^{(n+1)\wedge N}k_j\right)v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}x^i\prod_{j=1}^{(n+1)\wedge N}x_j^{k_j}$$
$$=\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}\sum_{i=0}^{\infty}\sum_{j=1}^{(n+1)\wedge N}\sum_{k_j=0}^{\infty}v_{i-1,k_l-1,\{k_j\}_{j=1,j\neq l}^{(n+1)\wedge N}}x^i\prod_{j=1}^{(n+1)\wedge N}x_j^{k_j}. \quad (1.6.12)$$

From this equation and using the definition of the generating function given by (1.6.11), we can obtain that

$$(1+b)\left(G\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right)-\frac{1}{1+b}\right)+c\sum_{j=1}^{(n+1)\wedge N}x_jG'_{x_j}\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right)$$
$$=\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_lG\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right). \quad (1.6.13)$$

Let $h(t)=G\left(x,\{tx_j\}_{j=1}^{(n+1)\wedge N}\right)$. By (1.6.10), we have the initial condition

$$h(0)=G\left(x,\{0\}_{j=1}^{(n+1)\wedge N}\right)=\sum_{i=0}^{\infty}v_{i,\{0\}_{j=1}^{(n+1)\wedge N}}x^i=\frac{1}{1+b}.$$

Now, substituting x_j with tx_j in (1.6.13), we get the following first-order differential equation:

$$h'(t)+h(t)\frac{1}{ct}\left((1+b)-t\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_l\right)=\frac{1}{ct} \quad (1.6.14)$$

where the initial value is

$$h(0)=\frac{1}{1+b}. \quad (1.6.15)$$

The solution of the above initial value problem (1.6.14) and (1.6.15) is

$$h(t)=t^{\frac{-(1+b)}{c}}e^{\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_l}{c}t}\frac{1}{c}\int_0^t s^{\frac{1+b}{c}-1}e^{-\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_l}{c}s}ds. \quad (1.6.16)$$

Substituting $t=1$ in (1.6.16), we obtain that the generating function of the sequence $v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$ is

$$G\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right)=h(1)=\frac{1}{c}\int_0^1 s^{\frac{1+b}{c}-1}e^{\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_l}{c}(1-s)}ds. \quad (1.6.17)$$

Now, let us denote by $g_{\Pi_n,\{\xi_{n,j}\}_{j=1}^{(n+1)\wedge N}}$ the joint generating function of the variables $\Pi_n(\lambda_n)$ and $\xi_{n,j}(\lambda_n)$, $j=1,\ldots,(n+1)\wedge N$. Using the definitions of the sequences $w_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$ and $v_{i,\{k_j\}_{j=1}^{(n+1)\wedge N}}$, by (1.6.13) we have

$$g_{\Pi_n,\{\xi_{n,j}\}_{j=1}^{(n+1)\wedge N}}\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right)=\mathbb{E}\left(x^{\Pi_n(\lambda_n)}\prod_{j=1}^{(n+1)\wedge N}x_j^{\xi_{n,j}(\lambda_n)}\right)$$

$$=\sum_{i=0}^{\infty}\sum_{j=1}^{(n+1)\wedge N}\sum_{k_j=0}^{\infty}\mathbb{P}\left(\Pi_n(\lambda_n)=i,\xi_{n,j}(\lambda_n)=k_j,j=1,\ldots,(n+1)\wedge N\right)x^i$$

$$\times\prod_{j=1}^{(n+1)\wedge N}x_j^{k_j}=bG\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right)+c\sum_{j=1}^{(n+1)\wedge N}x_jG'_{x_j}\left(x,\{x_j\}_{j=1}^{(n+1)\wedge N}\right)$$

$$=e^{\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}x_l}{c}}\frac{1}{c}$$

$$\times\int_0^1 s^{\frac{1+b}{c}-1}e^{-\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_l}{c}s}\left(b+(1-s)\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}xx_l\right)ds. \quad (1.6.18)$$

With $x=1$ in (1.6.18), we obtain that the generating function of the total offspring distribution of the generic n-clique is

$$f_n\left(\{x_j\}_{j=1}^{(n+1)\wedge N}\right)=e^{\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}x_l}{c}}$$

$$\times\frac{1}{c}\int_0^1 s^{\frac{1+b}{c}-1}e^{-\frac{\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}x_l}{c}s}\left(b+(1-s)\sum_{l=1}^{(n+1)\wedge N}p_{n,l-1}x_l\right)ds. \quad (1.6.19)$$

1.6.2 The probability of extinction

Consider the embedded multi-type Galton-Watson process which can be constructed in the following way. At the initial time, the ancestor alone constitutes the starting generation of the Galton-Watson process. During its life, the ancestor produces a random number of offspring. All of the offspring of this ancestor form the 1^{st} generation. Generally, the n^{th} generation is formed by the offspring of the members of the $(n-1)^{\text{th}}$ generation.

Under some reasonable conditions, the probability of extinction of our process is the same as the probability of extinction of the embedded multi-type Galton-Watson process, see Theorem 7.1 in Chapter 3 of [55]. In our case those conditions are satisfied. Let $\mathbb{M}$ be the matrix of the expected total offspring number of our process. Then

$$\mathbb{M}=\left(m_{i,j}(\infty)\right)_{i,j=1}^{N}=A(0)\left(p_{i,j}(\infty)\right)_{i,j=1}^{N}.$$

We see that the matrix $\mathbb{M}$ contains the expected offspring numbers of the embedded Galton-Watson process.

From the theory of multi-type branching processes, we know that the probability of extinction may depend on the type of the ancestor. Moreover, the vector of extinction probabilities is the solution of a vector equation.

Theorem 1.6.1. *Let us denote by s_i the probability of extinction when the ancestor of the network is an i-type object. Let $\mathbf{s} = (s_1, \ldots, s_N)$. Assume that $(p_{i,j}(\infty))_{i,j=1}^N$ is an irreducible acyclic Markov transition matrix. Denote by ϱ the Perron–Frobenius root of $\mathbb{M}$. If $\varrho \leq 1$, then $s_1 = s_2 = \cdots = s_N = 1$. If $\varrho > 1$, then $s_1 < 1, s_2 < 1, \ldots, s_N < 1$. In any case, $\mathbf{s}$ is the smallest non-negative solution of the vector equation*

$$\mathbf{s} = \mathbf{f}(\mathbf{s}), \tag{1.6.20}$$

where $\mathbf{f} = (f_1, \ldots, f_N)$ and the functions f_k are defined in (1.6.19).

Proof. Our conditions ensure that the matrix $\mathbb{M}$ is positively regular. Now, applying Theorem 7.1 in Chapter 1 of [55], we can obtain the desired result. □

1.7 Simulation results

In this section, we present some numerical results for our previously presented asymptotic theorems. We used the Julia environment because of the possibility of fast numerical computing allowed by the well-written dynamic structures. The code can be downloaded from GitHub, see [71].

According to Theorem 1.4.1, the numbers of n-cliques, when the process moves forward in time, are asymptotically close to a straight line on the logarithmic scale. To support numerically our Theorem 1.4.1, we studied the slope of the sequence of the simulated number of n-cliques being born up to time t on the logarithmic scale.

Example 1.7.1. *Now, we present an example of the Leslie model with $N = 5$, the transition matrix*

$$P_1 = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 & 0 \\ 0.2 & 0 & 0.8 & 0 & 0 \\ 0.3 & 0 & 0 & 0.7 & 0 \\ 0.4 & 0 & 0 & 0 & 0.6 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{pmatrix},$$

and with parameters of the hazard rate $b = 0.2$ and $c = 0.2$. In Figure 1.6, we illustrate the first 2^{19} birth steps of two different simulated processes. The five solid lines represent the number of n-cliques being born, for $n = 1, \ldots, 5$ on a logarithmic scale, while the dotted line's slope $\hat{\alpha} = 0.6462$ equals the numerical approximation of the Malthusian parameter. We can see that the plotted lines are parallel straight lines for large values of t, which in fact gives nice feedback to our asymptotic results.

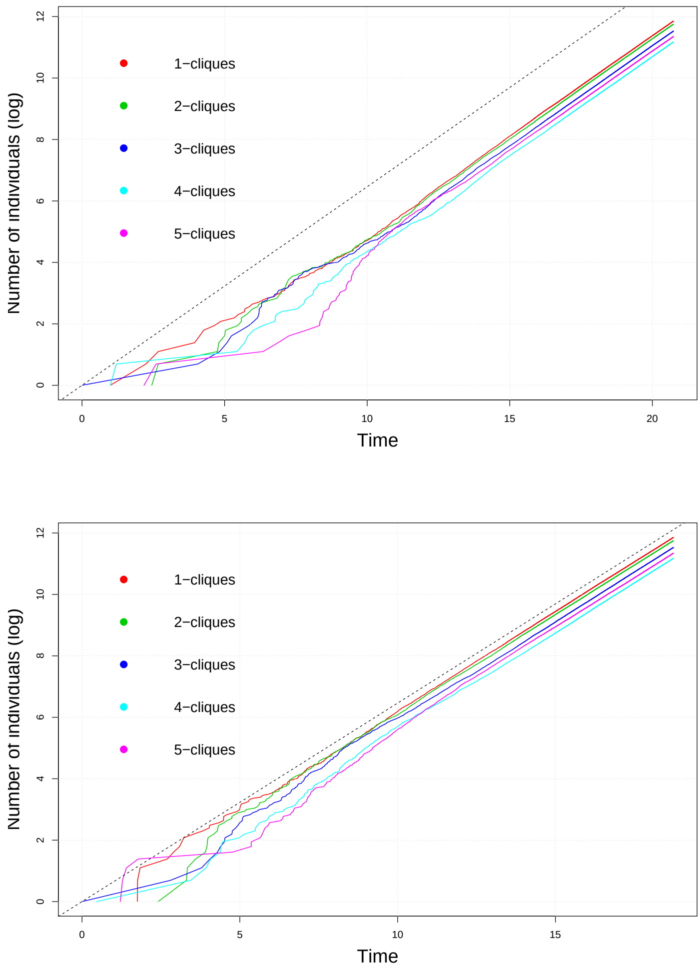


Figure 1.6: Two example processes for the Leslie model with transition matrix P_1, and parameters $b = 0.2$ and $c = 0.2$.

Then, to obtain statistically significant evidence, we used 100 simulated processes to construct a 99% confidence stripe for the trajectory of the number of n-cliques. For demonstration, in Figure 1.7, the 99% confidence stripes are presented for 4- and 5-cliques. The red lines are the borders of the stripes.

In Table 1.1, the boundaries of the 99% confidence intervals for α *are presented. Each fixed clique size gives a confidence interval. The columns labeled with* 0.5% *and* 99.5% *show the lower and the upper bounds calculated from simulations. As* *the results show, the numerical approximation* $\hat{\alpha}_1 = 0.6462$ *is contained by all confidence intervals.*

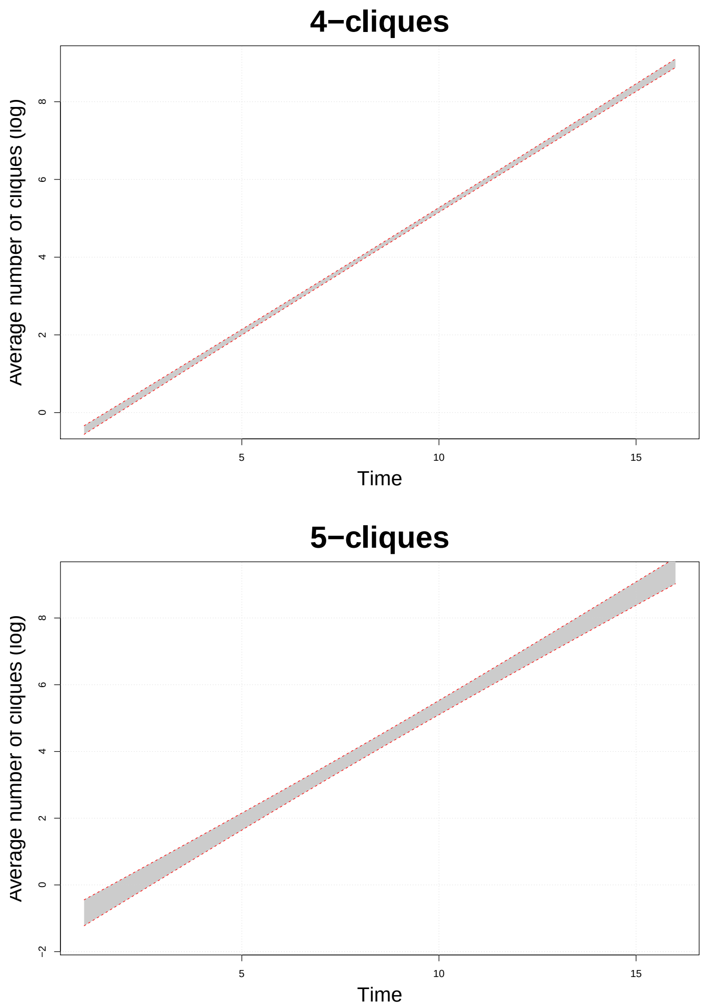


Figure 1.7: The 99% confidence stripes based on 100 simulations.

Table 1.1: All 99% confidence intervals for the slopes of the number of n-cliques include the approximation $\hat{\alpha}_1 = 0.6462$ of the Malthusian parameter.

Type	0.5%	99.5%
1	0.6435	0.6769
2	0.6454	0.7205
3	0.6437	0.8244
4	0.6167	0.6420
5	0.6392	0.7276

Example 1.7.2. *Now, we present another example with a transition matrix*

$$P_2 = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.8 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0.7 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \end{pmatrix},$$

and parameters $b = 0.4$ and $c = 0.4$. In this model, a newcomer joining an n-clique can contact any other group members, when $n = 1, \ldots, 4$. For $n = 5$, it is not possible that the newcomer joins all former clique members. In Figure 1.8, we show a simulated example of the process. The five solid lines represent the number of n-cliques being born, for $n = 1, \ldots, 5$ on a logarithmic scale, while the dotted line's slope equals $\hat{\alpha}_2 = 0.3391$. Table 1.2 contains the confidence intervals for the slope, using 100 simulated processes.

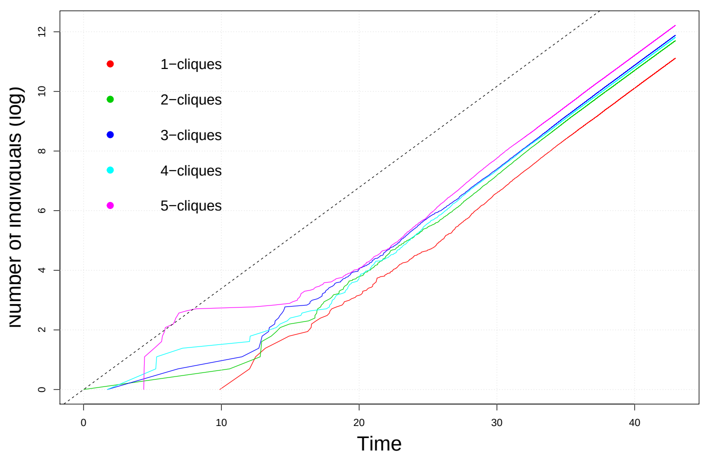


Figure 1.8: Example process with transition matrix P_2, $b = 0.4$, and $c = 0.4$.

Table 1.2: 99% confidence intervals, $\hat{\alpha}_2 = 0.3391$.

Type	0.5%	99.5%
1	0.3265	0.3623
2	0.3361	0.3547
3	0.3208	0.3490
4	0.3279	0.3411
5	0.3350	0.3548

Chapter 2

A discrete-time network evolution model based on cliques

2.1 Introduction

In this chapter, we consider the generalization of the model of Backhausz and Móri [5] to k-cliques. The chapter is based on the results of our paper [29].

Network theory is important both for real-life applications and theoretical research. It studies general properties of networks and offers models and methods to understand their evolution. Well-known large networks are e.g. the World Wide Web, the Internet, metabolic networks, and social networks. It has applications in logistics, electrical engineering, biology, economics, ecology, public health, sociology and many other fields. One can find several general facts on network theory in the book [7]. A random graph can describe an evolving network. The vertices of the graph are the nodes of the network and the edges of the graph are the connections among the nodes. Our aim is to introduce a new network evolution procedure and find its basic properties.

Backhausz and Móri [5] introduced a random graph evolution model with moderate edge density. Their model is the following. The starting graph is an empty one of size 2. At each step, two vertices are chosen uniformly at random. If the two vertices chosen are not connected, we connect them with 1 edge. If the two vertices are connected, we delete the connecting edge, add a new vertex to the graph, and connect the new vertex to both of the selected vertices. The result of this procedure is an evolving graph containing n edges after the n^{th} step. In [5], several asymptotic theorems are proved for the number of vertices and the degree of a fixed vertex. Also in [5], a short overview is presented about the edge densities of some well-known random graph models.

We study the following extension of the model of [5]. Instead of connections of two vertices, we consider connections of k vertices, where $k \geq 2$ is a fixed integer. So the main ingredients of our model are the k-cliques. The evolution of our graph is based on constructions and deletions of k-cliques. A k-clique is a sub-graph containing k vertices and any two different vertices are connected by 1 edge. When we form a k-clique, then we draw $\binom{k}{2}$ new edges among k vertices, and we add this new clique to the list of k-cliques.

The initial graph at time $n = 0$ contains k vertices and no edges. In the first step i.e. when the time is $n = 1$, we connect the k vertices to obtain a single k-clique. Then, in each step, we choose k vertices uniformly at random from the existing vertices. If they do not form a k-clique, then we construct a new k-clique on these vertices. In the other case, when the sub-graph consisting of the k vertices chosen is a k-clique, then that k-clique is deleted. Then a new vertex is added to the graph and two new k-cliques are created. The detailed description of this procedure is given in Section 2.2.

In Section 2.3, using martingale theory, we prove almost sure limit theorem for the number of vertices, then we show its asymptotic normality, see Theorem 2.3.1. In Section 2.4, we obtain an almost sure limit theorem for the degree of a fixed vertex, see Theorem 2.4.1. Then, we present an asymptotic normality result for degree of a fixed vertex. In sections 2.3 and 2.4, our results are extensions of the results of [5]. However, the results of Section 2.5 are new for any value of k, including the particular case of $k = 2$ studied in [5]. These new results are functional limit theorems. Theorem 2.5.1 is a functional limit theorem for the number of vertices. Theorem 2.5.2 and Proposition 2.5.1 are multidimensional functional limit results for the joint distribution of the degrees of several fixed vertices. For the proof, we apply general functional limit theorems for martingales. In Section 2.6, we present simulation results.

We remark, that instead of uniform choice, one can use the preferential attachment principle for certain sub-graphs, but then the asymptotic behaviour of the graph will be different, see e.g. [32].

2.2 The model

We study a discrete time network evolution model. Our network (i.e. the graph) can contain multiple edges. The evolution of the graph is based on constructions and deletions of k-cliques, where $k \geq 2$ is a fixed integer. A k-clique is a sub-graph containing k vertices and $\binom{k}{2}$ edges, i.e. any two different vertices are connected by 1 edge. When we form a k-clique, then we draw $\binom{k}{2}$ new edges among k vertices, and we add this new clique to the list of k-cliques. The $\binom{k}{2}$ new edges will be considered as the own edges of the clique, but the formerly existing edges are not considered as own edges of the clique at hand. When we delete a k-clique, then we delete it from the list of k-cliques, and we delete its $\binom{k}{2}$ own edges. But we do not delete its vertices and we do not delete those edges which were not own edges of the clique at hand.

The initial graph at time $n = 0$ contains k vertices and no edges. In the first step, i.e. when the time is $n = 1$, we connect the k vertices to obtain a single k-clique. The second step is the following. We choose two vertices uniformly at random, let us denote them by v_1 and v_2. Then we add a new vertex and construct two new k-cliques. The vertices of the first k-clique are the existing $k+1$ vertices but v_1, while the vertices of the second k-clique are the existing $k+1$ vertices but v_2. Then the original k-clique is deleted.

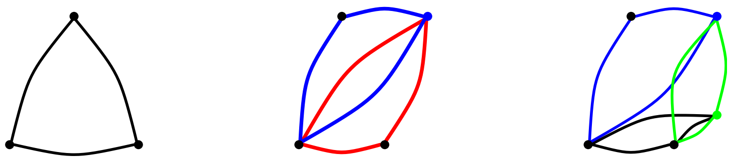


Figure 2.1: Three evolution steps of the network when the cliques are triangles

From now on, in each step, we choose k vertices uniformly at random from the existing vertices. If they do not form a k-clique, then we construct a new k-clique on these vertices (i.e. we connect them using $\binom{k}{2}$ new edges). In the other case, when the sub-graph consisting of the k vertices chosen is a k-clique, then that k-clique is deleted but its vertices are used to construct two new k-cliques as in the second step. That is a new vertex is added to the graph and using this new vertex and the k vertices of the just deleted k-clique, two new k-cliques are created in the same way as in the second step.

2.3 The number of vertices

We see that after n steps the graph has $n\binom{k}{2}$ edges. Let V_n denote the number of vertices in the model after n steps and the σ-field generated by the first n steps is denoted by $\mathcal{F}_n$. In the first theorem we prove that the magnitude of V_n is $n^{\frac{2}{k+1}}$ and V_n is asymptotically normal. Let us denote the normal distribution with mean m and variance σ^2 by $\mathcal{N}(m, \sigma^2)$.

Remark 2.3.1. *To prove our next theorem, we shall use Lemma 2.2 of Móri and Backhausz [5]:*

Let $g(n)$, $n = 0, 1, \dots,$ be an increasing positive sequence, and define $G(n) = \sum_{i=0}^{n-1} g(i)$. Assume $g(n) = O(G(n)^\delta)$ for some δ, $0 < \delta < 1$ (for example, every polynomial multiplied by a power of logarithms satisfies the condition). Roughly speaking, this means that g can only grow polynomially. Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration (an increasing sequence of σ-fields), and $A_n \in \mathcal{F}_n$, $n = 1, 2, \dots$. Let $S_n = \sum_{i=1}^{n} \mathbb{I}(A_i)$, where $\mathbb{I}(\,\cdot\,)$ stands for the indicator of the event in brackets. Finally, let $(\eta_n)_{n\geq 0}$ be an adapted sequence of non-negative random variables such that $H_n = \sum_{i=0}^{n-1} \eta_i \to \infty$ with probability 1. Suppose

$$P(A_n \mid \mathcal{F}_{n-1}) = \frac{\eta_{n-1}}{g(S_{n-1})}, \qquad n = 1, 2, \dots.$$

Then $G(S_n) \sim H_n$ a.s. as $n \to \infty$.

Theorem 2.3.1. *As $n \to \infty$, the following almost sure convergence holds for the number of vertices in the graph after n steps:*

$$\frac{V_n}{\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}} \to 1. \tag{2.3.1}$$

Furthermore, we have

$$\frac{1}{n^{\frac{1}{k+1}}}\left(V_n - \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{2k+1}\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}}\right) \tag{2.3.2}$$

as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution.

We shall also write (2.3.1) as

$$V_n \sim \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}, \qquad \text{almost surely as} \quad n \to \infty. \tag{2.3.3}$$

Proof. To prove the first statement, we apply Remark 2.3.1. Let us use the notation A_n for the event that we choose a k-clique in the n^{th} step and let $S_n = \sum_{i=1}^{n} \mathbb{I}_{A_i}$, where $\mathbb{I}_{A_i}$ is the indicator of the event A_i . We have started at time $n = 0$ with an empty graph on k vertices, so $V_n = k + S_n$. It is also easy to see that

$$\mathbb{P}\left(A_n | \mathcal{F}_{n-1}\right) = \frac{n-1}{\binom{V_{n-1}}{k}}.$$

We remark that this relation is true also for $n = 1$ and $n = 2$ as $V_0 = k$ and $A_1 = \emptyset$, $A_2 = \Omega$. By setting $\eta_n = n$ and $g(n) = \binom{k+n}{k}$ in Remark 2.3.1, we obtain

$$H_n = \sum_{i=0}^{n-1} \eta_i = 1 + \ldots + (n-1) = \frac{n(n-1)}{2} = \binom{n}{2} \to \infty,$$

$$G(n) = \sum_{i=0}^{n-1} g(i) = \binom{k}{k} + \binom{k+1}{k} + \ldots + \binom{k+n-1}{k} = \binom{k+n}{k+1}.$$

Since $g(n) = \binom{n+k}{k} \leq \frac{(n+k)^k}{k!}$ and $G(n) = \binom{n+k}{k+1} \geq \frac{n^{k+1}}{(k+1)!}$, there exists a $0 < \delta < 1$ such that $g(n) = \mathcal{O}(G(n)^\delta)$. To this end we just have to choose a δ with $\frac{k}{k+1} < \delta < 1$. Hence the conditions of Remark 2.3.1 are satisfied, so we have

$$G(S_n) = \binom{V_n}{k+1} \sim \binom{n}{2}. \tag{2.3.4}$$

It implies (2.3.1).

For the proof of the second statement, we use Corollary 3.1 of the monograph [40]. Let us take the martingale difference array $(X_{n,i}, \mathcal{F}_{n,i})$, $n \geq 1$, $1 \leq i \leq n$, where

$$X_{n,i} = \frac{1}{n^{\frac{2k+1}{k+1}}} \left[\binom{V_i}{k+1} - \binom{V_{i-1}}{k+1} - (i-1) \right],$$

and $\mathcal{F}_{n,i} = \mathcal{F}_i$. It is indeed a martingale difference, because by the evolution rules of our network

$$\begin{aligned} &\mathbb{E}\left(\frac{1}{n^{\frac{2k+1}{k+1}}} \left[\binom{V_i}{k+1} - \binom{V_{i-1}}{k+1} - (i-1) \right] \,|\, \mathcal{F}_{n,i-1} \right) = \\ &\quad = \frac{1}{n^{\frac{2k+1}{k+1}}} \left[\binom{V_{i-1}}{k+1} \cdot \frac{\binom{V_{i-1}}{k} - (i-1)}{\binom{V_{i-1}}{k}} + \binom{V_{i-1}+1}{k+1} \cdot \frac{i-1}{\binom{V_{i-1}}{k}} - \right. \\ &\qquad \left. - \binom{V_{i-1}}{k+1} - (i-1) \right] = \\ &\quad = \frac{1}{n^{\frac{2k+1}{k+1}}} \left[\binom{V_{i-1}}{k+1} + (i-1) \cdot \frac{-\binom{V_{i-1}}{k+1} + \binom{V_{i-1}+1}{k+1}}{\binom{V_{i-1}}{k}} - \right. \\ &\qquad \left. - \binom{V_{i-1}}{k+1} - (i-1) \right] = 0. \end{aligned} \tag{2.3.5}$$

For fixed n, the sequence $X_{n,i}$ is the difference of the martingale $\left(\binom{V_i}{k+1} - \binom{i}{2}\right) / n^{\frac{2k+1}{k+1}}$.

For the sums of the conditional variances, using (2.3.4), we have

$$\sum_{i=1}^{n}\mathbb{E}\left(X_{n,i}^2 \mid \mathcal{F}_{n,i-1}\right) =$$

$$= \frac{1}{n^{\frac{4k+2}{k+1}}}\sum_{i=1}^{n}\left[\mathbb{E}\left\{\left[\binom{V_i}{k+1}-\binom{V_{i-1}}{k+1}\right]^2 \mid \mathcal{F}_{i-1}\right\}-(i-1)^2\right] =$$

$$= \frac{1}{n^{\frac{4k+2}{k+1}}}\sum_{i=1}^{n}\left[\left[\binom{V_{i-1}+1}{k+1}-\binom{V_{i-1}}{k+1}\right]^2\cdot\frac{i-1}{\binom{V_{i-1}}{k}}-(i-1)^2\right] =$$

$$= \frac{1}{n^{\frac{4k+2}{k+1}}}\left[\sum_{i=1}^{n}\binom{V_{i-1}}{k}\cdot(i-1)-\frac{n(n-1)(2n-1)}{6}\right] \sim$$

$$\sim \frac{1}{n^{\frac{4k+2}{k+1}}}\left[\sum_{i=0}^{n-1}\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}\cdot\frac{i^{\frac{2k}{k+1}}\cdot i}{k!}-\frac{n(n-1)(2n-1)}{6}\right] \sim$$

$$\sim \frac{1}{n^{\frac{4k+2}{k+1}}}\left[\frac{\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}}{k!}\int_0^n x^{\frac{3k+1}{k+1}}\,dx-\frac{n(n-1)(2n-1)}{6}\right] \sim$$

$$\sim \frac{\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}}{k!}\cdot\frac{n^{\frac{4k+2}{k+1}}}{\frac{4k+2}{k+1}}\cdot\frac{1}{n^{\frac{4k+2}{k+1}}} = \frac{\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}}{k!\cdot\frac{2(2k+1)}{k+1}}.$$

Then applying similar calculations as in (2.3.5) and using (2.3.4), we have the bound

$$|X_{n,i}| = \frac{1}{n^{\frac{2k+1}{k+1}}}\left|\binom{V_i}{k+1}-\binom{V_{i-1}}{k+1}-(i-1)\right| \le$$

$$\le \frac{1}{n^{\frac{2k+1}{k+1}}}\left|\binom{V_i}{k}-(i-1)\right| \le \frac{1}{n^{\frac{2k+1}{k+1}}}\left|\binom{V_n}{k}-(i-1)\right| \sim$$

$$\sim \frac{1}{n^{\frac{2k+1}{k+1}}}\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}} n^{\frac{2k}{k+1}} = \left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}\frac{1}{n^{\frac{1}{k+1}}}.$$

It means that if $\varepsilon > 0$ is fixed and n is large enough, then $|X_{n,i}| < \varepsilon$, so the conditional Lindeberg condition is satisfied. Now Corollary 3.1 of [40] implies that

$$\frac{1}{n^{\frac{2k+1}{k+1}}}\left[\binom{V_n}{k+1}-\binom{n}{2}\right] \Rightarrow \mathcal{N}\left(0, \frac{\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}}{k!\cdot\frac{2(2k+1)}{k+1}}\right). \tag{2.3.6}$$

By using Taylor's expansion and using that $V_n \to \infty$ a.s., we get that

$$\left[\binom{V_n}{k+1}\right]^{\frac{1}{k+1}} = \frac{V_n}{[(k+1)!]^{\frac{1}{k+1}}}+\mathcal{O}(1)\text{ a.s.,}\quad\text{and}\quad \binom{n}{2}^{\frac{1}{k+1}} = \frac{n^{\frac{2}{k+1}}}{2^{\frac{1}{k+1}}}+\mathcal{O}(1).$$

With the above formulas, we obtain

$$V_n - \frac{[(k+1)!]^{\frac{1}{k+1}}}{2^{\frac{1}{k+1}}} n^{\frac{2}{k+1}} =$$

$$= \binom{V_n}{k+1}^{\frac{1}{k+1}} \cdot [(k+1)!]^{\frac{1}{k+1}} - [(k+1)!]^{\frac{1}{k+1}} \cdot \binom{n}{2}^{\frac{1}{k+1}} + \mathcal{O}(1) =$$

$$= [(k+1)!]^{\frac{1}{k+1}} \left[\binom{V_n}{k+1}^{\frac{1}{k+1}} - \binom{n}{2}^{\frac{1}{k+1}} \right] + \mathcal{O}(1).$$

Now we use $a - b = (a^{k+1} - b^{k+1})/(a^k + a^{k-1}b + \dots + b^k)$ with $a = \binom{V_n}{k+1}^{\frac{1}{k+1}}$ and $b = \binom{n}{2}^{\frac{1}{k+1}}$ and apply (2.3.4) to get

$$V_n - \frac{[(k+1)!]^{\frac{1}{k+1}}}{2^{\frac{1}{k+1}}} n^{\frac{2}{k+1}} = [(k+1)!]^{\frac{1}{k+1}} \frac{\binom{V_n}{k+1} - \binom{n}{2}}{(k+1)\left(\frac{n^2}{2}\right)^{\frac{k}{k+1}}} + \mathcal{O}(1) =$$

$$= \frac{[(k+1)!]^{\frac{1}{k+1}}}{k+1} \cdot 2^{\frac{k}{k+1}} \cdot n^{\frac{1}{k+1}} \cdot \frac{1}{n^{\frac{2k+1}{k+1}}} \left[\binom{V_n}{k+1} - \binom{n}{2} \right] + \mathcal{O}(1).$$

Finally, we use (2.3.6), and we obtain that the variance is

$$\frac{[(k+1)!]^{\frac{2}{k+1}}}{(k+1)^2} \cdot 2^{\frac{2k}{k+1}} \cdot \frac{\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}}}{k! \cdot \frac{2(2k+1)}{k+1}} = \frac{1}{2^{\frac{1}{k+1}}} \cdot [(k+1)!]^{\frac{1}{k+1}} \cdot \frac{1}{2k+1} =$$

$$= \frac{1}{2k+1} \left[\frac{(k+1)!}{2} \right]^{\frac{1}{k+1}},$$

so the limit distribution is $\mathcal{N}\left(0, \frac{1}{2k+1} \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}}\right)$.

□

2.4 The degree of a fixed vertex in the model

In this section, we investigate the asymptotic behaviour of a fixed vertex. Let us use the notation $d_n(v)$ for the degree of the vertex v at time n. At the start, the degree of the initial k vertices is zero and let us use $-(k-1), -(k-2), \dots, 0$ as the labels for these vertices. The new vertices will have labels $1, 2, \dots$ in the order they are added to the model.

Let us consider a step in the evolution of our graph. If the chosen k vertices do not form a k-clique, then they are connected to obtain a k-clique, which means that their degree is increased by $k-1$. If the sub-graph consisting of the chosen k vertices is a k-clique, then that clique is deleted, a new vertex is added and two new k-cliques are created. We choose two vertices $v_1 \neq v_2$ out of the k chosen ones. The first new k-clique is constructed with these $k+1$ vertices but v_1, the second one is created using these $k+1$ vertices but v_2. So these exceptional vertices v_1 and v_2 are participating only in one of the new cliques, which implies that their degrees do not change. The vertices of the deleted k-clique, except v_1 and v_2, are included in one more k-clique so their degrees are also increased by $k-1$. In conclusion, the degree of an existing vertex does not change or it is increased by $k-1$. This increase happens whenever the chosen k vertices does not form a k-clique or if the sub-graph with these vertices is a k-clique and the considered vertex is not one of the two exceptional vertices. The new vertex is participating in two cliques immediately, its initial degree is $2(k-1)$.

Let use the notation A_n for the event that in the n^{th} step v is among the k selected vertices, which do not form a k-clique or they form a k-clique but v is not one of the exceptional vertices. Then the degree of v is increased by $k-1$.

To sum up, for the degree of the initial vertices $v = -(k-1), -(k-2), \ldots, 0$ we have

$$d_n(v) = \sum_{i=1}^{n} (k-1)\mathbb{I}_{A_i}. \tag{2.4.1}$$

In the following theorem, we calculate the asymptotic degree of the vertices.

Theorem 2.4.1. *For the degree of vertices we have*

$$d_n(v) \sim k(k+1)\left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} n^{\frac{k-1}{k+1}} \tag{2.4.2}$$

almost surely as $n \to \infty$ *.*

Proof. For $v = 1, 2, \ldots$ we have

$$d_n(v) = 2(k-1) \cdot \mathbb{I}_{\{V_{n-1} \geq k+v\}} + \sum_{i=1}^{n} (k-1)\mathbb{I}_{A_i}, \tag{2.4.3}$$

where A_i is the event that $V_{i-1} \geq k+v$ (the vertex v already exists), and at the i^{th} step v is among the k selected vertices, but it is not one of the two exceptional vertices. Then

$$\mathbb{P}(A_i|\mathcal{F}_{i-1}) = \mathbb{I}_{\{V_{i-1} \geq k+v\}} \cdot \frac{\frac{d_{i-1}(v)}{k-1} \cdot \frac{\binom{k-1}{2}}{\binom{k}{2}} + \left[\binom{V_{i-1}-1}{k-1} - \frac{d_{i-1}(v)}{k-1}\right]}{\binom{V_{i-1}}{k}} =$$

$$= \mathbb{I}_{\{V_{i-1} \geq k+v\}} \cdot \frac{\frac{d_{i-1}(v)}{k-1}\left[\frac{\binom{k-1}{2}}{\binom{k}{2}} - 1\right] + \binom{V_{i-1}-1}{k-1}}{\binom{V_i-1}{k}} \leq \frac{\binom{V_{i-1}-1}{k-1}}{\binom{V_{i-1}}{k}} = \frac{k}{V_{i-1}}. \tag{2.4.4}$$

So it follows that

$$\sum_{i=1}^{n} \mathbb{P}(A_i|\mathcal{F}_{i-1}) \le ck \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \sum_{i=1}^{n} i^{-\frac{2}{k+1}} \le$$
$$\le ck \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \int_0^n x^{-\frac{2}{k+1}}\, dx = ck \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \frac{k+1}{k-1} \left[x^{\frac{k-1}{k+1}}\right]_0^n =$$
$$= c\frac{k(k+1)}{k-1} \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} n^{\frac{k-1}{k+1}} = \mathcal{O}\left(n^{\frac{k-1}{k+1}}\right).$$

Now the generalization of the Borel-Cantelli lemma (see Corollary VII-2-6 of [62]) implies that $\sum_{i=1}^{n} \mathbb{I}_{A_i} = \mathcal{O}\left(n^{\frac{k-1}{k+1}}\right)$, which also means that $d_n(v) = \mathcal{O}\left(n^{\frac{k-1}{k+1}}\right)$. In the first term of (2.4.4), by (2.3.1) we have

$$\frac{d_{n-1}(v)}{\binom{V_{n-1}}{k}} \sim \frac{\mathcal{O}(n^{\frac{k-1}{k+1}})}{\left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}} n^{\frac{2k}{k+1}}} \to 0. \tag{2.4.5}$$

If $n \to \infty$, then $\mathbb{I}_{\{V_{n-1} \ge k+v\}} \to 1$ a.s. in (2.4.3), therefore relations (2.4.4), (2.4.5) and (2.3.1) give

$$\mathbb{P}(A_n \mid \mathcal{F}_{n-1}) \sim \frac{\binom{V_{n-1}-1}{k-1}}{\binom{V_{n-1}}{k}} \sim$$
$$\sim \frac{k!}{(k-1)!} \left[\frac{(k+1)!}{2}\right]^{-\frac{1}{k+1}} \frac{n^{\frac{2(k-1)}{k+1}}}{n^{\frac{2k}{k+1}}} = k \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \frac{1}{n^{\frac{2}{k+1}}}.$$

By using the above formulas, and the generalization of the Borel-Cantelli lemma, it follows that

$$d_n(v) \sim \sum_{i=1}^{n} (k-1)\mathbb{I}_{A_i} \sim \sum_{i=1}^{n} (k-1)\mathbb{P}(A_i \mid \mathcal{F}_{i-1}) \sim$$
$$\sim (k-1)k \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \sum_{i=1}^{n} \frac{1}{i^{\frac{2}{k+1}}} \sim (k-1)k \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \int_0^n x^{-\frac{2}{k+1}}\, dx \sim$$
$$\sim (k-1)k \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \frac{k+1}{k-1} \cdot n^{\frac{k-1}{k+1}} \sim k(k+1) \left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} \cdot n^{\frac{k-1}{k+1}}.$$

The previous calculation is also true for the asymptotic behaviour of the initial vertices. We can start with the formula (2.4.1) and then the rest of the proof also works if we omit the indicator in (2.4.4). □

The following result for $k = 2$ implies the asymptotic normality of the degree of a fixed vertex (see [5]).

Remark 2.4.1. *As $n \to \infty$, we have*

$$\frac{1}{n^{\frac{k-1}{2(k+1)}}}\left(d_n(v) - \sum_{i=1}^{n}(k-1)\mathbb{E}(\mathbb{I}_{A_i} \mid \mathcal{F}_{i-1})\right) \Rightarrow$$
$$\Rightarrow \mathcal{N}\left(0,\ (k-1)k(k+1)\left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}}\right). \quad (2.4.6)$$

For the proof, let us consider the martingale difference

$$X_{n,i} = \frac{1}{n^{\frac{k-1}{2(k+1)}}}\left(\mathbb{I}_{A_i} - \mathbb{E}(\mathbb{I}_{A_i} \mid \mathcal{F}_{i-1})\right),$$

with $\mathcal{F}_{n,i} = \mathcal{F}_i$, $n \geq 1$, $1 \leq i \leq n$. For the sums of conditional variances we get

$$\frac{1}{n^{\frac{k-1}{k+1}}}\sum_{i=1}^{n}\mathbb{E}\left(X_{n,i}^2 \mid \mathcal{F}_{i-1}\right) = \frac{1}{n^{\frac{k-1}{k+1}}}\sum_{i=1}^{n}\left[\mathbb{E}(\mathbb{I}_{A_i} \mid \mathcal{F}_{i-1}) - (\mathbb{E}(\mathbb{I}_{A_i} \mid \mathcal{F}_{i-1}))^2\right] \sim$$
$$\sim \frac{1}{n^{\frac{k-1}{k+1}}}\sum_{i=1}^{n}\mathbb{E}(\mathbb{I}_{A_i} \mid \mathcal{F}_{i-1}) \sim \frac{k(k+1)}{k-1}\left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}}.$$

Since $|X_{n,i}| \leq \frac{2}{n^{\frac{k-1}{2(k+1)}}}$, the conditional Lindeberg condition is also satisfied, so we can apply Corollary 3.1 of [40]. It implies that

$$\sum_{i=1}^{n}X_{n,i} = \frac{1}{n^{\frac{k-1}{2(k+1)}}}\sum_{i=1}^{n}\left(\mathbb{I}_{A_i} - \mathbb{E}(\mathbb{I}_{A_i} \mid \mathcal{F}_{i-1})\right) \Rightarrow \mathcal{N}\left(0,\ \frac{k(k+1)}{k-1}\left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}}\right),$$

and if we multiply by $k-1$, we obtain (2.4.6).

2.5 Functional limit theorems

In this section, we prove a functional limit theorem (invariance principle) for the number of vertices in our graphs (Theorem 2.5.1) and multidimensional functional limit theorems for the degrees of vertices (Proposition 2.5.1 and Theorem 2.5.2). Let $[nt]$ denote the integer part of nt. To study the number of vertices, we shall consider the process $V_{[nt]}$, $t \in [0,1]$, where V_n is the number of vertices. The trajectories of $V_{[nt]}$ belong to the space $D[0,1]$, i.e. they have no discontinuity of the second kind.

Theorem 2.5.1.

$$\frac{\binom{V_{[nt]}}{k+1} - \binom{[nt]}{2}}{n^{\frac{2k+1}{k+1}}} \Rightarrow \int f(s)\,dW(s),$$

where

$$f(s) = \frac{\left(\frac{(k+1)!}{2}\right)^{\frac{k}{2k+1}}}{(k!)^{\frac{1}{2}}} \cdot s^{\frac{3k+1}{2(k+1)}},$$

$W(s)$, $s \in [0,1]$, is the Wiener process, and $\Rightarrow$ denotes weak convergence in the space $D[0,1]$ with respect to the Skorohod topology.

Proof. We shall apply Theorem A.1 of [1] with $\nu_n(t) = [nt]$, $t \in [0,1]$. Let us use the martingale difference array

$$X_{n,i} = \frac{1}{n^{\frac{2k+1}{k+1}}} \left[\binom{V_i}{k+1} - \binom{V_{i-1}}{k+1} - (i-1) \right],$$

where $n \geq 1$, $1 \leq i \leq n$. Now, we shall check conditions (i) and (ii) of Theorem A.1 of [1]. The requirement (i) is the conditional Lindeberg condition, that was seen earlier in the proof of Theorem 2.3.1. According to our previous calculations, it is satisfied. For the condition (ii), by using the proof of Theorem 2.3.1 and the substitution $y = \frac{x}{n}$, we obtain

$$\sum_{i=1}^{[nt]} \mathbb{E}\left(X_{n,i}^2 \mid \mathcal{F}_{n,i-1}\right) \sim \frac{1}{n^{\frac{4k+2}{k+1}}} \cdot \frac{1}{k!} \cdot \left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}} \int_0^{[nt]} x^{\frac{3k+1}{k+1}}\, dx \sim$$

$$\sim \frac{1}{n^{\frac{4k+2}{k+1}}} \cdot \frac{1}{k!} \cdot \left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}} \cdot n^{\frac{3k+1}{k+1}} \cdot n \int_0^t y^{\frac{3k+1}{k+1}}\, dy =$$

$$= \frac{1}{k!} \cdot \left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}} \int_0^t y^{\frac{3k+1}{k+1}}\, dy = \int_0^t \left(\frac{1}{\sqrt{k!}} \left[\frac{(k+1)!}{2}\right]^{\frac{k}{2(k+1)}} y^{\frac{3k+1}{2(k+1)}} \right)^2 dy,$$

so this condition is fulfilled with $f(x) = \frac{1}{\sqrt{k!}} \left[\frac{(k+1)!}{2}\right]^{\frac{k}{2(k+1)}} x^{\frac{3k+1}{2(k+1)}}$. Then Theorem A.1 of [1] implies our theorem. $\square$

From Theorem 2.5.1, with $t = 1$, we can obtain our previous result (2.3.2). To prove it, take $t = 1$, and we get that $\int_0^1 f(s)\, dW(s)$ has distribution $\mathcal{N}\left(0, \int_0^1 (f(s))^2\, ds\right)$, because the function f is not random. And the variance of this normal distribution is

$$\frac{1}{k!} \left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}} \int_0^1 s^{\frac{3k+1}{k+1}}\, ds = \frac{1}{k!} \cdot \frac{k+1}{2(2k+1)} \left[\frac{(k+1)!}{2}\right]^{\frac{k}{k+1}},$$

which implies (2.3.6) and from that we can obtain (2.3.2) as in the proof of Theorem 2.3.1.

Now, we turn to the degrees of vertices. We shall study the joint behaviour of m vertices, where m is a fixed positive integer.

Proposition 2.5.1. *Consider the following martingale differences for the vertex v_l, $l = 1, \ldots, m$:*

$$X_{n,i}^{(v_l)} = \frac{1}{n^{\frac{k-1}{2(k+1)}}} \left[\mathbb{I}_{A_i^{(v_l)}} - \mathbb{E}\left(\mathbb{I}_{A_i^{(v_l)}} | \mathcal{F}_{i-1} \right) \right], \tag{2.5.1}$$

where $A_i^{(v_l)}$ is the event that we choose the vertex v_l in the i^{th} step, but it is not one of the two exceptional vertices. Let us define for $t \in [0,1]$

$$Y_n^{(v_l)} = \sum_{i=1}^{[nt]} X_{n,i}^{(v_l)}.$$

Let v_1, v_2, …, v_m be different vertices. Then we have

$$\left(Y_n^{(v_1)}, \dots, Y_n^{(v_m)}\right) \Rightarrow \left(\int f\, dW_1, \dots, \int f\, dW_m\right), \tag{2.5.2}$$

where $W_1, \dots, W_m$ are independent Wiener processes, $f(x) = k^{\frac{1}{2}} \left[\frac{2}{(k+1)!}\right]^{\frac{1}{2(k+1)}} x^{-\frac{1}{k+1}}$ is the function in the condition (3.4) of [43], and $\Rightarrow$ denotes weak convergence with respect to the product Skorohod topology in the space $D[0,1] \times \dots \times D[0,1]$.

Proof. We use Theorem 3.3 (a) of [43] for this proof. First, we check condition (3.11) of [43]. For any pair of vertices v_1 and v_2 ($v_1 \neq v_2$), we have that

$$\sum_{i=1}^{[nt]} \mathbb{E}\left(X_{n,i}^{(v_1)} X_{n,i}^{(v_2)} \mid \mathcal{F}_{i-1}\right) =$$
$$= \frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} \mathbb{E}\left\{\left[\mathbb{I}_{A_i^{(v_1)}} - \mathbb{E}\left(\mathbb{I}_{A_i^{(v_1)}} | \mathcal{F}_{i-1}\right)\right]\left[\mathbb{I}_{A_i^{(v_2)}} - \mathbb{E}\left(\mathbb{I}_{A_i^{(v_2)}} | \mathcal{F}_{i-1}\right)\right] \mid \mathcal{F}_{i-1}\right\} =$$
$$= \frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} \left\{\mathbb{E}\left(\mathbb{I}_{A_i^{(v_1)}} \mathbb{I}_{A_i^{(v_2)}} \mid \mathcal{F}_{i-1}\right) - \mathbb{E}\left(\mathbb{I}_{A_i^{(v_1)}} \mid \mathcal{F}_{i-1}\right) \mathbb{E}\left(\mathbb{I}_{A_i^{(v_2)}} \mid \mathcal{F}_{i-1}\right)\right\}. \tag{2.5.3}$$

For the first term in (2.5.3), we have

$$0 \le \mathbb{P}\left(A_i^{(v_1)} A_i^{(v_2)} \mid \mathcal{F}_{i-1}\right) = \frac{\binom{V_{i-1}-2}{k-2} - d_i(v_1, v_2) + d_i(v_1, v_2) \cdot \frac{\binom{k-2}{2}}{\binom{k}{2}}}{\binom{V_{i-1}}{k}} =$$
$$= \frac{\binom{V_{i-1}-2}{k-2}}{\binom{V_{i-1}}{k}} - \frac{d_i(v_1, v_2)\left(1 - \frac{(k-2)(k-3)}{k(k-1)}\right)}{\binom{V_{i-1}}{k}} \le \frac{\binom{V_{i-1}-2}{k-2}}{\binom{V_{i-1}}{k}}. \tag{2.5.4}$$

Since

$$0 \le \mathbb{E}\left(\mathbb{I}_{A_i^{(v_l)}} \mid \mathcal{F}_{i-1}\right) \le \frac{k}{V_{i-1}} \quad (l = 1, \dots m),$$

and $\binom{V_{i-1}}{k} \sim \frac{V_i^k}{k!}$, for the sum of the second terms in (2.5.3), we obtain

$$
\begin{aligned}
0 \leq \frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} \mathbb{E}\left(\mathbb{I}_{A_i^{(v_1)}} \mid \mathcal{F}_{i-1}\right) \mathbb{E}\left(\mathbb{I}_{A_i^{(v_2)}} \mid \mathcal{F}_{i-1}\right) \leq \frac{1}{n^{\frac{k-1}{k+1}}} k^2 \sum_{i=1}^{[nt]} \frac{1}{V_{i-1}^2} \sim \\
\sim \frac{1}{n^{\frac{k-1}{k+1}}} k^2 \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \sum_{i=1}^{[nt]} \frac{1}{i^{\frac{4}{k+1}}} \leq \\
\leq c \frac{1}{n^{\frac{k-1}{k+1}}} k^2 \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \int_0^{[nt]} x^{-\frac{4}{k+1}}\, dx = \\
= c \frac{k^2(k+1)}{k-3} \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \cdot \frac{1}{n^{\frac{k-1}{k+1}}} [nt]^{\frac{k-3}{k+1}} \leq \\
\leq c \frac{k^2(k+1)}{k-3} \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \cdot t^{\frac{k-3}{k+1}} \frac{1}{n^{\frac{2}{k+1}}} \to 0,
\end{aligned}
$$

if $n \to \infty$. Here we have also used the property $[nt] \leq nt$.

Using (2.5.4), for the sum of the first terms in (2.5.3), we have

$$
\begin{aligned}
\frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} \mathbb{E}\left(\mathbb{I}_{A_i^{(v_1)}} \mathbb{I}_{A_i^{(v_2)}} \mid \mathcal{F}_{i-1}\right) \leq (k-1)k \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} i^{-\frac{4}{k+1}} \leq \\
\leq (k-1)k \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \frac{1}{n^{\frac{k-1}{k+1}}} \int_0^{[nt]} x^{-\frac{4}{k+1}}\, dx = \\
= \frac{(k-1)k(k+1)}{k+3} \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} \frac{1}{n^{\frac{k-1}{k+1}}} [nt]^{\frac{k-3}{k+1}} \leq \\
\leq \frac{(k-1)k(k+1)}{k+3} \left[\frac{(k+1)!}{2}\right]^{-\frac{2}{k+1}} t^{\frac{k-3}{k+1}} n^{-\frac{2}{k+1}} \to 0
\end{aligned}
$$

if $n \to \infty$. So the condition (3.11) of [43] is fulfilled.

Now, turn to condition (3.4) of [43]. For any vertex v,

$$\sum_{i=1}^{[nt]} \mathbb{E}\left\{ \left(X_{n,i}^{(v)} \right)^2 \mid \mathcal{F}_{i-1} \right\} =$$

$$= \frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} \left\{ \mathbb{E}\left(\mathbb{I}_{A_i^{(v)}} \mid \mathcal{F}_{i-1} \right) - \left(\mathbb{E}\left(\mathbb{I}_{A_i^{(v)}} \mid \mathcal{F}_{i-1} \right) \right)^2 \right\} \sim$$

$$\sim \frac{1}{n^{\frac{k-1}{k+1}}} \sum_{i=1}^{[nt]} \mathbb{E}\left(\mathbb{I}_{A_i^{(v)}} \mid \mathcal{F}_{i-1} \right) \sim \frac{1}{n^{\frac{k-1}{k+1}}} k \left[\frac{2}{(k+1)!} \right]^{\frac{1}{k+1}} \sum_{i=1}^{[nt]} \frac{1}{i^{\frac{2}{k+1}}} \sim$$

$$\sim \frac{1}{n^{\frac{k-1}{k+1}}} k \left[\frac{2}{(k+1)!} \right]^{\frac{1}{k+1}} \int_{i=0}^{[nt]} x^{-\frac{2}{k+1}} \, dx \sim k \left[\frac{2}{(k+1)!} \right]^{\frac{1}{k+1}} \cdot \frac{k+1}{k-1} \cdot t^{\frac{k-1}{k+1}} =$$

$$= k \left[\frac{2}{(k+1)!} \right]^{\frac{1}{k+1}} \int_0^t x^{-\frac{2}{k+1}} \, dx = \int_0^t \left(k^{\frac{1}{2}} \left[\frac{2}{(k+1)!} \right]^{\frac{1}{2(k+1)}} x^{-\frac{1}{k+1}} \right)^2 dx,$$

so the condition (3.4) of [43] is also satisfied with the function $f(x) = k^{\frac{1}{2}} \left[\frac{2}{(k+1)!} \right]^{\frac{1}{2(k+1)}} x^{-\frac{1}{k+1}}$.

We can also obtain (3.5) of [43], i.e. the conditional Lindeberg condition in the same way as in the proof of Remark 2.4.1. Now, Theorem 3.3. (a) of [43] implies our statement. □

For $k = 2$, we can formulate the above proposition in explicit form.

Theorem 2.5.2. *Let $k = 2$ and let $v_1, v_2, \ldots, v_m$ be different vertices. Let*

$$Z_n^{(v_l)} = \frac{1}{n^{\frac{1}{6}}} \left(d_{[nt]}(v_l) - 2 \cdot 3^{\frac{2}{3}} [nt]^{\frac{1}{3}} \right), \quad t \in [0,1],$$

for $l = 1, \ldots, m$. Then we have

$$\left(Z_n^{(v_1)}, \ldots, Z_n^{(v_m)} \right) \Rightarrow \left(\int f \, dW_1, \ldots, \int f \, dW_m \right), \tag{2.5.5}$$

where $W_1, \ldots, W_m$ are independent Wiener processes, $f(x) = 2^{\frac{1}{2}} \cdot 3^{-\frac{1}{6}} \cdot x^{-\frac{1}{3}}$, and $\Rightarrow$ denotes weak convergence with respect to the product Skorohod topology in the space $D[0,1] \times \cdots \times D[0,1]$.

Moreover, we have

$$\left(V_n^{(v_1)}, \ldots, V_n^{(v_m)} \right) \Rightarrow \mathcal{N}_m \left(\mathbf{0}, 2 \cdot 3^{\frac{2}{3}} \mathbf{I}_m \right),$$

where

$$V_n^{(v_l)} = \frac{1}{n^{\frac{1}{6}}} \left(d_n(v_l) - 2 \cdot 3^{\frac{2}{3}} n^{\frac{1}{3}} \right), \quad l = 1, \ldots, m,$$

$\mathcal{N}_m\left(\mathbf{0}, 2\cdot 3^{\frac{2}{3}}\mathbf{I}_m\right)$ *denotes the* m*-dimensional multivariate normal distribution and* $\mathbf{I}_m$ *is the identity matrix of size* $m\times m$*. So the joint distribution of the degrees of vertices is asymptotically normal and the degrees of different vertices are asymptotically independent.*

Proof. In Proposition 2.5.1 with $k=2$, we have

$$Y_n^{(v_l)} = \sum_{i=1}^{[nt]} X_{n,i}^{(v_l)} = \frac{1}{n^{\frac{1}{6}}}\sum_{i=1}^{[nt]}\left(\mathbb{I}_{A_i^{(v_l)}} - \mathbb{E}(\mathbb{I}_{A_i^{(v_l)}}|\mathcal{F}_{i-1})\right) =$$

$$= \frac{1}{n^{\frac{1}{6}}}\left(\sum_{i=1}^{[nt]}\mathbb{I}_{A_i^{(v_l)}} - \sum_{i=1}^{[nt]}\mathbb{E}(\mathbb{I}_{A_i^{(v_l)}}|\mathcal{F}_{i-1})\right),$$

where

$$\sum_{i=1}^{[nt]}\mathbb{I}_{A_i^{(v_l)}} = d_{[nt]}\left(v_l\right)$$

and

$$\sum_{i=1}^{[nt]}\mathbb{E}(\mathbb{I}_{A_i^{(v_l)}}|\mathcal{F}_{i-1}) = 2\cdot 3^{\frac{2}{3}}[nt]^{\frac{1}{3}} + o\left(n^{\frac{1}{6}}\right),$$

implied by equation (3.4) of [5]. The function f is

$$f(x) = 2^{\frac{1}{2}}\cdot 3^{-\frac{1}{6}}\cdot x^{-\frac{1}{3}}.$$

Now, Proposition 2.5.1 implies the first part of our theorem.

To finish the proof, we need the stochastic integral of $f(x)$ with respect to the standard Wiener process. If we consider this integral in general, for functions of the form cx^{α}, then

$$\int_0^1 cx^{\alpha}\, dW(x) \approx c\sum_{l=0}^{n-1}\left(\frac{l}{n}\right)^{\alpha}\left[W\left(\frac{l+1}{n}\right) - W\left(\frac{l}{n}\right)\right].$$

Since $W\left(\frac{l+1}{n}\right) - W\left(\frac{l}{n}\right)$ is an increment of the standard Wiener process on an interval of length $\frac{1}{n}$, it has distribution $\mathcal{N}\left(0,\frac{1}{n}\right)$. Then the distribution of this integral is $\mathcal{N}\left(0, c^2/(2\alpha+1)\right)$.

In this case of $k=2$, we have $c = 2^{\frac{1}{2}}\cdot 3^{-\frac{1}{6}}$ and $\alpha = -\frac{1}{3}$, so we obtain the distribution $\mathcal{N}_m\left(\mathbf{0}, 2\cdot 3^{\frac{2}{3}}\mathbf{I}_m\right)$. Here we used that the Wiener processes $W_1,\ldots,W_m$ are independent. □

2.6 Simulation studies

First, we performed simulation studies for numerical support of Theorem 2.3.1.

Example 2.6.1. *For $k = 2$, we generated 10.000 evolution steps of our graph. Then we repeated it 1000 times. Figure 2.2 shows the sequence* $\frac{V_n}{\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}}$ *for* $n = 1, 2, \ldots, 10000$. *On the upper part of the figure, we can see one realization of the above sequence, on the lower part, we can see 1000 trajectories. Both figures show convergence to 1, that is they support formula (2.3.1).*

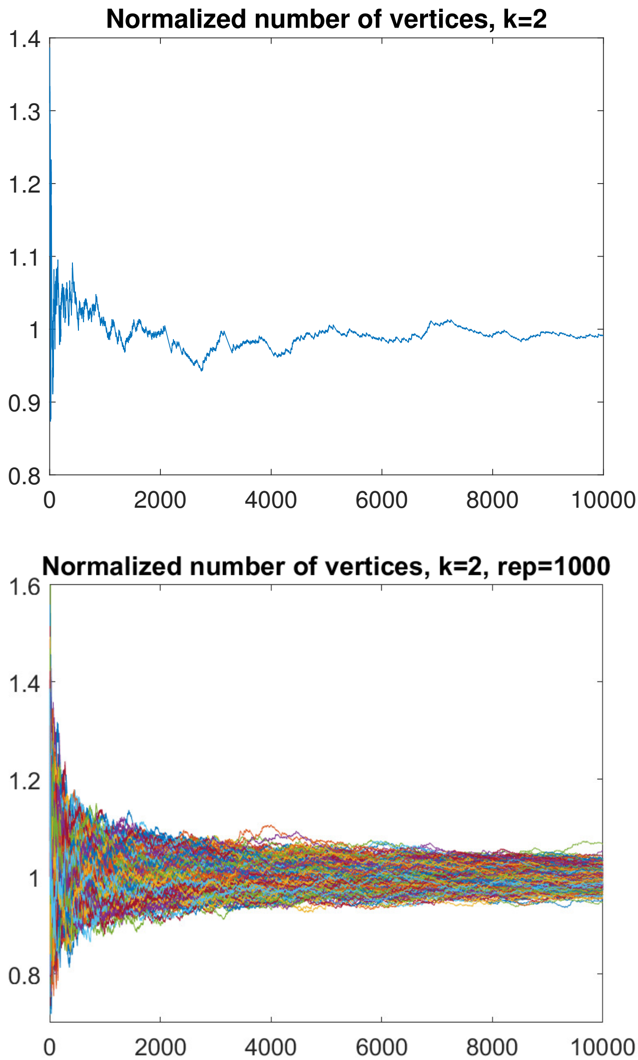


Figure 2.2: Generated realizations of formula (2.3.1), $k = 2$, $n = 1, 2, \ldots, 10000$. One realization at the top, 1000 realizations at the bottom.

Then we studied the asymptotic normality given by formula (2.3.2) *for* $k = 2$. *The sample was the* 1000 *realizations of* V_{10000}. *We standardized them according to formula* (2.3.2) *and constructed the density histogram. This density histogram is shown together with the standard normal density function at the top of Figure 2.3. The figure supports the asymptotic normality.*

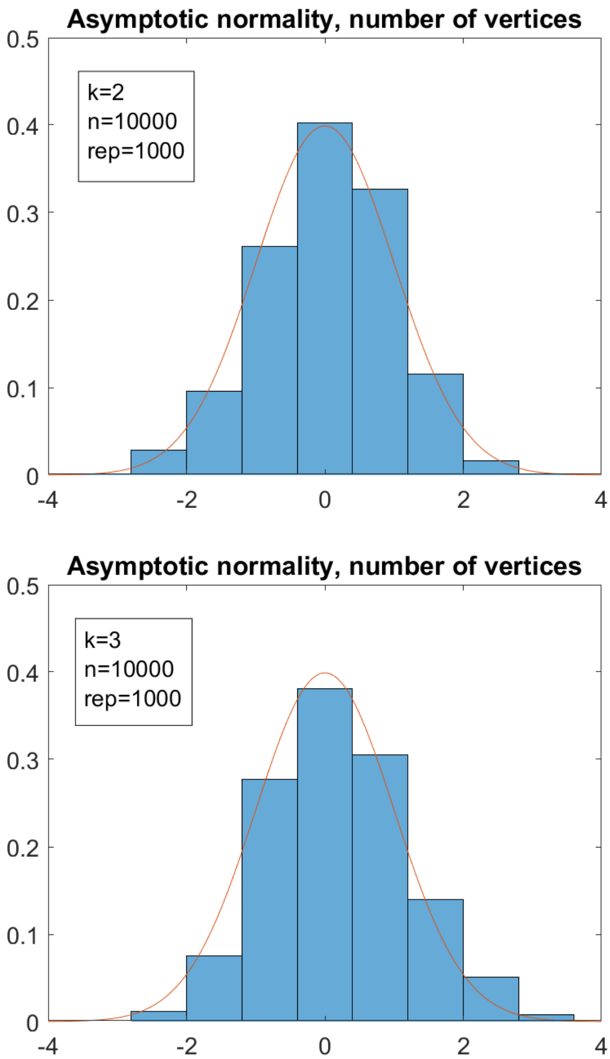


Figure 2.3: The histogram of 1000 realizations of V_{10000}, $k = 2$ at the top $k = 3$ at the bottom.

Example 2.6.2. *As in the previous case, we generated* 10.000 *evolution steps of our graph also for* $k = 3$. *Then we repeated the experiment* 1000 *times. Figure 2.4 shows the sequence* $\frac{V_n}{\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}}$ *for* $n = 1, 2, \ldots, 10000$. *On the upper part of the figure, one trajectory is presented, on the lower part,* 1000 *trajectories are visualized. Both figures show convergence to* 1, *so they support formula* (2.3.1).

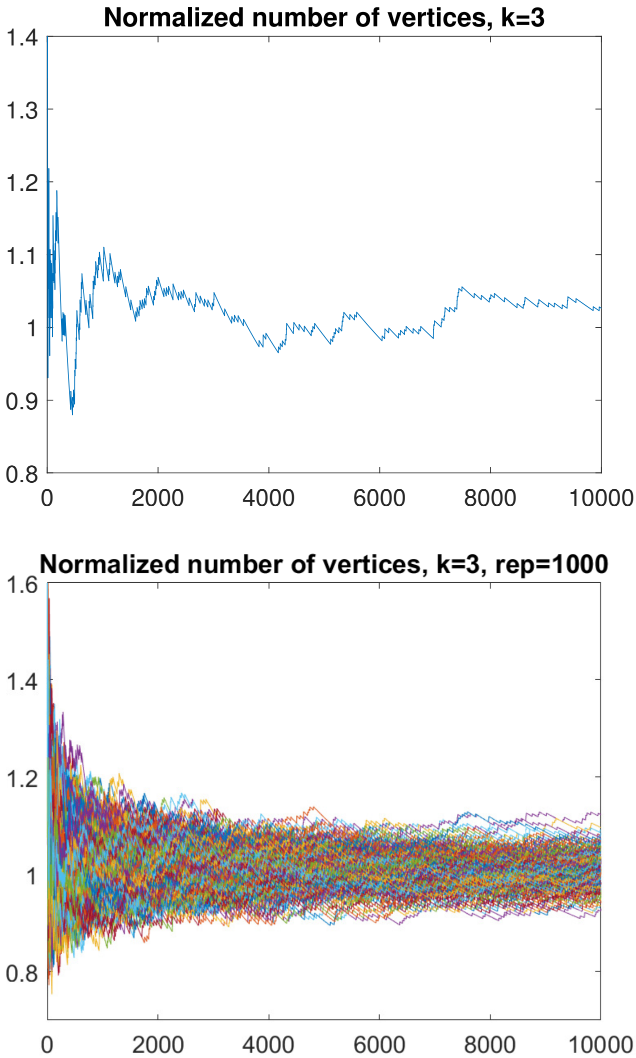


Figure 2.4: Generated realizations of formula (2.3.1), $k = 3$, $n = 1, 2, \ldots, 10000$. One realization above 1000 realizations below.

Then we studied the asymptotic normality given by formula (2.3.2) *with* $k = 3$. *The sample was the* 1000 *realizations of* V_{10000}. *We standardized them according to the formula* (2.3.2). *The density histogram is visualized together with the standard normal density function at the bottom of Figure 2.3. The figure shows asymptotic normality.*

Example 2.6.3. *To obtain numerical evidence for Theorem 2.4.1, we needed quite long computer experiments as the convergence is slow. For* $k = 2$ *we want to show the convergence*

$$d_n(v)/\left(2 \cdot 3^{\frac{2}{3}} n^{\frac{1}{3}}\right) \to 1$$

as $n \to \infty$. *We generated* 88.500.000 *evolution steps of our graph and visualized the sequence* $d_n(v)/\left(2 \cdot 3^{\frac{2}{3}} n^{\frac{1}{3}}\right)$, $n = 1, 2, \dots, 88.500.000$. *For large values of* n, *it should be close to* 1. *The computer experiment showed that this sequence approximates* 1 *slowly. Most of the sequences approximate 1 from below. The upper part of the Figure 2.5 shows the above sequence for* $v = 101$ *(solid black line),* $v = 301$ *(dashed blue line) and* $v = 1901$ *(dotted red line).*

Example 2.6.4. *To visualize Theorem 2.4.1 for* $k = 3$, *we performed* 20.000.000 *evolution steps. We want to show the convergence*

$$d_n(v)/\left(12^{\frac{3}{4}} n^{\frac{1}{2}}\right) \to 1 \tag{2.6.1}$$

as $n \to \infty$. *We visualized the sequence* $d_n(v)/\left(12^{\frac{3}{4}} n^{\frac{1}{2}}\right)$, $n = 1, 2, \dots, 20.000.000$. *For large values of* n, *it is close to* 1, *but the convergence to* 1 *is slow. The lower part of the Figure 2.5 shows the above sequence for* $v = 4$ *(solid blue line),* $v = 5$ *(dashed red line), and for* $v = 9$ *(dotted black line).*

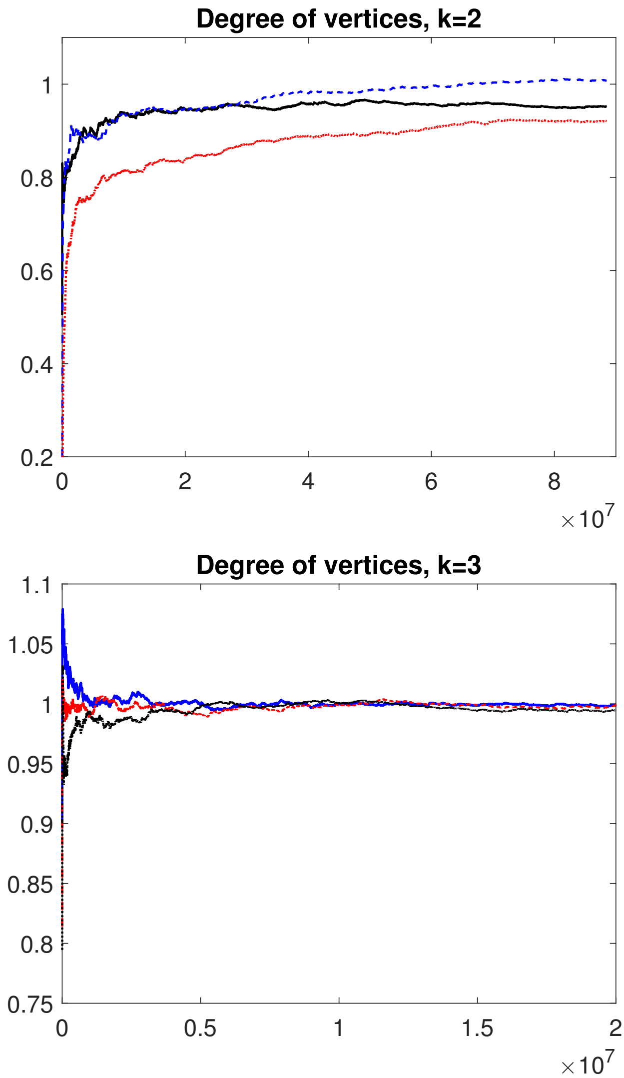


Figure 2.5: Generated realizations of the formula (2.4.2). On the upper part the normalized degree sequences of 3 vertices up to time 88.500.000 when $k = 2$. On the lower part the normalized degree sequences of 3 vertices up to time 20.000.000 when $k = 3$.

Chapter 3

On the convergence rate for the longest at most T-contaminated runs of heads

3.1 Introduction

In this chapter, we give a new, more precise approximation for the distribution of the length of the longest at most T-contaminated runs of heads. This chapter presents the results of our paper [26].

Consider the usual coin-tossing experiment. Let p be the probability of heads and $q = 1 - p$ be the probability of tails. Here p is a fixed number with $0 < p < 1$. We toss a coin N times independently. We write 1 for heads and 0 for tails. Therefore we consider independent identically distributed random variables $X_1, X_2, \ldots, X_N$ with distribution $P(X_i = 1) = p$ and $P(X_i = 0) = q = 1 - p$, $i = 1, 2, \ldots, N$.

Let T be a fixed non-negative integer. We shall study the length of at most T-contaminated (in other words at most T-interrupted) runs of heads. It means that there are at most T zeros in an m length sequence of ones and zeros.

There are several well-known results on the length of the pure head runs. Fair coins were studied in the paper of Erdős and Rényi [19]. Almost sure limit results for the length of the longest runs containing at most T tails were obtained in [21]. Földes [35] presented asymptotic results for the distribution of the number of T-contaminated head runs, the first hitting time of a T-contaminated head run having a fixed length, and the length of the longest T-contaminated head run. Móri [56] proved an almost sure limit theorem for the longest T-contaminated head run.

Gordon, Schilling, and Waterman [38] applied extreme value theory to obtain the asymptotic behaviour of the expectation and the variance of the length of the longest T-contaminated head run. Then accompanying distributions were obtained for the length of the longest T-contaminated head run. Novak [64] proved results on the accuracy of the approximation to the distribution of the length of the longest head run in a Markov chain.

We follow the lines of Arratia, Gordon, and Waterman [2], where Poisson approximation was used to find the asymptotic behaviour of the length of the longest at most T-contaminated head run. We use the basic results presented in [2], and give a new approximation for the distribution of the length of the longest at most T-contaminated head run. We show that for $T > 0$ the rate of the approximation in our new result is $\mathcal{O}\left(1/(\log(n))^2\right)$, where log denotes the logarithm to base $1/p$. We use the notation $f(n) = \mathcal{O}(h(n))$ for the property that $f(n)/h(n)$ is bounded as $n \to \infty$. We see that for $T > 0$ the rate of the approximation offered by [2] is $\mathcal{O}(\log(\log(n))/\log(n))$, so our result considerably improves the former result. In our opinion the much better rate $\mathcal{O}(\log(n)/n)$ presented without detailed proof in [2] is just a misprint, that is true only for $T = 0$. The main result is Theorem 3.3.1. For completeness, we also give a proof of the former result, see Proposition 3.2.1. In Section 3.4, we present some simulation results as well, to support our theorem.

For $T = 1$ and $T = 2$, our result is the same as the former result in [27], where a powerful lemma by Csáki, Földes and Komlós [16] was used in the proof.

3.2 The approximation of Arratia, Gordon and Waterman

Using the notation of [2], let $S_i = X_1 + \cdots + X_i$ and let $S_{n,t}$ be the largest increment in the sequence S_i in t steps, more precisely $S_{n,t}$ is the maximal number of heads in a window of length t starting in the first n tosses. Let $R_n(T)$ be the length of the longest at most T-interrupted runs of heads starting in the first n tosses. (One can see that $R_n(T)$ is the length of the longest precisely T-interrupted runs of heads starting in the first n tosses.) Then

$$\{R_n(T) < t\} = \{S_{n,t} < t - T\}.$$

According to Theorem 1 of [2] for the distribution of $S_{n,t}$, we have the following approximation. For positive integers n, s and t with $s \le t$, $s/t > p$,

$$|P(S_{n,t} < s) - e^{-EW}| \le 7tP(X_1 + \cdots + X_t = s) + P(X_1 + \cdots + X_t > s), \quad (3.2.1)$$

$$e^{-n\left(\frac{s}{t}-p\right)P(X_1+\cdots+X_t=s)} \cdot e^{-2n\left(1-\frac{s}{t}\right)P(X_1+\cdots+X_t=s)P(X_1+\cdots+X_t>s)} \le$$
$$\le e^{-EW} \le \quad e^{-n\left(\frac{s}{t}-p\right)P(X_1+\cdots+X_t=s)}. \quad (3.2.2)$$

In the above inequalities EW is the expectation of the random variable W defined in [2]. Here we do not use W itself, we only need the bounds above. We shall use inequalities (3.2.1) and (3.2.2) with $s = t - T$. Using notation $\alpha = n\left(\frac{s}{t} - p\right) P(X_1 + \cdots + X_t = s)$ and $\beta = 2n\left(1 - \frac{s}{t}\right) P(X_1 + \cdots + X_t = s) P(X_1 + \cdots + X_t > s)$, the above inequality is of the form

$$e^{-\alpha} e^{-\beta} \leq e^{-EW} \leq e^{-\alpha}. \tag{3.2.3}$$

Throughout this chapter, the approximation of $e^{-\alpha}$ will serve as the main term.

Now, we shall analyse that approximation of $R_n(T)$ which was proposed in [2]. The centering constant in [2] is

$$c_n(T) = \log n + T \log\log n - \log(T!) + \log(q^{T+1} p^{-T}). \tag{3.2.4}$$

Let x be a fixed number so that $c_n(T) + x = t$ is an integer. We want to estimate $P(R_n(T) - c_n(T) < x) = P(S_{n,t} < t - T)$. In the following we shall use both $\exp(x)$ and e^x for the usual exponential function.

Proposition 3.2.1. *Let $[c_n(T)]$ be the integer part of $c_n(T)$ and $\{c_n(T)\} = c_n(T) - [c_n(T)]$ be its fractional part.*

If $T = 0$, then for any integer l,

$$P(R_n(T) - [c_n(0)] < l) = \exp\left(-p^{l - \{c_n(0)\}}\right)\left(1 + \mathcal{O}\left(\frac{\log n}{n}\right)\right). \tag{3.2.5}$$

If $T > 0$, then for any integer l,

$$P(R_n(T) - [c_n(T)] < l) = \exp\left(-p^{l - \{c_n(T)\}}\right)\left(1 + \mathcal{O}\left(\frac{\log\log n}{\log n}\right)\right). \tag{3.2.6}$$

Remark 3.2.1. *In Corollary 3 of [2], the same remainder term $\mathcal{O}\left(\frac{\log n}{n}\right)$ is given for the case $T > 0$, too. However, in our opinion, it contains only a part of the remainder terms.*

Proof of Proposition 3.2.1. As our remainder term and the remainder term offered by [2] are different, we give the details of the more or less simple calculation. First, we calculate the right hand side of inequality (3.2.1) for $s = t - T$ and $t = c_n(T) + x$, where x is chosen so that t is an integer.

$$P(X_1 + \cdots + X_t = t - T) = \binom{t}{T} p^t (q/p)^T \leq \kappa \frac{(\log n)^T}{(1/p)^{\log n + T \log\log n}} = \mathcal{O}\left(\frac{1}{n}\right).$$

Here and in what follows κ is an appropriate finite positive constant. Therefore

$$7tP(X_1 + \cdots + X_t = t - T) = \mathcal{O}\left(\frac{\log n}{n}\right).$$

For $T > 0$, we have

$$\begin{aligned} P(X_1 + \cdots + X_t > t - T) &\leq T\binom{t}{t - T + 1} p^{t-T+1} \leq \kappa t^{T-1} p^t \\ &\leq \kappa \frac{(\log n)^{T-1}}{n(\log n)^T} = \mathcal{O}\left(\frac{1}{n \log n}\right). \end{aligned}$$

So we obtain

$$|P(S_{n,t} < t - T) - e^{-EW}| = \mathcal{O}\left(\frac{\log n}{n}\right). \tag{3.2.7}$$

This last formula is valid for $T = 0$, too.

Now we turn to the other parts of the approximation. First consider $T = 0$. Then the main term of the approximation, i.e. $e^{-\alpha}$ in formula (3.2.3) is

$$e^{-\alpha} = e^{-n\left(\frac{t}{t}-p\right)P(X_1+\cdots+X_t=t)} = e^{-p^{-\log(nq)+t}}.$$

We have to approximate $P(R_n(0) - [c_n(0)] < l)$, where l is an integer, $c_n(0) = \log n + \log q$ and $[.]$ denotes the integer part. So we should apply the previous equality with $t = [c_n(0)] + l$, so we obtain

$$e^{-\alpha} = e^{-p^{l-\{c_n(0)\}}},$$

where $\{.\}$ denotes the fractional part. We see, that if $T = 0$, then $\beta = 0$, so in inequality (3.2.3) we have equality. So for $T = 0$ this part of the approximation is precise, i.e. the main term does not contain a remainder part.

Now we consider the approximation of the main term for $T > 0$.

$$e^{-\alpha} = e^{-n\left(\frac{t-T}{t}-p\right)P(X_1+\cdots+X_t=t-T)} = e^{-n\left(q-\frac{T}{t}\right)\binom{t}{T}q^T p^{t-T}}.$$

Now denote by L the base $1/p$ logarithm of the negative of the exponent, that is $L = \log \alpha$. So

$$L = \log n + \log(q - T/t) + \log(t(t-1)\ldots((t-T+1)) - \log T! + T\log q + T - t.$$

We shall use $t = c_n(T) + x$. Applying Taylor's expansion of the logarithm function, $\log(x_0 + y) = \log x_0 + \frac{y}{cx_0} - \frac{y^2}{2c\tilde{x}_0^2}$, where $\tilde{x}_0$ is between x_0 and $x_0 + y$, and where $c = \ln(1/p)$, we get

$$L = \log n + \log q - \frac{T}{cqt} - \mathcal{O}\left(\frac{1}{t^2}\right) + \log t^T - \frac{t^{T-1}\binom{T}{2}}{ct^T} + \mathcal{O}\left(\frac{1}{t^2}\right) -$$

$$- \log T! + T\log q + T - t =$$

$$= \log n + T\log t - \frac{1}{ct}\left(\frac{T}{q} + \binom{T}{2}\right) + \mathcal{O}\left(\frac{1}{t^2}\right) - \log T! + (T+1)\log q + T - t.$$

We insert $t = c_n(T) + x = \log n + T\log\log n + E$, where E is defined by the equation at hand so it does not depend on n. Using again Taylor's expansions of the logarithm function as $\log(x_0 + y) = \log x_0 + \frac{y}{cx_0} - \frac{y^2}{2cx_0^2} + \frac{y^3}{3c\tilde{x}_0^3}$, where $\tilde{x}_0$ is between x_0 and $x_0 + y$, and for the $1/t$ function, as $\frac{1}{x_0+y} = \frac{1}{x_0} - \frac{y}{x_0^2} + \frac{y^2}{\tilde{x}_0^3}$, where

$\tilde{x}_0$ is between x_0 and $x_0 + y$, we get

$$\begin{aligned}
L &= \log n + \\
&+ T\left(\log\log n + \frac{T\log\log n + E}{c\log n} - \frac{(T\log\log n + E)^2}{2c(\log n)^2} + \mathcal{O}\left(\frac{(\log\log n)^3}{(\log n)^3}\right)\right) - \\
&- \frac{1}{c}\left(\frac{T}{q} + \binom{T}{2}\right)\left(\frac{1}{\log n} - \frac{T\log\log n + E}{(\log n)^2} + \mathcal{O}\left(\frac{(\log\log n)^2}{(\log n)^3}\right)\right) + \\
&+ \mathcal{O}\left(\frac{1}{t^2}\right) - \log T! + (T+1)\log q + T - t.
\end{aligned}$$

Now, using that $t = c_n(T) + x$ and inserting the value of $c_n(T)$, we obtain

$$L = -x + \frac{T^2\log\log n}{c\log n} + \mathcal{O}\left(\frac{1}{\log n}\right),$$

which implies that

$$L = -x + \mathcal{O}\left(\frac{\log\log n}{\log n}\right),$$

and this rate is not improvable. We remark, that this relation is valid for $T = 1$, too.

Therefore, by applying the Taylor series expansion $e^y = 1 + y + e^{\tilde{y}}\frac{y^2}{2}$ twice, where $\tilde{y}$ is between 0 and y, we obtain

$$e^{-\alpha} = e^{-(1/p)^L} = e^{-p^x}\left(1 - \ln\left(\frac{1}{p}\right)\frac{T^2\log\log n}{c\log n} + \mathcal{O}\left(\frac{1}{\log n}\right)\right) \tag{3.2.8}$$

$$= e^{-p^{l-\{c_n(T)\}}}\left(1 + \mathcal{O}\left(\frac{\log\log n}{\log n}\right)\right), \tag{3.2.9}$$

and this rate is not improvable.

Now we consider the $e^{-\beta}$ part. Here

$$\beta = 2\frac{T}{t}\sum_{i=t-T+1}^{t}\binom{t}{i}p^i q^{t-i} n\binom{t}{T}p^{t-T}q^T$$

with $t = c_n(T) + x = \log n + T\log\log n + E$. The largest term in the above sum is the first one, and it is

$$\binom{t}{T-1}p^t\left(\frac{q}{p}\right)^{T-1} = \mathcal{O}\left(\frac{1}{n\log n}\right).$$

Then

$$\binom{t}{T}p^{t-T}q^T = \mathcal{O}\left(\frac{1}{n}\right).$$

Using Taylor's expansion

$$\frac{T}{t} = \mathcal{O}\left(\frac{1}{\log n}\right).$$

So $\beta = \mathcal{O}(1/n(\log n)^2)$, and

$$e^{-\beta} = 1 - \mathcal{O}\left(\frac{1}{n(\log n)^2}\right).$$

Therefore

$$e^{-\alpha}e^{-\beta} = e^{-p^{l-\{c_n(T)\}}}\left(1+\mathcal{O}\left(\frac{\log\log n}{\log n}\right)\right)\left(1-\mathcal{O}\left(\frac{1}{n(\log n)^2}\right)\right) = e^{-p^{l-\{c_n(T)\}}}\left(1+\mathcal{O}\left(\frac{\log\log n}{\log n}\right)\right). \quad (3.2.10)$$

□

3.3 A new approximation

Theorem 3.3.1. *Let $T \geq 1$ be an integer. Let*

$$\begin{aligned} \tilde{c}_n(T) &= \log(qn) + T\log(\log(qn)) + \quad (3.3.1)\\ &+ T^2\frac{\log(\log(qn))}{c\log(qn)} - \frac{T}{cq_0\log(qn)} - \frac{T^3}{2c}\left(\frac{\log(\log(qn))}{\log(qn)}\right)^2 + \\ &+ T^2\frac{\log(\log(qn))}{cq_0(\log(qn))^2} + T^3\frac{\log(\log(qn))}{(c\log(qn))^2} + \\ &+ \left(T\log\left(\frac{q}{p}\right) - \log(T!)\right)\left(1+\frac{T}{c\log(qn)} - T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right), \end{aligned}$$

where log *denotes the logarithm to base $1/p$, $c = \ln(1/p)$,* ln *denotes the natural logarithm to base e, and $q_0 = \frac{2q}{2+Tq-q}$. Let $[\tilde{c}_n(T)]$ denote the integer part of $\tilde{c}_n(T)$, while $\{\tilde{c}_n(T)\}$ denotes the fractional part of $\tilde{c}_n(T)$, i.e. $\{\tilde{c}_n(T)\} = \tilde{c}_n(T) - [\tilde{c}_n(T)]$.*

Then

$$\begin{aligned} P(R_n(T) - [\tilde{c}_n(T)] < l) &= \quad (3.3.2)\\ &= \exp\left(-p^{(l-\{\tilde{c}_n(T)\})\left(1-\frac{T}{c\log(qn)}+T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right)}\right)\left(1+\mathcal{O}\left(\frac{1}{(\log n)^2}\right)\right) \end{aligned}$$

for any integer l.

Proof. We use the same approach as in the previous section. First, we calculate the right hand side of inequality (3.2.1) for $s = t - T$ and $t = \tilde{c}_n(T) + x$, where x is chosen so that t is an integer. As

$$\tilde{c}_n(T) = \log(n) + T\log(\log(n)) + \mathcal{O}(1),$$

so

$$P(X_1 + \cdots + X_t = t - T) = \binom{t}{T}p^t(q/p)^T \leq \kappa\frac{(\log n)^T}{(1/p)^{\log n + T\log\log n}} = \mathcal{O}\left(\frac{1}{n}\right).$$

Therefore

$$7tP(X_1+\cdots+X_t=t-T)=\mathcal{O}\left(\frac{\log n}{n}\right).$$

Similarly

$$P(X_1+\cdots+X_t>t-T)\le \kappa t^{T-1}p^t=\mathcal{O}\left(\frac{1}{n\log n}\right).$$

So

$$|P(S_{n,t}<t-T)-e^{-EW}|=\mathcal{O}\left(\frac{\log n}{n}\right). \tag{3.3.3}$$

Now, we turn to the approximation of the main term $e^{-\alpha}$. Denote by L again the base $1/p$ logarithm of the negative of the exponent, so

$$\begin{aligned} L &= \log\alpha = \\ &= \log n+\log(q-T/t)+\log(t(t-1)\dots((t-T+1))-\log T!+ \\ &+ T\log q+T-t. \end{aligned}$$

We shall apply it for $t=\tilde{c}_n(T)+x$. Therefore

$$\begin{aligned} L &= \\ &= \log\left(q-\frac{T}{t}\right)+\log n+\log\left(t^T-\frac{T(T-1)}{2}t^{T-1}+\mathcal{O}(t^{T-2})\right)-t+ \\ &\quad +\log((q/p)^T)-\log(T!)= \\ &= \log\left(q-\frac{T}{t}\right)+\log n+\log(t^T)-\frac{\frac{T(T-1)}{2}t^{T-1}}{ct^T}+\mathcal{O}\left(\frac{1}{t^2}\right)-t+ \\ &\quad +\log((q/p)^T)-\log(T!)= \\ &= \log q-\frac{T}{cqt}+\log n+T\log t-\frac{\frac{T(T-1)}{2}}{ct}-t+ \\ &\quad +\log((q/p)^T)-\log(T!)+\mathcal{O}\left(\frac{1}{(\log n)^2}\right)= \\ &= \log(qn)-\frac{T}{cq_0t}+T\log t-t+\log((q/p)^T)-\log(T!)+\mathcal{O}\left(\frac{1}{(\log n)^2}\right), \end{aligned}$$

where we applied Taylor's expansion of the log function up to second order and used the notation $q_0=\frac{2q}{2+Tq-q}$.

Introduce notation

$$\begin{aligned} D &= -\frac{T^3}{2c}\left(\frac{\log(\log(qn))}{\log(qn)}\right)^2+T^2\frac{\log(\log(qn))}{cq_0(\log(qn))^2}+T^3\frac{\log(\log(qn))}{(c\log(qn))^2}+ \\ &+ \left(T\log\left(\frac{q}{p}\right)-\log(T!)\right)\left(\frac{T}{c\log(qn)}-T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right), \end{aligned} \tag{3.3.4}$$

$$B=T^2\frac{\log(\log(qn))}{c\log(qn)}-\frac{T}{cq_0\log(qn)}+D \tag{3.3.5}$$

and

$$A = T\log(\log(qn)) + B.$$

Then $t = \tilde{c}_n(T) + x = \tilde{c}_n(T) + l - \{\tilde{c}_n(T)\}$, where l is an integer, so

$$t = T\log\left(\frac{q}{p}\right) - \log(T!) + \log(qn) + A + l - \{\tilde{c}_n(T)\}.$$

Inserting this value of t into the term $-t$ of L, we obtain

$$L \;=\; -\frac{T}{cq_0 t} + T\log t - A - l + \{\tilde{c}_n(T)\} + \mathcal{O}\left(\frac{1}{(\log n)^2}\right).$$

Then use Taylor's expansion for the function $1/t$ to get

$$\begin{aligned} L \;&=\; -\frac{T}{cq_0\log(qn)} + T^2\frac{\log(\log(qn))}{cq_0(\log(qn))^2} + \\ &+\; T\log\big(\log(qn) + T\log(\log(qn)) + B + \log((q/p)^T) - \log(T!) + l - \\ &-\; \{\tilde{c}_n(T)\}\big) - A - l + \{\tilde{c}_n(T)\} + \mathcal{O}\left(\frac{1}{(\log n)^2}\right). \end{aligned}$$

Now, by Taylor's expansion for the $\log(x)$ function, we obtain

$$\begin{aligned} L = &-\; \frac{T}{cq_0\log(qn)} + T^2\frac{\log(\log(qn))}{cq_0(\log(qn))^2} + T\log(\log(qn)) + \\ &+\; \frac{T\big(T\log(\log(qn)) + B + \log((q/p)^T) - \log(T!) + l - \{\tilde{c}_n(T)\}\big)}{c\log(qn)} - \\ &-\; \frac{1}{2}\frac{T\big(T\log(\log(qn)) + B + \log((q/p)^T) - \log(T!) + l - \{\tilde{c}_n(T)\}\big)^2}{c(\log(qn))^2} - \\ &-\; A - l + \{\tilde{c}_n(T)\} + \mathcal{O}\left(\frac{1}{(\log n)^2}\right). \end{aligned}$$

Now we can omit B from the quadratic term. Then we apply that $A =$

$T\log(\log(qn)) + B$, so we get

$$\begin{aligned}
L = & - \frac{T}{cq_0\log(qn)} + \frac{T^2\log(\log(qn))}{cq_0(\log(qn))^2} + \frac{T^2\log(\log(qn))}{c\log(qn)} \\
& + \frac{T(\log((q/p)^T) - \log(T!))}{c\log(qn)} + \frac{T^3\log(\log(qn))}{(c\log(qn))^2} - \frac{T^2}{q_0(c\log(qn))^2} + \\
& + \frac{TD}{c\log(qn)} + \frac{T(l - \{\tilde{c}_n(T)\})}{c\log(qn)} - \frac{1}{2}\frac{T^3(\log(\log(qn)))^2}{c(\log(qn))^2} - \\
& - \frac{1}{2}\frac{T\left(\log((q/p)^T) - \log(T!) + l - \{\tilde{c}_n(T)\}\right)^2}{c(\log(qn))^2} - \\
& - \frac{2T}{2}\frac{T\log(\log(qn))\left(\log((q/p)^T) - \log(T!) + l - \{\tilde{c}_n(T)\}\right)}{c(\log(qn))^2} - \\
& - B - l + \{\tilde{c}_n(T)\} + \mathcal{O}\left(\frac{1}{(\log n)^2}\right) = \\
= & (l - \{\tilde{c}_n(T)\})\left(\frac{T}{c\log(qn)} - \frac{T^2\log(\log(qn))}{c(\log(qn))^2} - 1\right) + \mathcal{O}\left(\frac{1}{(\log n)^2}\right).
\end{aligned}$$

So

$$e^{-\alpha} = e^{-p^{(l-\{\tilde{c}_n(T)\})\left(1-\frac{T}{c\log(qn)}+\frac{T^2\log(\log(qn))}{c(\log(qn))^2}\right)+\mathcal{O}\left(\frac{1}{(\log n)^2}\right)}}.$$

Using Taylor's expansion again,

$$e^{-\alpha} = e^{-p^{(l-\{\tilde{c}_n(T)\})\left(1-\frac{T}{c\log(qn)}+\frac{T^2\log(\log(qn))}{c(\log(qn))^2}\right)}}\left(1 + \mathcal{O}\left(\frac{1}{(\log n)^2}\right)\right).$$

Now turn to the $e^{-\beta}$ part, where

$$\beta = 2\frac{T}{t}\sum_{i=t-T+1}^{t}\binom{t}{i}p^i q^{t-i} n\binom{t}{T}p^{t-T}q^T$$

and $t = \tilde{c}_n(T) + x$. Simple calculations show that $\beta \leq \kappa(1/n(\log n)^2)$, and so

$$e^{-\beta} = 1 + \mathcal{O}\left(\frac{1}{n(\log n)^2}\right).$$

Therefore

$$e^{-\alpha}e^{-\beta} = e^{-p^{(l-\{\tilde{c}_n(T)\})\left(1-\frac{T}{c\log(qn)}+\frac{T^2\log(\log(qn))}{c(\log(qn))^2}\right)}}\left(1 + \mathcal{O}\left(\frac{1}{(\log n)^2}\right)\right).$$

□

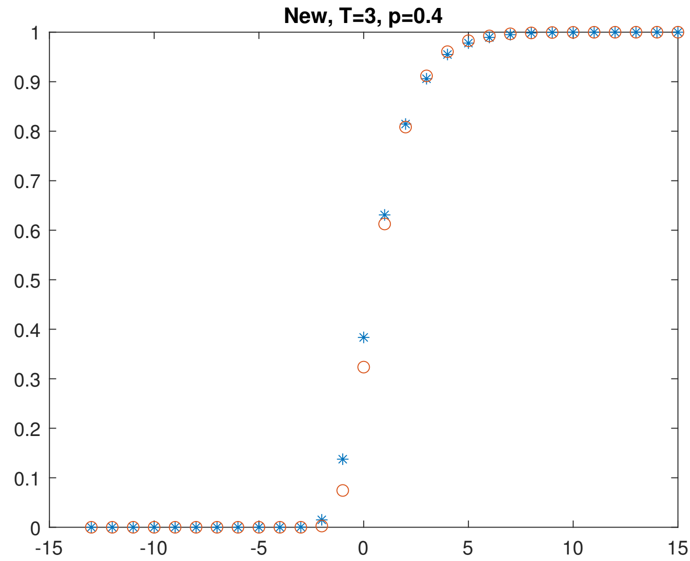
3.4 Simulation results

We performed several computer simulation studies for certain fixed values of p and T. Here we present the results of six simulations. The length of each simulated sequence was $N = 10^6$ and $s = 2000$ was the number of repetitions of the N-length sequences in each case.

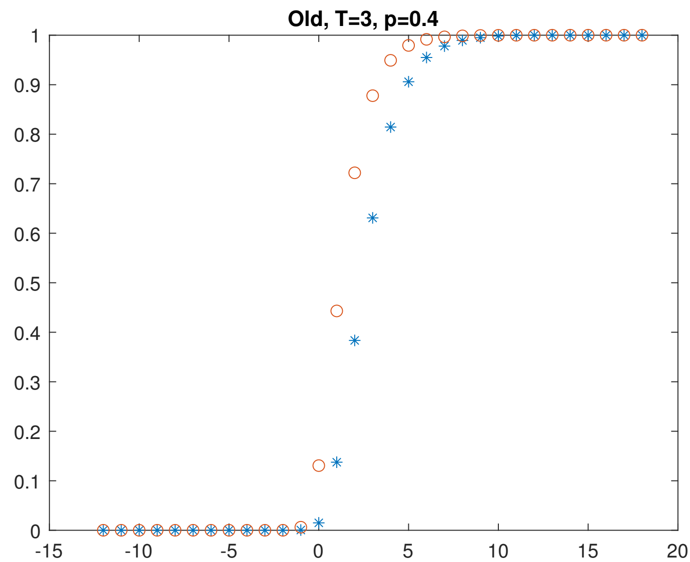
Figures 3.1, 3.2, and 3.3 present the results of the simulations when the number of contaminations was $T = 3$. The upper part of each figure shows the empirical distribution function of the longest at most T contaminated run and its approximation suggested by our Theorem 3.3.1. Asterisk (i.e. $*$) denotes the result of the simulation, i.e. the empirical distribution of the longest at most T contaminated run and circle ($\circ$) denotes the approximation offered by Theorem 3.3.1. The lower part of each figure shows the approximation by the former result. Asterisk denotes the result of the simulation again, and circle denotes approximation offered by Proposition 3.2.1. The simulation results support that our new theorem offers a better approximation than the previous one.

We also carried out experiments with 1 and 2-contaminations. Figure 3.4 shows the results of simulations with $T = 1$ contaminations and $p = 0.5$. Finally, figures 3.5 and 3.6 shows our simulation results with contaminations $T = 2$ with $p = 0.5$, $p = 0.6$ values, respectively.

For the simulation studies, we created our own programs. We used the Matlab software and we divided the task into two parts. The first program generated the data and collected the values of the lengths of the runs. The second program processed the results of the first one and created the figures. For the computer experiments, we used a laptop with "Intel Core i7 9th Gen" processor.

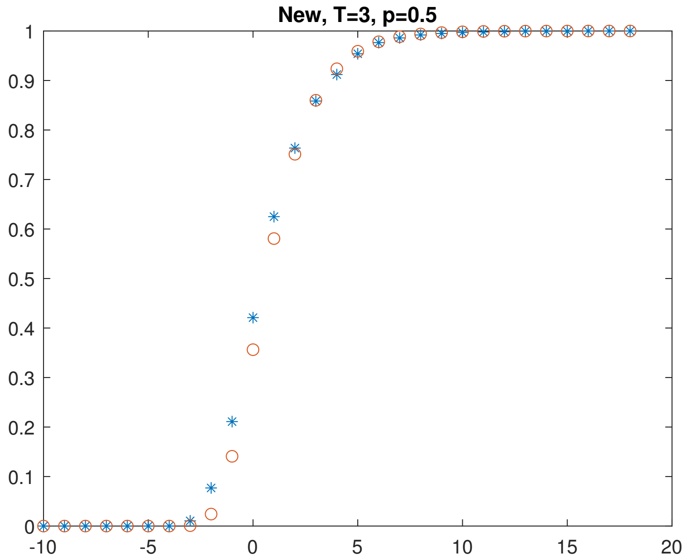


(a) New theorem

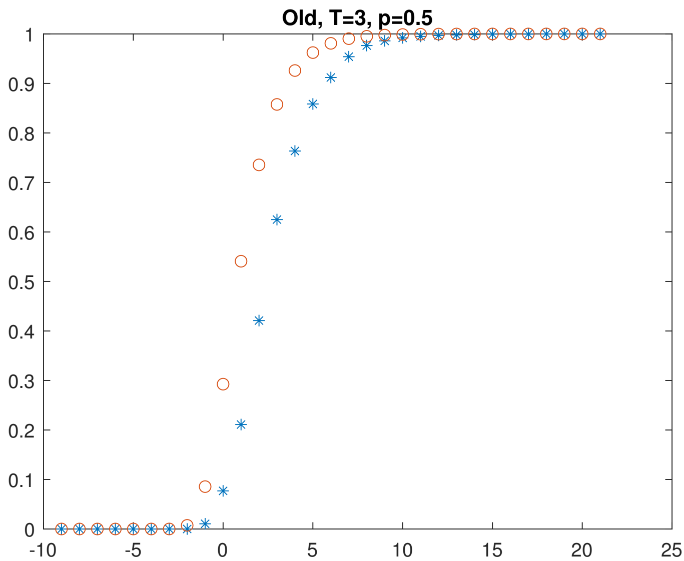


(b) Previous result

Figure 3.1: Longest at most $T = 3$ contaminated run when $p = 0.4$

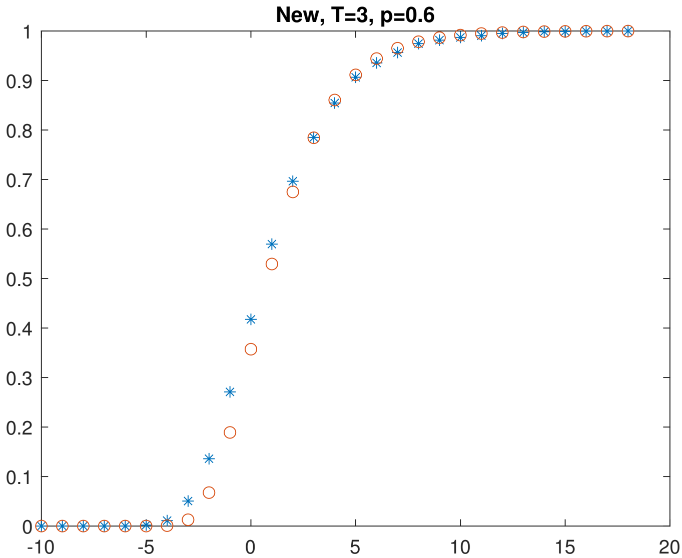


(a) New theorem

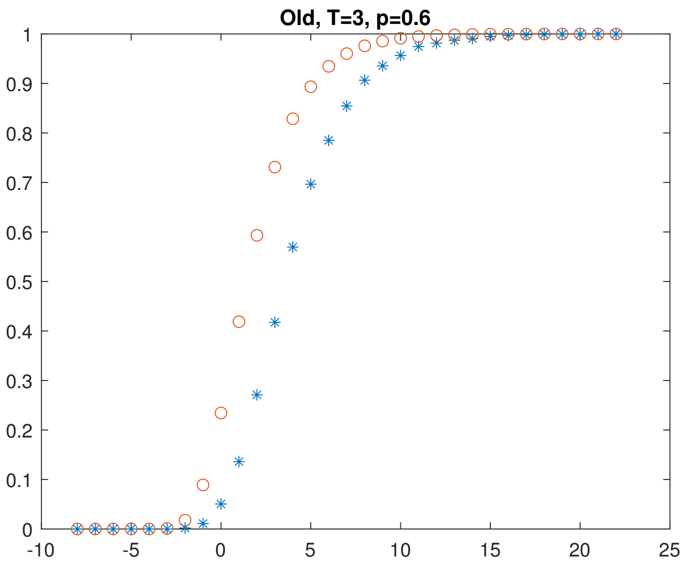


(b) Previous result

Figure 3.2: Longest at most $T = 3$ contaminated run when $p = 0.5$

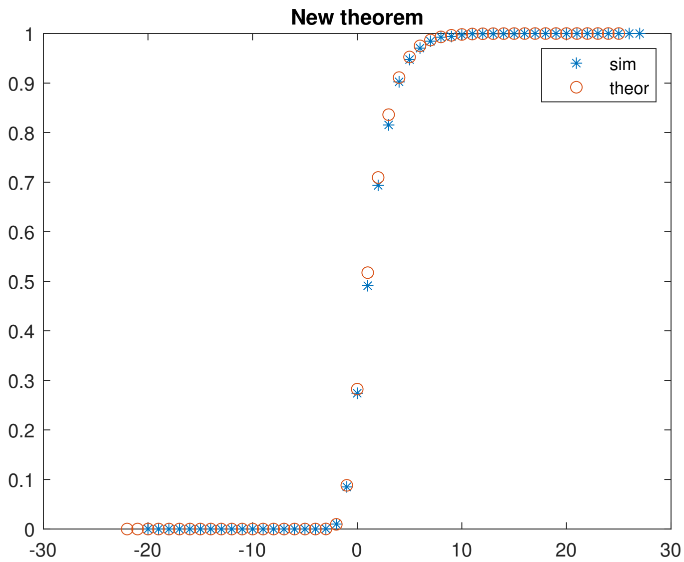


(a) New theorem

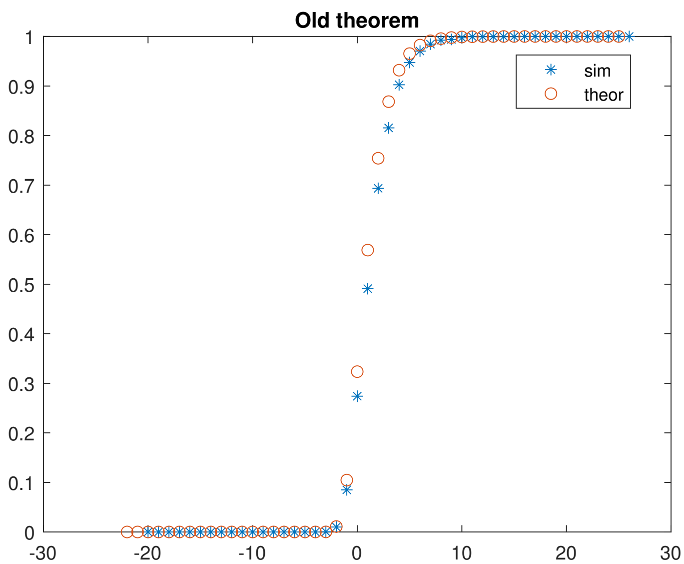


(b) Previous result

Figure 3.3: Longest at most $T = 3$ contaminated run when $p = 0.6$

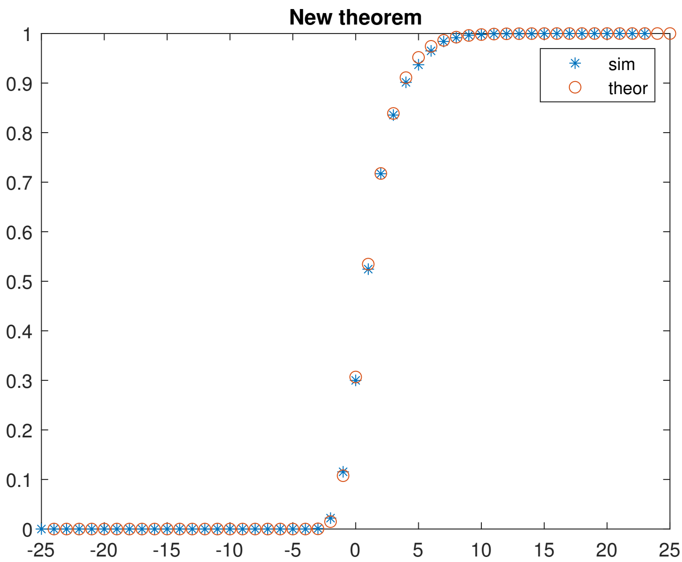


(a) New theorem

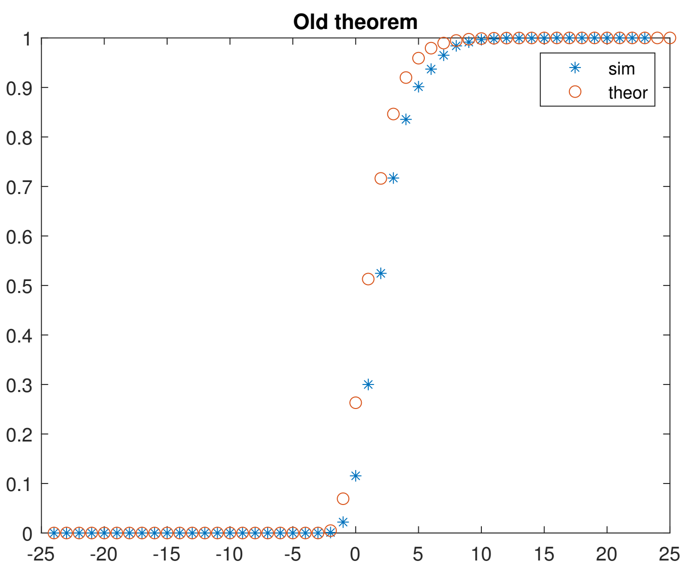


(b) Previous result

Figure 3.4: Longest at most $T = 1$ contaminated run when $p = 0.5$

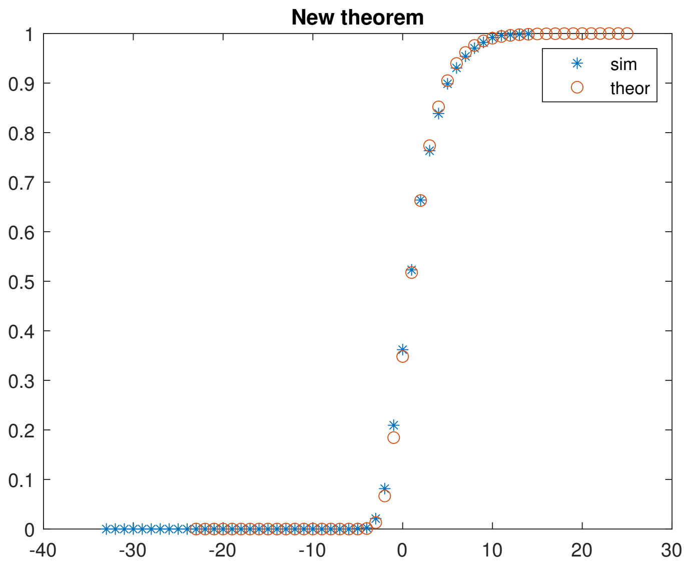


(a) New theorem

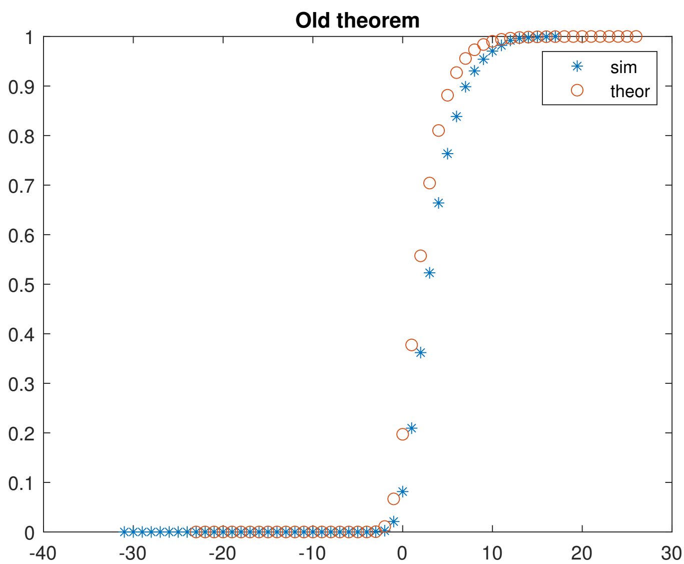


(b) Previous result

Figure 3.5: Longest at most $T = 2$ contaminated run when $p = 0.5$



(a) New theorem



(b) Previous result

Figure 3.6: Longest at most $T = 2$ contaminated run when $p = 0.6$

Summary

In the dissertation, we presented asymptotic results for some probabilistic models. We introduced two new network evolution models based on cliques of nodes. We considered a continuous time model that was governed by a general multi-type Crump-Mode-Jagers branching process. The type of the individuals was given by the clique size. We also studied a parametrized family of discrete-time network evolution models. The evolution of the graph was based on constructions and deletions of k-cliques. Finally, we showed a result in connection with coin tossing, we gave a new approximation for the distribution of the length of the longest at most T-contaminated head runs.

In the Introduction, we provided a brief review of the literature related to the topics and models of the thesis.

Chapter 1 includes the results of our paper [24]. We introduced a continuous-time network evolution model that described N-interactions, so we considered the cliques of nodes. The model was based on teams as individuals, i.e. cliques of size $1, 2, \ldots, N$, where N is an arbitrarily large but fixed number. We have used a multi-type branching process to describe the evolution of the network. In terms of multi-type branching processes, an n-clique is considered as an individual of type-n. At the initial time $t = 0$ there is one team, and the size of this team can be $1, 2, \ldots, N$. This team is called the ancestor. This ancestor team produces offspring teams which can also be cliques of sizes $1, 2, \ldots, N$. Then these children teams also produce their own children teams, and so on. Any team has its Poisson process of rate 1. These reproduction processes of the teams of size n are independent and identically distributed.

The mathematical description of the evolution of the generic n-clique is the following. Let $\Pi_n(t)$ denote the Poisson process with parameter 1 corresponding to the generic n-clique. The jumping times of $\Pi_n(t)$ are the reproduction times. When $\Pi_n(t)$ jumps, then a new vertex appears and we connect it to certain vertices of the generic n-clique. The new vertex will be connected to j vertices of the generic n-clique with probability $q_{n,j}$, where $0 \leq q_{n,j} \leq 1$, $j = 0, 1, \ldots, n$, and $\sum_{j=0}^{n} q_{n,j} = 1$. (We assume that $q_{N,N} = 0$ because the largest team is of size N.) The probability that an i-type ancestor produces a j-type child is $p_{i,j} = q_{i,j-1}$, $j = 1, 2, \ldots, i+1$. The life length λ_i of an i-clique is given by the $l_i(t) = b + c\xi_i(t)$

hazard rate, where $b \geq 0$, $c > 0$ are fixed constants and $\xi_i(t)$ is the number of all children of the generic i-clique up to time t. $m_{i,j}(t) = \mathbb{E}\xi_{i,j}(t)$ is the expected number of type-j offspring of a type-i parent up to time t. Let us use the notation M for the matrix of the $m^*_{ij}(t)$ Laplace transforms of the functions $m_{ij}(t)$.

After the mathematical description of our continuous-time network evolution model, the most important results were the following:

Assume that the matrix $(p_{i,j})_{i,j=1}^N$ is irreducible and acyclic. We introduce a matrix $A(\kappa)$ depending on a and b. Assume that $A(0) > 1$, that is the process is supercritical. Let α be the Malthusian parameter determined by the matrix M. If α is the Malthusian parameter, then we denote by $\mathbf{v} = (v_1, \ldots, v_N)^\top$ the right eigenvector of $M(\alpha)$ corresponding to eigenvalue 1 and normalized as $v_1 + \cdots + v_N = 1$. Let $\mathbf{u} = (u_1, \ldots, u_N)^\top$ be the left eigenvector of $M(\alpha)$ satisfying $u_1v_1 + \cdots + u_Nv_N = 1$. $A(\alpha)$ is given by the Laplace transforms and

$$-A'(\alpha) = D(\alpha) = \sum_{l,j=1}^{N} u_l v_j \left(-m^*_{l,j}(\alpha)\right)'.$$

In the following results, the quantity ${}_kW$ is a.s. non-negative, $\mathbb{E}({}_kW) = 1$, and ${}_kW$ is a.s. positive on the event of survival. Let n be fixed, $1 \leq n \leq N$.

1. Let us use the notation ${}_kT(t)$ for the number of all n-cliques being born up to time t if the ancestor of the network was a k-clique, $k = 1, 2, \ldots, N$. Then

$$\lim_{t\to\infty} e^{-\alpha t}{}_kT(t) = {}_kW \frac{v_k u_n}{\alpha\left(-A'(\alpha)\right)}$$

almost surely for $k = 1, 2, \ldots, N$.

2. Let us use the notation ${}_k\hat{T}(t)$ for the number of all n-cliques alive at time t if the ancestor of the network was a k-clique, $k = 1, 2, \ldots, N$. Then

$$\lim_{t\to\infty} e^{-\alpha t}{}_k\hat{T}(t) = {}_kW \frac{v_k u_n A(\alpha)}{\left(-A'(\alpha)\right)}$$

almost surely for $k = 1, 2, \ldots, N$.

Let $\mathbb{M}$ be the matrix of the expected total offspring number of our process. Then $\mathbb{M} = (m_{i,j}(\infty))_{i,j=1}^N = A(0)\,(p_{i,j}(\infty))_{i,j=1}^N$.

3. Let us denote by s_i the probability of extinction when the ancestor of the network is an i-type object. Let $\mathbf{s} = (s_1, \ldots, s_N)$. Assume that $(p_{i,j}(\infty))_{i,j=1}^N$ is an irreducible acyclic Markov transition matrix. Denote by ϱ the Perron–Frobenius root of $\mathbb{M}$. If $\varrho \leq 1$, then $s_1 = s_2 = \cdots = s_N = 1$. If $\varrho > 1$, then $s_1 < 1, s_2 < 1, \ldots, s_N < 1$. In any case, $\mathbf{s}$ is the smallest non-negative solution of the vector equation

$$\mathbf{s} = \mathbf{f}(\mathbf{s}),$$

where $\mathbf{f} = (f_1, \ldots, f_N)$ and the functions f_k are the generating functions of the offspring distributions (defined in (1.6.19)).

The first and second result describe the asymptotic behavior of the k-cliques being born and being alive at time t. We have similar results in [24] for the degree of a fixed vertex as well. We also have empirical support for our theorems, simulaton results were presented in Section 1.7.

Chapter 2 was based on our paper [29]. We studied a discrete-time network evolution model that is based on constructions and deletions of k-cliques, where $k \geq 2$ is a fixed integer.

The initial graph at time $n = 0$ contains k vertices and no edges. In the first step, i.e. when the time is $n = 1$, we connect the k vertices to obtain a single k-clique. The second step is the following. We choose two vertices uniformly at random, let us denote them by v_1 and v_2. Then we add a new vertex and construct two new k-cliques. The vertices of the first k-clique are the existing $k+1$ vertices but v_1, while the vertices of the second k-clique are the existing $k+1$ vertices but v_2. Then the original k-clique is deleted.

Later on, in each step, we choose k vertices uniformly at random from the existing vertices. If they do not form a k-clique, then we construct a new k-clique on these vertices (i.e. we connect them using $\binom{k}{2}$ new edges). In the other case, when the sub-graph consisting of the k vertices chosen is a k-clique, then that k-clique is deleted but its vertices are used to construct two new k-cliques as in the second step. That is a new vertex is added to the graph and using this new vertex and the k vertices of the just deleted k-clique, two new k-cliques are created in the same way as in the second step.

After the description of the evolution of the model, we obtained the limit of the number of vertices as well as the asymptotic normality of the number of vertices. We also calculated the asymptotic degree of the vertices.

4. As $n \to \infty$, the following almost sure convergence holds for the number of vertices in the graph after n steps:

$$\frac{V_n}{\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}} \to 1.$$

We also write this property as

$$V_n \sim \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}, \qquad \text{almost surely as} \quad n \to \infty.$$

5. We have

$$\frac{1}{n^{\frac{1}{k+1}}}\left(V_n - \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{2k+1}\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}}\right)$$

as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution.

6. If we use the notation $d_n(v)$ for the degree of a fixed vertex v at time n, then we have

$$d_n(v) \sim k(k+1)\left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} n^{\frac{k-1}{k+1}}.$$

We performed simulation studies as well, to support the previous three results.

We also proved a functional limit theorem (invariance principle) for the number of vertices in our graphs and multidimensional functional limit theorems for the degrees of vertices. Let $[nt]$ denote the integer part of nt. To study the number of vertices, we shall consider the process $V_{[nt]}$, $t \in [0,1]$, where V_n is the number of vertices. The trajectories of $V_{[nt]}$ belong to the space $D[0,1]$, i.e. they have no discontinuity of the second kind.

7. $$\frac{\binom{V_{[nt]}}{k+1} - \binom{[nt]}{2}}{n^{\frac{2k+1}{k+1}}} \Rightarrow \int f(s)\, dW(s),$$

   where

   $$f(s) = \frac{\left(\frac{(k+1)!}{2}\right)^{\frac{k}{2k+1}}}{(k!)^{\frac{1}{2}}} \cdot s^{\frac{3k+1}{2(k+1)}},$$

   $W(s)$, $s \in [0,1]$, is the Wiener process, and $\Rightarrow$ denotes weak convergence in the space $D[0,1]$ with respect to the Skorohod topology.

We also obtained multidimensional functional limit theorems for the degrees of vertices. We studied the joint behaviour of m vertices, where m is a fixed positive integer.

8. Consider the following martingale differences for the vertex v_l, $l = 1, \ldots, m$:

   $$X_{n,i}^{(v_l)} = \frac{1}{n^{\frac{k-1}{2(k+1)}}}\left[\mathbb{I}_{A_i^{(v_l)}} - \mathbb{E}\left(\mathbb{I}_{A_i^{(v_l)}} | \mathcal{F}_{i-1}\right)\right],$$

   where $A_i^{(v_l)}$ is the event that we choose the vertex v_l in the i^{th} step, but it is not one of the two exceptional vertices. Let us define for $t \in [0,1]$

   $$Y_n^{(v_l)} = \sum_{i=1}^{[nt]} X_{n,i}^{(v_l)}.$$

   Let $v_1, v_2, \ldots, v_m$ be different vertices. Then we have

   $$\left(Y_n^{(v_1)}, \ldots, Y_n^{(v_m)}\right) \Rightarrow \left(\int f\, dW_1, \ldots, \int f\, dW_m\right),$$

   where $W_1, \ldots, W_m$ are independent Wiener processes, $f(x) = k^{\frac{1}{2}}\left[\frac{2}{(k+1)!}\right]^{\frac{1}{2(k+1)}} x^{-\frac{1}{k+1}}$ is the function in the condition (3.4) of [43], and $\Rightarrow$ denotes weak convergence with respect to the product Skorohod topology in the space $D[0,1] \times \cdots \times D[0,1]$.

9. Let $k = 2$ and let $v_1, v_2, \ldots, v_m$ be different vertices. Let

$$Z_n^{(v_l)} = \frac{1}{n^{\frac{1}{6}}} \left(d_{[nt]}(v_l) - 2 \cdot 3^{\frac{2}{3}} [nt]^{\frac{1}{3}} \right), \quad t \in [0,1],$$

for $l = 1, \ldots, m$. Then we have

$$\left(Z_n^{(v_1)}, \ldots, Z_n^{(v_m)} \right) \Rightarrow \left(\int f \, dW_1, \ldots, \int f \, dW_m \right),$$

where $W_1, \ldots, W_m$ are independent Wiener processes, $f(x) = 2^{\frac{1}{2}} \cdot 3^{-\frac{1}{6}} \cdot x^{-\frac{1}{3}}$, and $\Rightarrow$ denotes weak convergence with respect to the product Skorohod topology in the space $D[0,1] \times \cdots \times D[0,1]$.

Moreover, we have

$$\left(V_n^{(v_1)}, \ldots, V_n^{(v_m)} \right) \Rightarrow \mathcal{N}_m \left(\mathbf{0}, 2 \cdot 3^{\frac{2}{3}} \mathbf{I}_m \right),$$

where

$$V_n^{(v_l)} = \frac{1}{n^{\frac{1}{6}}} \left(d_n(v_l) - 2 \cdot 3^{\frac{2}{3}} n^{\frac{1}{3}} \right), \quad l = 1, \ldots, m,$$

$\mathcal{N}_m \left(\mathbf{0}, 2 \cdot 3^{\frac{2}{3}} \mathbf{I}_m \right)$ denotes the m-dimensional multivariate normal distribution and $\mathbf{I}_m$ is the identity matrix of size $m \times m$. So the joint distribution of the degrees of vertices is asymptotically normal and the degrees of different vertices are asymptotically independent.

Our results (4.)–(6.) are extensions of the results of [5]. However, the results (7.)–(9.) are new for any value of k, including the particular case of $k = 2$ studied in [5]. (7.) is a functional limit theorem for the number of vertices in our graph. (8.) and (9.) are multidimensional functional limit results for the joint distribution of the degrees of several fixed vertices.

Chapter 3 contains the results of our paper [26]. We considered the coin tossing experiment. Let p be the probability of heads and $q = 1 - p$ be the probability of tails. Here p is a fixed number with $0 < p < 1$. Let us take independent identically distributed random variables $X_1, X_2, \ldots, X_N$ with distribution $P(X_i = 1) = p$ and $P(X_i = 0) = q = 1 - p$, $i = 1, 2, \ldots, N$.

We gave a new approximation for the distribution of the length of the longest at most T-contaminated head run. We showed that for $T > 0$ the rate of the approximation in our new result is $\mathcal{O}\left(1/(\log(n))^2\right)$, where log denotes the logarithm to base $1/p$. Let $R_n(T)$ be the length of the longest at most T-interrupted runs of heads starting in the first n tosses. The main result in this topic is the following:

10. Let $T \geq 1$ be an integer. Let

$$\begin{aligned}\tilde{c}_n(T) &= \log(qn) + T\log(\log(qn)) + \\ &+ T^2\frac{\log(\log(qn))}{c\log(qn)} - \frac{T}{cq_0\log(qn)} - \frac{T^3}{2c}\left(\frac{\log(\log(qn))}{\log(qn)}\right)^2 + \\ &+ T^2\frac{\log(\log(qn))}{cq_0(\log(qn))^2} + T^3\frac{\log(\log(qn))}{(c\log(qn))^2} + \\ &+ \left(T\log\left(\frac{q}{p}\right) - \log(T!)\right)\left(1 + \frac{T}{c\log(qn)} - T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right),\end{aligned}$$

where log denotes the logarithm to base $1/p$, $c = \ln(1/p)$, ln denotes the natural logarithm to base e, and $q_0 = \frac{2q}{2+Tq-q}$. Let $[\tilde{c}_n(T)]$ denote the integer part of $\tilde{c}_n(T)$, while $\{\tilde{c}_n(T)\}$ denotes the fractional part of $\tilde{c}_n(T)$, i.e. $\{\tilde{c}_n(T)\} = \tilde{c}_n(T) - [\tilde{c}_n(T)]$. Then

$$\begin{aligned}&P(R_n(T) - [\tilde{c}_n(T)] < l) = \\ &= \exp\left(-p^{(l-\{\tilde{c}_n(T)\})\left(1-\frac{T}{c\log(qn)}+T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right)}\right)\left(1+\mathcal{O}\left(\frac{1}{(\log n)^2}\right)\right)\end{aligned}$$

for any integer l.

We saw that for $T > 0$ the rate of the approximation offered by [2] was $\mathcal{O}\left(\log(\log(n))/\log(n)\right)$, so our result considerably improved the former result. Simulation studies were also presented to provide empirical support for our theorem.

Összefoglalás

A disszertációban aszimptotikus eredményeket mutattunk be néhány valószínűségi modellre vonatkozóan. Bevezettünk két új hálózatfejlődési modellt, amelyek klikkeken alapulnak, melyek alapegységei a klikkek. Vizsgáltunk egy folytonos idejű, általános, többtípusú Crump-Mode-Jagers-féle elágazó folyamat által vezérelt modellt. Az egyedek típusát a klikkméret adta. Diszkrét idejű hálózatfejlődési modelleknek egy paraméteres családját is tanulmányoztuk. A gráf fejlődése itt k-klikkek konstrukcióján és törlésén alapult. Végül egy érmedobással kapcsolatos eredményt is bemutattunk: adtunk egy új közelítést a leghosszabb, legfeljebb T-szennyezett fej sorozatok hosszának eloszlására.

A Bevezetésben áttekintettük a dolgozat témáival, illetve az abban szereplő modellekkel kapcsolatos irodalmi előzményeket.

Az 1. fejezet [24] cikkünk eredményeit tartalmazza. Bemutattunk egy folytonos idejű hálózatfejlődési modellt, amely N-interakciókat ír le, így a csúcsok által alkotott klikkeket tekintettük. A modell alapegységei, egyedei a csapatok, azaz $1, 2, \ldots, N$-klikkek ahol N tetszőlegesen nagy, de rögzített szám. Egy többtípusú elágazó folyamat segítségével írtuk le a hálózat fejlődését. A többtípusú elágazó folyamatok értelmezésében az n-klikkeket n típusú egyedeknek tekintjük. Kezdetben, a $t = 0$ időpillanatban csupán egyetlen klikk létezik, melynek mérete az $1, 2, \ldots, N$ bármelyike lehet. Ezt a klikket ősnek nevezzük. Az ős utódokat hoz létre, melyek szintén lehetnek $1, 2, \ldots, N$-klikkek. Ezen utódok is létrehozzák a saját utódaikat, stb. Minden klikkhez tartozik egy 1 paraméterű Poisson folyamat. Az n-klikkek ezen folyamatai függetlenek és azonos eloszlásúak. Ha $\Pi_n(t)$ értéke eggyel növekszik, akkor megjelenik egy új csúcs és ez az n-klikk j csúcsához $q_{n,j}$ valószínűséggel csatlakozik, ahol $0 \leq q_{n,j} \leq 1$, $j = 0, 1, \ldots, n$, és $\sum_{j=0}^{n} q_{n,j} = 1$. (Feltesszük, hogy $q_{N,N} = 0$, mert N a legnagyobb klikkméret.) Annak valószínűsége, hogy egy i típusú ős j típusú utódot képez, $p_{i,j} = q_{i,j-1}$, $j = 1, 2, \ldots, i + 1$.

Egy i-klikk λ_i élettartamát (azaz szaporodóképes fázisának hosszát) az $l_i(t) = b + c\xi_i(t)$ kockázati ráta határozza meg, ahol $b \geq 0$, $c > 0$ rögzített konstansok és $\xi_i(t)$ az i-klikk összes utódainak száma a t időpillanatig. $m_{i,j}(t) = \mathbb{E}\xi_{i,j}(t)$ az i típusú egyed j típusú utódainak átlagos száma a t ideig. Jelölje M az $m_{ij}(t)$ függvények $m^*_{ij}(t)$ Laplace transzformáltjainak mátrixát.

Folytonos idejű hálózatfejlődési modellünk matematikai leírása után legfontosabb eredményeink a következők voltak:

Tegyük fel, hogy a $(p_{i,j})_{i,j=1}^N$ mátrix irreducibilis és aciklikus. Bevezetünk egy $A(\kappa)$ mátrixot, mely a-tól és b-től függ. Tegyük fel, hogy $A(0) > 1$, azaz a folyamat szuperkritikus. Legyen α az M által meghatározott Malthusi paraméter. Ekkor jelöljük $\mathbf{v} = (v_1, \ldots, v_N)^\top$-nel az $M(\alpha)$ mátrix 1 sajátértékhez tartozó sajátvektorát, amely $v_1 + \cdots + v_N = 1$ módon van normálva. Legyen $\mathbf{u} = (u_1, \ldots, u_N)^\top$ az $M(\alpha)$ azon bal sajátvektora, melyre $u_1 v_1 + \cdots + u_N v_N = 1$. $A(\alpha)$ a Laplace-transzformáltak által meghatározott és

$$-A'(\alpha) = D(\alpha) = \sum_{l,j=1}^{N} u_l v_j \left(-m_{l,j}^*(\alpha)\right)'.$$

A következő eredményekben a ${}_kW$ mennyiségek m.b. nemnegatívak, $\mathbb{E}({}_kW) = 1$, illetve ${}_kW$ m.b. pozitív a túlélés eseménye mellett. Legyen n rögzített, $1 \leq n \leq N$.

1. Jelölje ${}_kT(t)$ a t ideig megszületett n-klikkek számát, ha az ős k-klikk, $k = 1, 2, \ldots, N$. Ekkor

$$\lim_{t\to\infty} e^{-\alpha t}{}_kT(t) = {}_kW \frac{v_k u_n}{\alpha\left(-A'(\alpha)\right)}$$

majdnem biztosan $k = 1, 2, \ldots, N$ esetén.

2. Jelölje ${}_k\hat{T}(t)$ a t időpillanatban életben lévő n-klikkek számát, ha az ős k-klikk, $k = 1, 2, \ldots, N$. Ekkor

$$\lim_{t\to\infty} e^{-\alpha t}{}_k\hat{T}(t) = {}_kW \frac{v_k u_n A(\alpha)}{\left(-A'(\alpha)\right)}$$

majdnem biztosan $k = 1, 2, \ldots, N$ esetén.

Legyen $\mathbb{M}$ az összes utódok átlagos számaiból álló mátrix. Ekkor $\mathbb{M} = (m_{i,j}(\infty))_{i,j=1}^N = A(0)\,(p_{i,j}(\infty))_{i,j=1}^N$.

3. Jelölje s_i a kihalás valószínűségét, ha az ős i típusú egyed volt. Legyen $\mathbf{s} = (s_1, \ldots, s_N)$. Tegyük fel, hogy $(p_{i,j}(\infty))_{i,j=1}^N$ irreducubilis és aciklikus mátrix. Legyen ϱ az $\mathbb{M}$ Perron-Frobenius gyöke. Ha $\varrho \leq 1$, akkor $s_1 = s_2 = \cdots = s_N = 1$. Ha $\varrho > 1$, akkor $s_1 < 1, s_2 < 1, \ldots, s_N < 1$. Az $\mathbf{s}$ mindegyik esetben az alábbi vektoregyenlet legkisebb nemnegatív megoldása:

$$\mathbf{s} = \mathbf{f}(\mathbf{s}),$$

ahol $\mathbf{f} = (f_1, \ldots, f_N)$ és az f_k függvények az utódeloszlás generátorfüggvényei (1.6.19)).

Első két eredményünk leírja a t időpontig megszületett összes k-klikkek számának, illetve a t időpontban életben lévő k-klikkek számának aszimptotikus viselkedését.

Hasonló eredményeink vannak [24] cikkünkben egy rögzített csúcs fokszámára is. Tételeink empirikus alátámasztásához szimulációkat is végeztünk, melyek eredményei a 1.7 szakaszban szerepelnek.

A második fejezet [29] cikkünk alapján íródott. Egy diszkrét idejű hálózatfejlődési modellt vizsgáltunk, amelynek egyedei a k-klikkek és a hálózat fejlődése k-klikkek konstrukcióján és törlésén alapul, ahol $k \geq 2$ rögzített egész.

Az $n = 0$ időpontban a kezdeti hálózat k csúcsú üres gráf, mely nem tartalmaz éleket. Az első lépésben, azaz az $n = 1$ időpillanatban összekötjük a k db csúcsot, és így kapunk egyetlen k-klikket. A második lépés a következő. Kiválasztunk két csúcsot véletlenszerűen, jelölje őket v_1 és v_2. Ezután hozzáadunk a gráfhoz egy új csúcsot és két k-klikket készítünk. Az első k-klikk csúcsai a meglévő $k + 1$ csúcs, kivéve v_1-et, míg a második k-klikk csúcsait a már létező $k + 1$ csúcs adja v_2-t kivéve. Ezután az eredeti k-klikket töröljük.

Ezután minden egyes lépésben kiválasztunk a már meglévő csúcsok közül k db-ot véletlenszerűen (diszkrét egyenletes eloszlás szerint). Amennyiben ezek nem alkotnak k-klikket, akkor konstruálunk egy új k-klikket ezen csúcsokkal (azaz összekötjük őket $\binom{k}{2}$ új éllel). A másik esetben, ha ez a k csúcsból álló részgráf egy k-klikk, akkor ez a k-klikk törlésre kerül, viszont a csúcsait felhasználjuk két új k-klikk készítéséhez úgy, mint a második lépésben. Azaz egy új csúcsot adunk hozzá a gráfhoz és ezt, illetve az épp kitörölt klikk k csúcsát felhasználva két új k-klikket készítünk olyan módon, mint a második lépésben.

Modellünk leírását követően bizonyítottunk egy határérték-tételt a csúcsok számára, illetve a gráf csúcsai számának aszimptotikus normalitását is beláttuk. Ezután megadtunk egy aszimptotikus tételt modellünk csúcsainak számára is.

4. Ha $n \to \infty$, a gráf csúcsainak számára teljesül az alábbi majdnem biztos konvergencia-tulajdonság:

$$\frac{V_n}{\left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}} \to 1.$$

Ezt az alábbi formában is írjuk:

$$V_n \sim \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}}, \qquad \text{majdnem biztosan} \quad n \to \infty.$$

5. Az alábbi aszimptotikus normalitási tulajdonság teljesül:

$$\frac{1}{n^{\frac{1}{k+1}}} \left(V_n - \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}} n^{\frac{2}{k+1}} \right) \Rightarrow \mathcal{N}\left(0, \frac{1}{2k+1} \left[\frac{(k+1)!}{2}\right]^{\frac{1}{k+1}}\right)$$

ha $n \to \infty$, ahol $\Rightarrow$ az eloszlásbeli konvergenciát jelöli.

6. Ha a $d_n(v)$ jelölést használjuk a rögzített v csúcs fokszámára az n időpontban, akkor

$$d_n(v) \sim k(k+1)\left[\frac{2}{(k+1)!}\right]^{\frac{1}{k+1}} n^{\frac{k-1}{k+1}}$$

Szimulációkat is végeztünk, hogy kísérleti úton is alátámasszuk előbbi három, csúcsok számáról és fokszámokról szóló eredményünket.

Bebizonyítottunk egy funkcionális határeloszlást-tételt (invariancia-elvet) a gráf csúcsainak számára és többdimenziős funkcionális határaloszlás-tételeket csúcsok fokszámaira. jelölje $[nt]$ az nt mennyiség egész részét. A csúcsok számának tanulmányozásához tekintsük a $V_{[nt]}$, $t \in [0,1]$ folyamatot, ahol V_n a csúcsok száma. A $V_{[nt]}$ folyamat trajektóriái benne vannak a $D[0,1]$, azaz nincs másodfajú szakadásuk.

7.
$$\frac{\binom{V_{[nt]}}{k+1} - \binom{[nt]}{2}}{n^{\frac{2k+1}{k+1}}} \Rightarrow \int f(s)\, dW(s),$$

ahol
$$f(s) = \frac{\left(\frac{(k+1)!}{2}\right)^{\frac{k}{2k+1}}}{(k!)^{\frac{1}{2}}} \cdot s^{\frac{3k+1}{2(k+1)}},$$

$W(s)$, $s \in [0,1]$, a Wiener-folyamat, és $\Rightarrow$ jelöli a gyenge konvergenciát a $D[0,1]$ a Szkorohod-topológiában.

Többdimenziós funkcionális határeloszlás-tételeket is kaptunk a csúcsok fokszámaira. A gráf m db csúcsának együttes viselkedését tanulmányoztuk, ahol m egy rögzített egész.

8. Tekintsük a v_l, $l = 1,\ldots,m$ csúcs esetén a következő martingáldifferenciákat:
$$X_{n,i}^{(v_l)} = \frac{1}{n^{\frac{k-1}{2(k+1)}}}\left[\mathbb{I}_{A_i^{(v_l)}} - \mathbb{E}\left(\mathbb{I}_{A_i^{(v_l)}} | \mathcal{F}_{i-1}\right)\right],$$

ahol $A_i^{(v_l)}$ az az esemény, hogy kiválasztjuk a v_l a csúcsot az i-edik lépésben, de az nem a két kivételes csúcs egyike. Definiáljuk $t \in [0,1]$-re az
$$Y_n^{(v_l)} = \sum_{i=1}^{[nt]} X_{n,i}^{(v_l)}.$$

mennyiséget. Legyenek v_1, v_2, $\ldots$, v_m különböző csúcsok. Ekkor
$$\left(Y_n^{(v_1)},\ldots,Y_n^{(v_m)}\right) \Rightarrow \left(\int f\, dW_1,\ldots,\int f\, dW_m\right),$$

ahol $W_1,\ldots,W_m$ független Wiener-folyamatok, $f(x) = k^{\frac{1}{2}}\left[\frac{2}{(k+1)!}\right]^{\frac{1}{2(k+1)}} x^{-\frac{1}{k+1}}$ pedig Helland [43] cikkének (3.4) feltételében szereplő függvény, és $\Rightarrow$ jelöli a gyenge konvergenciát a szorzat Szkorohod-topológiában a $D[0,1] \times \cdots \times D[0,1]$ téren.

9. Legyen $k = 2$ és v_1, v_2, $\dots$, v_m különböző csúcsok. Ha

$$Z_n^{(v_l)} = \frac{1}{n^{\frac{1}{6}}} \left(d_{[nt]}(v_l) - 2 \cdot 3^{\frac{2}{3}} [nt]^{\frac{1}{3}} \right), \quad t \in [0,1],$$

$l = 1, \dots, m$-re, akkor

$$\left(Z_n^{(v_1)}, \dots, Z_n^{(v_m)} \right) \Rightarrow \left(\int f \, dW_1, \dots, \int f \, dW_m \right).$$

Itt $W_1, \dots, W_m$ független Wiener-folyamatok, $f(x) = 2^{\frac{1}{2}} \cdot 3^{-\frac{1}{6}} \cdot x^{-\frac{1}{3}}$, és $\Rightarrow$ a gyenge konvergenciát jelöli a szorzat Szkorohod topológiában a $D[0,1] \times \cdots \times D[0,1]$ térben.

Továbbá

$$\left(V_n^{(v_1)}, \dots, V_n^{(v_m)} \right) \Rightarrow \mathcal{N}_m \left(\mathbf{0}, 2 \cdot 3^{\frac{2}{3}} \mathbf{I}_m \right),$$

ahol

$$V_n^{(v_l)} = \frac{1}{n^{\frac{1}{6}}} \left(d_n(v_l) - 2 \cdot 3^{\frac{2}{3}} n^{\frac{1}{3}} \right), \quad l = 1, \dots, m,$$

$\mathcal{N}_m \left(\mathbf{0}, 2 \cdot 3^{\frac{2}{3}} \mathbf{I}_m \right)$ m-dimenziós, többváltozós normális eloszlás, és $\mathbf{I}_m$ az $m \times m$ típusú egységmátrix. Tehát a csúcsok fokszámainak együttes eloszlása aszimptotikusan normális és különböző csúcsok fokszámai aszimptotikusan függetlenek.

A (4.)–(6.) eredményeink Backhausz és Móri [5] eredményeinek kiterjesztései. Viszont (7.)–(9.) minden k-ra újak, beleértve a [5] cikk által tanulmányozott $k = 2$ speciális esetet is. A (7.) eredmény egy funkcionális határeloszlás-tétel a gráfunk csúcsainak számára. (8.) és (9.) pedig több csúcs együttes eloszlására vonatkozó funkcionális határeloszlás-tételek.

A 3. fejezet [26] cikkünk eredményeit tartalmazza. Az fej-írás érmedobási kísérletet vizsgáljuk. Legyen p a fej dobás valószínűsége, $q = 1 - p$ pedig az írásé. Itt p rőgzített, $0 < p < 1$. Tekintsük az $X_1, X_2, \dots, X_N$ független, azonos eloszlású valószínűségi változókat, $P(X_i = 1) = p$, $P(X_i = 0) = q = 1 - p$, $i = 1, 2, \dots, N$ eloszlással.

Új közelítést adtunk a leghosszabb, legfeljebb T-szennyezett fej sorozat hosszának eloszlására. Megmutattuk, hogy $T > 0$-ra a becslésünk hibája $\mathcal{O}\left(1/(\log(n))^2\right)$, ahol log az $1/p$ alapú logaritmust jelöli. Legyen továbbá $R_n(T)$ a leghosszabb, legfeljebb T-szennyezett, első n dobásnál kezdődő sorozat hossza.

A fejezet fő eredménye az alábbi:

10. Legyen $T \geq 1$ egész és legyen

$$\begin{aligned}
\tilde{c}_n(T) &= \log(qn) + T\log(\log(qn)) + \\
&+ T^2\frac{\log(\log(qn))}{c\log(qn)} - \frac{T}{cq_0\log(qn)} - \frac{T^3}{2c}\left(\frac{\log(\log(qn))}{\log(qn)}\right)^2 + \\
&+ T^2\frac{\log(\log(qn))}{cq_0(\log(qn))^2} + T^3\frac{\log(\log(qn))}{(c\log(qn))^2} + \\
&+ \left(T\log\left(\frac{q}{p}\right) - \log(T!)\right)\left(1 + \frac{T}{c\log(qn)} - T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right),
\end{aligned}$$

ahol log jelöli az $1/p$ alapú logaritmust, $c = \ln(1/p)$, illetve ln a természetes, e alapú logaritmus, és

$$q_0 = \frac{2q}{2 + Tq - q}.$$

Jelölje $[\tilde{c}_n(T)]$ a $\tilde{c}_n(T)$ egész részét, $\{\tilde{c}_n(T)\}$ pedig $\tilde{c}_n(T)$ tört részét, azaz $\{\tilde{c}_n(T)\} = \tilde{c}_n(T) - [\tilde{c}_n(T)]$. Ekkor

$$\begin{aligned}
&P(R_n(T) - [\tilde{c}_n(T)] < l) = \\
&= \exp\left(-p^{(l-\{\tilde{c}_n(T)\})\left(1 - \frac{T}{c\log(qn)} + T^2\frac{\log(\log(qn))}{c(\log(qn))^2}\right)}\right)\left(1 + \mathcal{O}\left(\frac{1}{(\log n)^2}\right)\right)
\end{aligned}$$

minden l egész számra.

Láttuk, hogy $T > 0$-ra a közelítés hibája Arratia, Gordon és Waterman [2] munkájában $\mathcal{O}(\log(\log(n))/\log(n))$, volt, tehát tételünk jelentősen javította ezt a korábbi eredményt. Ezen tételünket szintén szimulációkkal támasztottuk alá.

Bibliography

[1] Aalen, O. O., Weak convergence of stochastic integrals related to counting processes. *Z. für Wahrscheinlichkeitstheorie und Verw. Gebiete*, **38**, 261–277, 1977.
https://doi.org/10.1007/BF00533158

[2] Arratia, R.; Gordon, L.; Waterman, M. S., The Erdős-Rényi Law in Distribution, for Coin Tossing and Sequence Matching. *Ann. Statist.* **18**, 2, 539–570, 2008.
https://doi.org/10.1214/aos/1176347615

[3] Athreya, K. B., Ghosh, A. P., Sethuraman, S., Growth of preferential attachment random graphs via continuous-time branching processes, *P. Indian AS-Math. Sci.* **118**, 473–494, 2008.
https://doi.org/10.1007/s12044-008-0036-2

[4] Backhausz, Á., Móri, T. F., A random graph model based on 3-interactions, *Ann. Univ. Sci. Budapest. Sect. Comput.* **36**, 41–52, 2012.
https://doi.org/10.71352/ac.36.041

[5] Backhausz, Á., Móri, T. F., A random graph of moderate density. *Electron. Commun. in Probab.* **27**, 1–12, 2022.
https://doi.org/10.1214/21-ECP444

[6] S. Banerjee, X. Huang, Degree centrality and root finding in growing random networks, *Electron. J. Probab.*, **28**, 1–39, 2023.
https://doi.org/10.1214/23-EJP930

[7] A.-L. Barabási, *Network science*, Cambridge University Press: Cambridge, UK, 2018.

[8] Barabási, A.-L.; Albert, R., Emergence of scaling in random networks. *Science*, **286**, 509–512, 1999.
https://doi.org/10.1126/science.286.5439.509

[9] Billingsley, P., *Convergence of Probability Measures*, Wiley, 1968.

[10] Bollobás, B., *Random graphs*, Cambridge University Press, 2001.

[11] Bollobás, B., Riordan, O., *Random graphs and branching processes,* In: B. Bollobás, R. Kozma, D. Miklós (Eds.), Handbook of Large-Scale Random Networks, Bolyai Society Mathematical Studies, Springer: Berlin, Heidelberg, **18**, 2008.
https://doi.org/10.1007/978-3-540-69395-6_1

[12] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001), The degree sequence of a scale-free random graph process. *Random Struct. Alg.* **18**, 279-290, 2001. https://doi.org/10.1002/rsa.1009

[13] Broido, A. D., Clauset, A., Scale-free networks are rare. *Nat Commun* **10**, 1017, 2019.
https://doi.org/10.1038/s41467-019-08746-5

[14] Cooper, C., Frieze, A., Vera., J., Random Deletion in a Scale-Free Random Graph Process. *Internet Math.* **1** (4) 463 - 483, 2003/2004.

[15] Cooper, C., Frieze, A., A general model of web graphs. *Random Struct. Alg.*, **22**, 311-335, 2003.
https://doi.org/10.1002/rsa.10084

[16] Csáki, E.; Földes, A.; Komlós, J., Limit theorems for Erdős-Rényi type problems. *Studia Sci. Math. Hungar.* **22**, 321–332, 1987.

[17] M. Deijfen, R. Fitzner, Birds of a feather or opposites attract - effects in network modelling, *Internet Mathematics* **1**, 1, 2017.
https://doi.org/10.24166/im.11.2017

[18] Durrett, R., *Random graph dynamics.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge Univ. Press, 2007.

[19] Erdős, P.; Rényi, A. On a new law of large numbers, *J. Analyse Math.* **23**, 103-111, 1970.

[20] Erdős, P., Rényi, A., On the Evolution of Random Graphs. *Publ. of the Math. Inst. of the Hung. Acad. of Sci.* **5**, 17-61, 1960.

[21] Erdős, P.; Révész, P. On the length of the longest head-run. In *Topics in information theory (Second Colloq., Keszthely, 1975)*, Colloq. Math. Soc. János Bolyai, Vol. 16, North-Holland, Amsterdam, 219–228, 1977.

[22] Fazekas, I., Merging to semistable processes. *Theory of Probability and its Applications* **56**, 4, Pages 621–633, 2011.
https://doi.org/10.1137/S0040585X97985662

[23] Fazekas, I., Barta, A., A continuous-time network evolution model describing 2- and 3-interactions, *Mathematics,* **9**, 3143, 2021.
https://doi.org/10.3390/math9233143

[24] Fazekas, I., Barta, A., Fórián, L., Porvázsnyik, B., A continuous-time network evolution model describing N-interactions, *AIMS Mathematics* **9**, 12, 35721-35742, 2024.
https://doi.org/10.3934/math.20241695

[25] Fazekas, I., Barta, A., Noszály, Cs., Porvázsnyik, B., A continuous-time network evolution model describing 3-interactions. *Comm. in Stat. - Theory and Methods*, **52**, 11, 4001–4020, 2023. https://doi.org/10.1080/03610926.2021.1985141

[26] Fazekas, I., Fazekas, B., Fórián, L., On the convergence rate for the longest at most T-contaminated runs of heads, *Entropy* **27**, 1, 33, 2025. https://doi.org/10.3390/e27010033

[27] Fazekas, I., Fazekas, B., Suja, M. O., Convergence rate for the longest T-contaminated runs of heads. *Statist. Probab. Lett.* **208**, 110059, 2024. https://doi.org/10.1016/j.spl.2024.110059

[28] Fazekas, I., Fazekas, B., Suja, M. O., A limit theorem for runs containing two types of contaminations, Period. Math. Hungar. **90**, 1, 78-93, 2025. https://doi.org/10.1007/s10998-024-00600-6

[29] Fazekas, I., Fórián, L., A family of network evolution models with moderate density. *Stat. Papers* **66**, 105, 2025. https://doi.org/10.1007/s00362-025-01712-y

[30] Fazekas, I., Karácsony, Zs., Libor, Zs., Longest runs in coin tossing. Comparison of recursive formulae, asymptotic theorems, computer simulations. Acta Universitatis Sapientiae, Mathematica 2/2 215-228, 2010. https://doi.org/10.5485/TMCS.2011.0290

[31] Fazekas, I., Noszály, Cs., Perecsényi, A., A population evolution model and its applications to random networks, *Stat. Probabil. Lett.*, **143**, 17-27, 2018. https://doi.org/10.1016/j.spl.2018.07.012.

[32] Fazekas, I., Noszály, Cs., Perecsényi, A., The N-star network evolution model. *J. Appl. Probab.* **56**, 2, 416–440, 2019. https://doi.org/10.1017/jpr.2019.21

[33] Fazekas, I., Porvázsnyik, B., Scale-free property for degrees and weights in an N-interactions random graph model, *J. Math. Sci.* **214**, 69–82, 2016. https://doi.org/10.1007/s10958-016-2758-5

[34] Feng, Q., Li, X., Hu, Z., Asymptotic degree distribution in a homogeneous evolving network model, *Stat. Probabil. Lett.* **193**, 109740, 2023. https://doi.org/10.1016/j.spl.2022.109740

[35] Földes, A. The limit distribution of the length of the longest head-run. *Period. Math. Hungar.* **10**, 4, 301–310, 1979.

[36] Gao, F., van der Vaart, A., Castro, R., van der Hofstad, R., Consistent estimation in general sublinear preferential attachment trees, *Electron. J. Stat.* **11**, 3979–3999, 2017. https://doi.org/10.1214/17-EJS1356

[37] Gnedenko, B., V., Kolmogorov, A. N., *Limit distributions for sums of independent random variables.* Cambridge, Addison-Wesley, 1954.

[38] Gordon, L.; Schilling, M. F.; Waterman, M. S., An extreme value theory for long head runs. *Probab. Theory Relat. Fields* **72**, 2, 1986, 279–287, 1986.

[39] Haccou, P., Jagers, P., Vatunin, V. A., *Branching processes: Variation, growth, and extinction of populations*, Cambridge University Press: Cambridge, UK, 2005.

[40] Hall, P., Heyde, C. C., Martingale Limit Theory and its Application. *Academic Press: New York*, 1980.
https://doi.org/10.1016/C2013-0-10818-5

[41] Hao, C. X., Hu, Z. C., Ma, T., On the length of the longest consecutive switches. *J. Sichuan Univ.: Nat. Sci. Ed.* **58**, 021002, 2021.
10.19907/j.0490-6756.2021.021002

[42] Hao, C.X., Ma, T. Limiting Behaviors for Longest Consecutive Switches in an IID Bernoulli Sequence, *Acta. Math. Sin.-English Ser.* **39**, 1138–1146, 2023.
https://doi.org/10.1007/s10114-023-1355-z

[43] Helland, I. S., Central Limit Theorems for Martingales with Discrete or Continuous Time. *Scandinavian Journal of Statistics* **9**, No. 2, 79–94, 1982.

[44] van der Hofstad, R., *Random Graphs and Complex Networks. Volume One.* Volume 43 in Cambridge Series in Statistical and Probabilistic Mathematics, 2017., ISBN 978-1-107-17287-6

[45] van der Hofstad, R., *Random Graphs and Complex Networks. Volume Two.* Volume 54 in Cambridge Series in Statistical and Probabilistic Mathematics, 2024., ISBN 978-1-107-17400-9

[46] Holmgren, C., Janson, S., Fringe trees, Crump-Mode-Jagers branching processes and m-ary search trees, *Probab. Surv.* **14**, 53–154, 2017.
https://doi.org/10.1214/16-PS272

[47] Hoppen, C., Kohayakawa, Y., Moreira, C. G., Ráth, B., Rudini, M. S., Limits of Permutation Sequences, *Journal of Combinatorial Theory, Series B*, 103 / 1, 93-113, 2013.
https://doi.org/10.1016/j.jctb.2012.09.003

[48] Iksanov, A., Kolesko, K., Meiners, M., Asymptotic fluctuations in supercritical Crump-Mode-Jagers processes, *Ann. Probab.* **52**, 1538–1606, 2024.
https://doi.org/10.1214/24-AOP1697

[49] Iksanov, A., Meiners, M., Rate of convergence in the law of large numbers for supercritical general multi-type branching processes, *Stoch. Proc. Appl.* **125**, 708–738, 2015.
https://doi.org/10.1016/j.spa.2014.10.004

[50] Jagers, P., *Branching processes with biological applications*, Wiley: London, 1975.

[51] Janson, S., Luczak, T., Rucinski, A., *Random Graphs.* Wiley, 2000., ISBN:9780471175414
https://doi.org/10.1002/9781118032718

[52] Kopocinski, B., On the distribution of the longest succes-run in Bernoulli trials, *Mat. Stos.* **20**, 34, 3–13, 1991.
https://doi.org/10.14708/MA.V20I34.1781

[53] Lovász, L., *Large Networks and Graph Limits.* American Mathematical Society, 2012., ISBN: 978-0-8218-9085-1

[54] Lovász, L., Szegedy, B., Limits of dense graph sequences, *Journal of Combinatorial Theory, Series B*, **96**, 6, 2006, 933-957, 2006.
https://doi.org/10.1016/j.jctb.2006.05.002

[55] Mode, C. J., *Multitype branching processes: Theory and applications*, American Elsevier: New York, USA, 1971.
https://doi.org/10.1137/1014092

[56] Móri, T. F., The a.s. limit distribution of the longest head run. *Canad. J. Math.* , **45**, 6, 1245–1262, 1993.
https://doi.org/10.4153/CJM-1993-070-x

[57] Móri T. F., On random trees. *Stud. Sci. Math. Hun.* **39**, 1, 2002.
https://doi.org/10.1556/SScMath.39.2002.1-2.9

[58] Móri, T. F., Rokob, S., A random graph model driven by time-dependent branching dynamics, *Ann. Univ. Sci. Budapest. Sect. Comp.*, **46**, 191–213, 2017.
https://doi.org/10.71352/ac.46.191

[59] Móri, T. F., Rokob, S., Moments of general time dependent branching processes with applications, *Acta Math. Hung.* **159**, 131–149, 2019.
https://doi.org/10.1007/s10474-019-00976-9

[60] Nerman, O., *On the convergence of supercritical general branching processes*, PhD Theses, University of Göteborg, 1979.

[61] Nerman, O., On the convergence of supercritical general (C-M-J) branching processes, *Z. Wahrsch. Verw. Gebiete,* **57**, 365–395, 1981.
https://doi.org/10.1007/BF00534830

[62] Neveu, J., Discrete-Parameter Martingales. *North-Holland: Amsterdam*, 1975.

[63] Newman, M., Barabási, A.-L., Watts, D. J., *The Structure and Dynamics of Networks.* Princeton University Press, 2006., ISBN 9780691113579

[64] Novak, S. Y.m On the length of the longest head run. *Statist. Probab. Lett.* **130**, 111–114, 2017.

[65] Ostroumova, L., Ryabchenko, A., Samosvat, E., *Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient.* In: International Workshop on Algorithms and Models for the Web-Graph, Springer, 2013.

[66] Révész, P.: Mennyire véletlen a véletlen?: Akadémiai székfoglaló, 1982. november 1. Akadémiai Kiadó, Budapest, 1984.

[67] S. Rosengren, A multi-type preferential attachment tree, *Internet Math.*, **1**, 2018.
https://doi.org/10.24166/im.05.2018

[68] Rudas, A., Tóth, B., Valkó, B., *Random trees and general branching processes*, *Random Struct. Algor.*, **31**, 186–202, 2007.
https://doi.org/10.1002/rsa.20137

[69] Rudas, A., Tóth, B., *Random tree growth with branching processes — A survey*, In: B. Bollobás, R. Kozma, D. Miklós (Eds.), Handbook of Large-Scale Random Networks, Bolyai Society Mathematical Studies, Springer: Berlin, Heidelberg, **18**, 2008.
https://doi.org/10.1007/978-3-540-69395-6_4

[70] Seneta, E., *Non-negative matrices and Markov chains*, Springer: New York, USA, 1981.
https://doi.org/10.1007/0-387-32792-4

[71] *Simulation codes for the article 'A continuous-time network evolution model describing N-interactions'*, 2024. Available from:
`https://github.com/bartaa89/n_interact`.

List of publications and conference talks

List of publications related to the dissertation

1. Fazekas, I., Fórián, L., A family of network evolution models with moderate density. *Statistical Papers* **66**, 105, 2025. (SJR: Q2) https://doi.org/10.1007/s00362-025-01712-y

2. Fazekas, I., Fazekas, B., Fórián, L., On the convergence rate for the longest at most T-contaminated runs of heads. *Entropy* **27**, 1, 33, 2025. (SJR: Q2) https://doi.org/10.3390/e27010033

3. Fazekas, I., Barta, A., Fórián, L., Porvázsnyik, B., A continuous-time network evolution model describing N-interactions. *AIMS Mathematics* **9**, 12, 35721-35742, 2024. (SJR: Q2) https://doi.org/10.3934/math.20241695

List of other publications

4. Fazekas, I., Fórián, L., Barta, A., Deep learning from noisy labels with some adjustments of a recent method. *Infocomm. J.* **15**, 9-12, 2023. (SJR: Q3) https://doi.org/10.36244/ICJ.2023.5.2

5. Fazekas, I., Barta, A., Fórián, L., Ensemble noisy label detection on MNIST. *Ann. Math. Inform.* **53**, 125-137, 2021. (SJR: Q3) https://doi.org/10.33039/ami.2021.03.015

6. Vincze, Cs., Varga, A., Oláh, M., Fórián, L., On computable classes of equidistant sets: equidistant functions. *Miskolc Math Notes* **19**, 1, 677-698, 2018. (SJR: Q3) https://doi.org/10.18514/MMN.2018.2145

7. Vincze, Cs., Varga, A., Oláh, M., Fórián, L., Lőrinc, S., On computable classes of equidistant sets: finite focal sets. *Involve* **11**, 2, 271-282, 2018. https://doi.org/10.2140/involve.2018.11.271

List of conference talks

1. Fazekas, I., Barta, A., Fórián, L., Ensemble noisy label detection on MNIST, *The 1st Conference on Information Technology and Data Science*, University of Debrecen, Debrecen, 6–8 November, 2020
2. Fazekas, I., Barta, A., Fórián, L., Label noise handling on MNIST with neural networks, *8th International Conference on Mathematics and Informatics*, Sapientia Hungarian University of Transylvania, Târgu Mureş (Marosvásárhely), 9-10 September, 2021
3. Fazekas, I., Fórián, L., Barta, A., Deep learning from noisy labels with some adjustments of a recent method, *The 12th International Conference on Applied Informatics*, Eszterházy Károly Catholic University, Eger, 2-4 March, 2023
4. Fazekas, I., Fórián, L., A family of random graph evolution models with moderate density, *The 2024 IEEE 3rd Conference on Information Technology and Data Science*, University of Debrecen, Debrecen, 26-27 August, 2024
5. Fazekas, I., Fórián, L., A family of network evolution models with moderate density, *10th International Mathematics and Informatics Conference on Distance and Critical Points*, Sapientia Hungarian University of Transylvania, Târgu Mureş (Marosvásárhely), 8-12 September, 2025