AARHUS UNIVERSITY

FACULTY OF NATURAL SCIENCES

RECOMMENDATION
PHD DEGREE

Péter László Juhász

Born on 22 March 1992

PhD thesis

Title of PhD thesis: Stochastic Models of Higher-Order Networks
Graduate programme: Mathematics
Submitted on 03 October 2025

Supervisors

Main supervisor: Associate Professor Christian Hirsch,
Department of Mathematics, Aarhus University
Co-supervisor: Associate Professor Claudia Strauch,
Department of Mathematics, Aarhus University

Assessment Committee

Member of assessment committee: Professor Remco van der Hofstad, Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands
Member of assessment committee: Prof. Dr. Daniel Hug, Department of Mathematics, Karlsruhe Institute of Technology, Germany
Chair: Associate Professor Markus Kiderlen,
Department of Mathematics, Aarhus University

Introduction

The thesis *Stochastic Models of Higher-Order Networks* by Péter Juhász is written in English with a Danish abstract included. The thesis consists of 282 pages and includes the following four research papers:

Ch. Hirsch, B. Jahnel, S.K. Jhawar, **P. Juhász**:
Poisson approximation of fixed-degree nodes in weighted random connection models. *Stochastic Process. Appl.* 183 (2025)

Ch. Hirsch, **P. Juhász**:
On the topology of higher-order age-dependent random connection models. *Methodol. Comput. Appl. Probab.* 27, article number 44, (2025)

SOLIDUM PETIT IN PROFUNDIS · UNIVERSITAS ARHUSIENSIS ·

M. Brun, Ch. Hirsch, **P. Juhász,** M. Otto:
Random connection hypergraphs. *arXiv:2407.16334v2* (2025)

Ch. Hirsch, B. Jahnel, **P. Juhász**:
Functional limit theorems for edge counts in dynamic random connection hypergraphs*. arXiv:2507.16270* (2025)

According to the coauthor statements, the PhD student made a proportional contribution to all 4 papers.

Assessment

The thesis of Péter Juhász is devoted to the investigation of random graph models building on geometry and higher-order structure. Such random graph models have attracted considerable attention in modern probability theory. The inspiration for them comes from real-world networks, such as collaboration networks, where one can expect that proximity between scientists makes it more likely that they collaborate, and where groups of scientists beyond pairs tend to collaborate on joint papers. The thesis focuses on structural and topological properties of such graphs, including their higher-order degree distributions and so-called Betti numbers, in the large graph limit. Considerable effort is devoted to the simulations of such random graphs, to test whether the theoretical large graph limits describe networks of reasonable size already well, and their fit to real-world collaboration networks.

The thesis contains 5 parts: firstly, an introduction summarizing the entire thesis and positioning it in the broad domain of network science, and then the 4 papers A-D, all co-authored by the PhD candidate. We now discuss the various parts one by one.

The introduction describes the field at large, and highlights the main contributions made in the thesis. It argues that most real-world networks have *power-law degree distributions*, and that the *geometric embedding* plays a central role in such networks. The geometric embedding gives rise to random graph models that have much larger *clustering* than most classical models (the latter not incorporating geometry). Further, it makes a clear case that *higher-order connectivity*, as described by cliques in the network, should be taken into account to realistically model real-world networks. This chapter is very useful for the reader to gain an overview of the main contributions of the thesis, and how it contributes to the field of spatial complex networks. It is didactically appealing that the chapter closes with a simple example of a spatial network, the Boolean model, to showcase the work. It would have been nice, though, to follow the analysis of the Boolean model up with a longer discussion of how this model relates to the models studied in the thesis.

Paper A studies the structure of the *low-degree vertices* in the weighted random connection model when it is close to being connected. In such cases, there is a bounded number of low-degree vertices. It is well known that, for many random graph models, the connectivity transition (where the random graph becomes connected) occurs at the same location as the transition where the last isolated vertex disappears. Thus, studying the low-degree vertices is certainly worthwhile. The paper studies the spatial locations and intensity of these low-degree vertices, which are described by a suitable Poisson point process with an intensity that depends on the degree. Further, the precise structure and scaling of these limits turns out to depend sensitively on the scaling of the weights close to zero. The weight structure in general is very important for various random graph properties, but it is rare that their *small values* are relevant, so this is a highly interesting finding. At the same time, the resulting networks have average degrees that tend to infinity, unlike many real-world networks, which are *sparse*, in the sense that the average degree is quite small even in large networks.

The proofs rely on Poisson process approximations, building on a theorem by *Bobrowski, Schulte, Yogeswaran*, which estimates the total variation distance between the law of a spatial process and a Poisson process limit in terms of the total variation between their intensities and four error terms, which are then estimated one by one. For readers not aware of the *Bobrowski, Schulte, Yogeswaran* result, it would have been helpful to give a formal definition of the four error terms in the present paper.

Paper B continues with the study of the age-dependent random connection model, where now the weights are interpreted as *ages*. Paper B considers the relevant case of sparse random graphs, where the total number of edges scales linearly in the network size. It investigates the *higher-order structure* of such networks, focussing on the higher-order degree distributions, the total number of edges, as well as topological features captured by the Betti numbers. The main results pertain to the power-law structure of the higher-order degree distributions, central and stable limit theorems for the total number of edges in the random graph, as well as for the *Betti numbers* in such graphs. These Betti numbers describe the number of connected components, the number of loops, the number of voids, etc. They have been defined in Definition 2.13. This definition is phrased in terms of homology groups, which, for the average probabilist reader of the thesis, could have been explained in a little more detail.

The proofs rely heavily on stochastic geometry techniques, using the fact that the model, as well as its limit, are described in terms of spatial Poisson processes.

Paper B closes with two applications. The first concerns simulations of the models, to check how well networks of reasonable size are well-described by the theoretical infinite-size graph limit of the model. The limit turns out to be already visible for reasonable network sizes, which is a highly useful result in practice. The second concerns how well the model fits real-world collaboration networks. This part contains several aspects, in that model parameters need to be estimated (which is notoriously hard, particularly for the power-law exponent of degree distributions), and then the modelling of the models with the estimated parameters. The general results indicate that the age-dependent random connection model does *not* model real-world networks well. This is not surprising though. Indeed, since such networks are generally large, one can expect, as in the famous quote by *Box*, that *all models are wrong*. So, one should not be disheartened by the failure of a simple and stylized model to model the real-world network, as such models could still be useful to understand real-world networks. Understandably, Paper B does not go further into such an analysis, as this is fundamentally highly challenging. The failure of the models to realistically describe real-world collaboration models is based on the limit theorems in Paper B, particularly the central limit theorems of the number of edges and Betti numbers. Thus, this gives a nice application of these important results.

Paper C dives deeper in the topic of *higher-order connectivity*. It proposes a new model, called the weighted random connection hypergraph. In this model, there are two types of vertices, corresponding to the vertices of the graph, as well as the groups through which they are connected using hyperedges. One can thus see the model as originating from a *bipartite random graph*, where the left vertices correspond to the original vertices in the hypergraph, while the right vertices correspond to the *groups* in which the left vertices participate. In the world of weighted random connection models, this is a novel perspective, and it is well motivated by the used case of collaboration networks.

The results of Paper C are related to those of Paper B, and again pertain to the power-law structure of the higher-order degree distributions, central and stable limit theorems for the total number of edges in the random graph, as well as for the Betti numbers in such graphs. Due to the novel definition of the graph, the proofs are technically demanding. As in Paper B, the paper closes with nice simulations and a comparison to real-world networks. Again, and arguably unsurprisingly, the conclusion is that also these models do not pass a statistical test for model fit.

Paper D extends the static random connection model from Paper C and introduces a dynamic component. The new model under investigation is based on two spaces. The first space is composed of quadruples $p = (x, u, b, l)$, encoding a position x, a weight u, a birth time b and a lifetime l. The second space consists of triples $p' = (z, w, r)$ involving a position z, a weight w and a time instance r. An edge between p and p' arises if the spatial components of p and p' are close enough, depending on the two weights u, w and additional tuning parameters γ, γ', and if the time instance r is included in the lifespan $[b, b + l]$. Randomness enters the model via a (partly homogeneous) Poisson process P of points p, where the lifetime follows an exponential distribution, and an independent homogeneous Poisson process P' of points p'. The central object of study in paper D is the random number of edges $S_n(t)$ between points p of P and p' of P' at time t, where the location x of p is in the bounded domain $[0, n]$ and $b \leq r \leq t \leq b + l$.

The two main results of this impressive analysis (not just in terms of its length) are two functional limit theorems in the Skorokhod space $(D[0,1], \mathbb{R})$. In the first regime, where the parameters γ, γ' that govern the connection condition are less than ¼, the standardized random element $S_n(.)$ converges in distribution to a Gaussian process with explicitly determined covariance structure. If $\gamma{>}1/2$ and $\gamma'{<}$¼, and hence in the infinite variance case, a non-standard scaling of $S_n(.)$ is required and then the weak limit in $(D[0,1], \mathbb{R})$ is shown to exist and described as a stable process (that is not Markov). These two profound results, which nicely complement each other, are obtained via an impressive combination of various tools and methods from the literature. Both results require model specific explicit calculations and estimates, some of which are presented in Section 8 and established in Section 12 (and used throughout the argument).

To derive the first functional limit theorem, a multivariate central limit theorem is established via a result by *Schulte and Yukich '19*, which requires to provide bounds for error terms involving integrals over moments of (first and second order) cost operators. The proof of tightness is based on a criterion by *Davydov '96*. The second functional limit theorem is accomplished by means of various decomposition and approximation arguments none of which are standard results, but require deviations and modifications of ideas described, e.g., in *Resnick's '07* monograph.

The work is structured so as to help the reader to enter the material step by step, starting with a gentle introduction, a description of the model and the results along with ideas of proof. For this reason, one first learns about a sequence of lemmas the proofs of which are deferred to a later section. In view of the length of the analysis, this may be considered a natural idea. The innovative work certainly opens the door for future investigations dealing with other functionals or (genuine) spatial modelling beyond the real line.

Summary

The thesis is very compelling, both in terms of the technical mathematical content, as well as the more applied model fit and simulation. Particularly the last paper is impressive, comprising some 90 pages in the thesis, and clearly highlighting the technical abilities of the candidate.
The thesis contains several new cutting-edge results. It is written in a concise style with an introduction giving a good overview of aims and objectives. The thesis documents clearly Péter's strong ability to treat complex spatial network models and to communicate his results concisely and coherently.

Recommendation

The PhD thesis is eligible for conferral of the PhD degree and can thus proceed to defence.

Date: 20 November 2025

On the authority of the assessment committee consisting of
Remco van der Hofstad, Daniel Hug, and Markus Kiderlen

Markus Kiderlen

Chair of the assessment committee,
Markus Kiderlen