



<https://openaccess.leidenuniv.nl>

License: Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: OpenAccess@library.leidenuniv.nl

Article details

Garlaschelli D., Hofstad R. van der, Hollander F. den & Mandjes M. (2018), Complex Networks, *Journal of Statistical Physics* 173(3-4): 439-447. Doi: 10.1007/s10955-018-2166-y



Special Issue of Journal of Statistical Physics Devoted to Complex Networks

Diego Garlaschelli^{1,2} · Remco van der Hofstad³ · Frank den Hollander⁴ · Michel Mandjes⁵

Published online: 1 November 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018



Diego Garlaschelli

Remco van der Hofstad

Frank den Hollander

Michel Mandjes

Complex networks constitutes a multi-disciplinary research field that is rapidly growing and diversifying. It is a crossroad of concepts, ideas and techniques that bring together scientists from different disciplines, all facing the major challenges that arise from dealing with the measurement, simulation, modelling and analysis of (typically very large) real-world net-

✉ Frank den Hollander
denholla@math.leidenuniv.nl

Diego Garlaschelli
diego.garlaschelli@imtlucca.it; garlaschelli@lorentz.leidenuniv.nl

Remco van der Hofstad
rhofstad@win.tue.nl

Michel Mandjes
M.R.H.Mandjes@uva.nl

¹ IMT School for Advanced Studies, Networks Unit, Piazza San Francesco 19, 55100 Lucca, Italy

² Lorentz Institute for Theoretical Physics, University of Leiden, Niels Bohrweg 2, NL-2333 CA Leiden, The Netherlands

³ Department of Mathematics and Computer Science, Eindhoven University of Technology, Groene Loper 5, 5612 AZ Eindhoven, The Netherlands

⁴ Mathematical Institute, University of Leiden, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

⁵ Korteweg-de Vries Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

works, as well as with questions of their optimisation and control. Networks are center stage in probability theory, combinatorics, algorithmics, statistical physics, population genetics, and complexity science. Core topics of interest are structure and function of networks, scaling properties of networks, and applications of networks in physics, biology and engineering.

Complex networks is a vibrant research area, which will continue to be high on the international research agenda for the next few decades, and to bring exciting new links between different disciplines from different viewpoints. It is one of the truly multidisciplinary scientific areas where theory, provided by mathematicians, theoretical physicists, computer scientists and others, reaches out to address applications in other scientific disciplines where ideas from network science are fruitful, such as neuroscience, social science, economics and biology.

The idea for this special issue arose from the outreach activities of NETWORKS, a 10-year research program on complex networks funded by the Dutch Ministry of Education, and administered by the Netherlands Organization for Scientific Research (NWO). It contains 30 papers written by leading researchers in the field. We are grateful to Joel Lebowitz for inviting us to prepare this special issue, and to Kellyann Giudice from the bureau of the Journal of Statistical Physics for her efficient support during the preparatory process. We thank the authors of the papers for their contribution and for their patience during the refereeing process.

What follows is a brief description of each of the papers, organized according to three classes of topics.

(I) Structure of Static and Dynamic Random Graphs

- *On edge exchangeable random graphs* by S. Janson analyses a model for edge exchangeable random graphs obtained by merging multiple edges. It is shown that the model can produce dense, sparse and extremely sparse random graphs. One example yields a power-law degree distribution. Examples are given where the random graph is dense and converges a.s. in the sense of graph limit theory, or where a.s. every graph limit is the limit of some subsequence, or where there is convergence to a non-integrable generalized graphon.
- *Local neighbourhoods for first-passage percolation on the configuration model* by S. Dereich and M. Ortgiese analyzes first-passage percolation on the configuration model. A network is generated, each edge is assigned a weight that corresponds to the passage time of a message along this edge, independently two vertices are chosen uniformly at random (to be thought of as a sender and a recipient), and all edges along the geodesic connecting the two vertices are coloured in red (in the case that both vertices are in the same component). Local limit theorems are proved for the coloured graph around the recipient, in multiple limiting regimes. The limiting objects are described in terms of continuous-time branching processes that describe the local structure of vertices in edge-weighted configuration models. The proofs rely on extensions of local weak convergence techniques, extended so as to be able to deal with the extra coloured path.
- *The local limit of the uniform spanning tree on dense graphs* by J. Hladky, A. Nachmias and T. Tran considers a connected graph in which almost all vertices have linear degrees, and a uniform spanning tree T of this connected graph. For any fixed rooted tree F of given height r the authors compute the asymptotic density of vertices v for which the r -ball around v in T is isomorphic to F . From this property, it is proven that if $\{G_n\}$ is a sequence of such graphs converging to a graphon W , then the uniform spanning tree of G_n locally converges to a multi-type branching process defined in terms of W .
- *Load balancing in hypergraphs* by P. Delgosha and V. Anantharam considers a simple locally finite hypergraph on a countable vertex set, with each edge representing one unit of load that should be distributed among the vertices defining the edge. The authors analyze

the properties of balanced allocations of load, extending the concept of balancedness from finite hypergraphs to their local weak limits.

- *Limiting properties of random graph models with vertex and edge weights* by S. Foss and T. Konstantopoulos provides an overview of results about longest or heaviest paths on random directed graphs on the integers. First-order asymptotics are derived for heaviest paths that allow for weights both on the edges and the vertices, where the weights on the edges are signed. For sparse graphs convergence is shown of the weighted random graph to a certain weighted graph that can be constructed in terms of Poisson processes. Applications range from ecology to parallel computing.
- *Covariance structure behind breaking of ensemble equivalence in random graphs* by D. Garlaschelli, F. den Hollander and A. Roccaverde looks at random graphs subject to topological constraints. In the microcanonical ensemble the constraint is met by every realisation of the graph (“hard constraint”), in the canonical ensemble it is met only on average (“soft constraint”). Breaking of ensemble equivalence may occur when the size of the graph tends to infinity, signalled by a non-zero specific relative entropy of the two ensembles. It is shown that for the configuration model in the dense regime, the relative entropy is correctly described by a general formula recently conjectured by T. Squartini and D. Garlaschelli, which predicts that the specific relative entropy is determined by the scaling of the determinant of the matrix of canonical covariances of the constraints.
- *Near critical preferential attachment networks have small giant components* by M. Eckhoff, P. Mörters and M. Ortgièse studies the size of the giant component in preferential attachment models where every new vertex connects independently to the vertices already in the network. This model is believed to be in the same universality class as the usual preferential attachment model where every new vertex enters the network with a fixed number of edges, yet has an interesting percolation phase transition when the degree power-law exponent τ satisfies $\tau > 3$. This phase transition means that there is a value $\rho_c > 0$ such that for small edge densities $\rho < \rho_c$ every component of the graph comprises an asymptotically vanishing proportion of vertices, while for large edge densities $\rho > \rho_c$ there is a unique giant component comprising an asymptotically positive proportion of vertices. The authors study the decay in the size of the giant component as the critical edge density is approached from above. They show that the size decays very rapidly, like $\exp(-c/\sqrt{\rho - \rho_c})$ with an explicit constant $c > 0$ depending on the details of the model. Thus, the phase transition is of infinite order, which is in contrast to the behaviour of the class of rank-one models of scale-free networks, including the configuration model, where the decay is polynomial.
- *Weighted exponential random graph models: scope and large network limits* by S. Bhamidi, S. Chakraborty, S. Cranmer and B. Desmarais investigates how to generalize the exponential random graph to a context where the edges have weights associated to them. Such weights could be related to debts in financial networks, or to the multiplicity of interactions in social networks. Due to the importance of exponential random graphs as key models for network structure, such generalizations are needed to be able to extend the powerful exponential random graph results beyond binary edge settings. Analogously to fundamental results derived for standard (unweighted) exponential random graph models in the work of S. Chatterjee and P. Diaconis, the authors derive limiting results for the structure of these models as the number of nodes goes to infinity. Their results are applicable for a wide variety of base measures including measures with unbounded support. Further, they derive sufficient conditions for continuity of functionals in the specification of the model including conditions on nodal covariates.

- *The tail does not determine the size of the giant* by M. Deijfen, S. Rosengren and P. Trapman studies the dependence on the degree distribution of the size of the giant component in configuration models. The largest component is unique and satisfies a law of large numbers when the degree distribution does. The limiting proportion of vertices in the giant component is given by a well-known expression involving the generating function of the degree distribution. The authors argue that the distribution over small degrees is more important for the size of the giant component than the distribution over large degrees, as captured by a power law for the degree sequence. Thus, the tail behavior of the degree distribution does not play the same crucial role for the size of the giant as it does for many other properties of the graph, such as the percolation phase transition. Upper and lower bounds for the component size are derived for an arbitrary distribution over small degrees, up to a truncation value L , and given expected degree. Numerical implementations show that these bounds are close already for small values of L , which shows asymptotic independence of the precise tail of the degree distribution. On the other hand, examples illustrate that, for a fixed degree tail, the component size can vary substantially depending on the distribution over small degrees.
- *Triadic closure in configuration models with unbounded degree fluctuations* by R. van der Hofstad, J.S.H. van Leeuwen, and C. Stegehuis studies the clustering spectrum $c(k)$ in the erased configuration model in the scale-free regime, where the degree power-law exponent τ satisfies $\tau \in (2, 3)$, in the large-graph limit. Here, $c(k)$ is the proportion of edges present between the neighbors of all vertices of degree k . Alternatively, $c(k)$ equals the probability that two neighbors of a degree- k node are neighbors themselves. The authors show that $c(k)$ progressively falls off with k as well as in the graph size n and eventually for $k = \Omega(\sqrt{n})$ settles on a power law $c(k) \sim n^{5-2\tau} k^{-2(3-\tau)}$. Such a fall-off has been observed in many real-world networks. The results agree with recent results for the rank-1 inhomogeneous random graphs (or the hidden-variable model) and also give the expected number of triangles in the configuration model when counting triangles only once despite the presence of multi-edges. The proof shows that only triangles consisting of triplets with uniquely specified degrees contribute to the triangle counting. These uniquely specified degrees can be interpreted in terms of a variational problem. The proofs rely on concentration methods (first and second moment methods), and a careful analysis of the limiting asymptotics given in terms of a specific integral.
- *Soft communities in similarity space* by G. García-Pérez, M.Á. Serrano and M. Boguñá studies a generalization of the so-called S^1 model, which is a model producing graphs by assigning nodes uniformly random angular and radial coordinates, and connecting pairs of nodes with a probability that decreases with their angular distance. The generalization consists in replacing the uniform distribution of angular coordinates with a heterogeneous one featuring groups of nodes that are clustered, i.e. closer than expected in the angular dimension. This modification can give rise to community structure in the resulting realizations of the graph. The authors explore this modification by assuming a certain dependence of radial coordinates on the angular ones and discuss possible applications of the model.
- *Network Geometry and Complexity* by D. Mulder and G. Bianconi discusses a class of network models that extend traditional ‘pairwise’ graph models to simplicial complexes and cell-complexes that account for higher-order interactions. While an ordinary edge in a graph encodes a 2-way interaction between two nodes, a d -simplex in a simplicial complex encodes $(d+1)$ -way interactions taking place simultaneously among all pairs of a set of $d+1$ nodes. In such models, nodes ($d=0$) and edges ($d=1$) generalize to triangles ($d=2$), tetrahedra ($d=3$), and so on. Further generalizing this concept, the

authors introduce a model of cell-complexes formed by gluing convex polytopes along their faces and study some of its properties, most notably community structure, degree distribution, hyperbolicity and spectral dimension.

- *Sparse maximum-entropy random graphs with a given power-law degree distribution* by P. van der Hoorn, G. Lippner and D. Krioukov studies properties of the so-called hypersoft configuration model (HSCM). While the ‘hard’ configuration model is a microcanonical ensemble where each node has a fixed degree and the ‘soft’ configuration model is a canonical ensemble where each node has a fixed expected degree, the HSCM is a further relaxed ‘hypercanonical’ ensemble (similar to inhomogeneous random graphs) where only the degree distribution and expected degree are fixed. The authors prove properties of the HSCM with given power-law degree distribution, in particular its unbiasedness, i.e., the sufficiently fast convergence of its entropy to that of a certain maximum-entropy graphon.

(II) Random Processes on Random Networks

- *Mixing time bounds via bottleneck sequences* by L. Addario-Berry and M.I. Roberts provides new upper bounds for mixing times of general finite Markov chains. These bounds are used to show that the total variation mixing time is robust under rough isometry for bounded degree graphs that are roughly isometric to trees. The mixing time is related to bottlenecks in the graph. The main idea of the paper is that the key quantities are the number and the strength of the bottlenecks that can be lined up in a row. A bottleneck sequence is defined to quantify this concept.
- *Corrected mean-field model for random sequential adsorption on random geometric graphs* by S. Dhara, J. van Leeuwen and D. Mukherjee introduces a solvable model for random sequential adsorption of non-overlapping congruent spheres in the d -dimensional Euclidean space with $d \geq 2$. In the model considered, spheres arrive sequentially at uniformly chosen locations in space and are accepted only when there is no overlap with previously deposited spheres. The authors succeed in analyzing the fraction of accepted spheres in an asymptotic regime.
- *Mean field analysis of personalized PageRank with Implications for local graph clustering* by K. Avrachenkov, A. Kadavankandy and N. Litvak focusses on establishing a mean-field limit for a model of Personalized PageRank on the Erdős-Rényi random graph with a denser planted Erdős-Rényi subgraph. In the asymptotic regimes considered, the values of Personalized PageRank concentrate around the mean-field value. Also attention is paid to optimizing the damping factor. The results have practical implications, as they help to understand the applicability of Personalized PageRank and its limitations for local graph clustering.
- *Replica bounds by combinatorial interpolation for diluted spin systems* by M. Lelarge and M. Oulamara consider the free energy of diluted random constraints satisfaction problems. A simplified proof is given for bounds that were derived earlier in the literature for a general degree distribution and a general degree distribution. As a corollary, this leads to new bounds for the size of the largest independent set (also known as hard core model) in a large random regular graph. The proof uses a combinatorial interpolation argument based on biased random walks.
- *Central limit theorem for exponentially quasi-local statistics of spin models on Cayley graphs* by T.R.R. Annapareddy, S. Vadlamani and D. Yogeshwaran derives central limit theorems for local statistics and exponentially quasi-local statistics of spin models on discrete Cayley graphs with polynomial growth, and for random fields on discrete

Cayley graphs taking values in a countable space. The results are illustrated with specific examples of lattice spin models and statistics arising in computational topology, statistical physics and random networks. Examples of clustering spin models include quasi-associated spin models with fast decaying covariances like the off-critical Ising model, level sets of Gaussian random fields with fast decaying covariances like the massive Gaussian free field and determinantal point processes with fast decaying kernels. Examples of local statistics include intrinsic volumes, face counts, component counts of random cubical complexes. Examples of exponentially quasi-local statistics include nearest neighbour distances in spin models and Betti numbers of sub-critical random cubical complexes.

- *Random forests and networks analysis* by L. Avena, F. Castell, A. Gaudillièrre and C. Melot analyzes Wilson's algorithm, a simple and efficient algorithm based on loop-erased random walks to sample weighted trees or forests spanning a given graph. Earlier work by the authors focused on applications of spanning rooted forests on finite graphs. The main conclusions are reviewed by collecting related theorems, algorithms, heuristics and numerical experiments. An overview of determinantal structures and efficient sampling procedures is followed by four applications: (1) a random-walk-based notion of well-distributed points in a graph; (2) a description of metastable dynamics in finite settings by means of Markov intertwining dualities; (3) coarse graining schemes for networks and associated processes; (4) wavelets-like pyramidal algorithms for graph signals.
- *Tackling information asymmetry in networks: a new entropy-based ranking index* by P. Barucca, G. Caldarelli and T. Squartini introduces an entropy-based ranking index to deal with asymmetries in the knowledge of network connections between agents in socio-economic systems. Finding the most important players in a network is of paramount interest, whether 'importance' is to be understood in a functional or in a structural way. A significant part of the information is captured by the network of connections between agents, which leads to an identification of the most central nodes. The assumption that the entire network knows the connectivity patterns leads to a standard hidden variable (or 'soft' configuration model). Instead, the authors assume that a node only knows its local connections (e.g., its ego-network), and quantify the importance of a node based on the amount of information that this knowledge gives. The different interlinkages patterns that agents establish may lead to asymmetries in the knowledge of the network structure. Since this entails a different ability of quantifying relevant systemic properties (e.g. the risk of financial contagion in a network of liabilities), agents capable of providing a better estimate of (otherwise) unaccessible network properties ultimately have a competitive advantage. A novel index, InfoRank, is introduced to measure the quality of the information possessed by each node, and the Shannon entropy of the ensemble conditioned on the node-specific information is computed. The performance of this novel ranking procedure is tested in terms of the reconstruction accuracy of the (unaccessible) network structure. It is shown that InfoRank outperforms other popular centrality measures, such as PageRank, closeness- and degree-centrality, in identifying the 'most informative' and thus most central nodes.
- *Large deviations for the annealed Ising model on inhomogeneous random graphs: spins and degrees* by S. Dommers, C. Giardinà, C. Giberti and R. van der Hofstad investigates properties of the Ising model on generalized random graphs (sometimes also called the hidden-variable model or the soft configuration model). Extending their recent work with M.L. Prioriello, the authors consider the annealed Ising model, obtained by taking the expectation with respect to the random graph in the numerator and denominator in the

Boltzmann distribution of the Ising model on the graph. This annealed Ising model is believed (and in many settings proved) to be in the same universality class as the original Ising model, but is in many cases easier to study. The authors prove a large deviation principle for the total spin and the number of edges, as well as detailed results on how the annealing over the Ising model changes the degrees of the vertices in the graph. Interestingly, the average degree of the graph under the annealed Ising model is larger than that under the original graph, showing the enormous effect the annealing has on the arising graph structure.

- *Weighted distances in scale-free configuration models* by E. Adriaans and J. Komjáthy studies the topology of weighted configuration models in the regime where the degree distribution of the graph has infinite variance. The weights are chosen to be i.i.d. random variables, giving rise to first-passage percolation on the configuration model. Contrary to the setting where the degree distribution has finite variance, for which the behavior is universal, there are different phases depending on the tail of the distribution function close to zero. The authors investigate the weighted distance or traversal time between two uniformly chosen vertices, called typical weighted distance. It was known that if the underlying age-dependent branching process approximating the local neighborhoods of vertices produces infinitely many individuals in finite time, called an explosive branching process, then typical weighted distances converge in distribution to a bounded random variable. The authors focus on a non-explosive branching process for which the degree distribution has a power-law tail with exponent $\tau \in (2, 3)$. They determine the first order of magnitude of typical distances in this regime for arbitrary edge-weight distributions. They prove that typical weighted distances tend to infinity with the amount of vertices and, by choosing an appropriate weight distribution, can be tuned to be any growing function that is $O(\log \log n)$, where n is the number of vertices in the graph. Thus, typical weighted distances can be anything in between the typical graph distance and a bounded random variable.
- *Eigenvector localization in real networks and its implications for epidemic spreading* by C. Castellano and R. Pastor-Satorras studies the properties of the largest eigenvalue (LEV) of the adjacency matrix of a graph in conjunction with the localization of the associated principal eigenvector (PEV). The authors extend previous results that had identified two possible values for the LEV, corresponding to two competing localization patterns of the PEV, namely either on the star formed by the node with largest degree and its neighbors or on the dense subgraph known as the K -core with maximum index. They find that, in a more general setting, the previous picture is not necessarily correct, as the PEV can be localized on both subgraphs simultaneously, or even on a set of nodes with no overlap with these subgraphs. They finally show that this result has non-intuitive consequences for optimal immunization strategies against the spreading of epidemics on networks.

(III) Applications of Networks

- *Queues on a dynamically evolving graph* by M. Mandjes, N. Starreveld and R. Bekker considers a population process on a queueing network with a structure that changes over time. Examples in which this model applies include communication networks, road traffic networks, various physics-motivated networks, and chemical reaction networks. For this system an in-depth analysis is performed, leading to recursions for the moments and an algorithm to evaluate the (time-dependent and stationary) moments and a diffusion limit for the joint queue length process.
- *The supermarket model with bounded queue lengths in equilibrium* by G. Brightwell, M. Fairthorne and M. Luczak considers a multi-server queueing model in which each arriving customer selects multiple servers uniformly at random, and joins the queue of a least-loaded server amongst those chosen. In various asymptotic regimes the behavior of the resulting model (usually referred to as the supermarket model) is analyzed. In a specific regime it can be proven that, in equilibrium, with high probability the queues remain bounded.
- *Customer sojourn time in GI/GI/1 feedback queue in the presence of heavy tails* by S. Foss and M. Miyazawa looks at a GI/GI/1 with feedback, in which the service times are (intermediately) regularly varying. The focus is on identifying sojourn-time tail asymptotics in two cases: the customer arrives in an empty system, and the customer arrives in the system in the stationary regime. The proofs reflect the principle-of-a-single-big-jump. In the case of Poisson arrivals, more explicit formulae are derived.
- *Metastability of queueing networks with mobile servers* by F. Baccelli, A. Rybko, S. Shlosman and A. Vladimirov studies a symmetric queueing network with moving servers, operating under a FIFO service discipline. It is observed that the mean-field limit dynamics exhibits unexpected behavior, which can be explained as a metastability phenomenon. Large enough finite symmetric networks on regular graphs are proved to be transient for arbitrarily small inflow rates. However, the limiting non-linear Markov process has at least two stationary solutions.
- *From ecology to finance (and back?): a review on entropy-based null models for the analysis of bipartite networks* by M.J. Straka, G. Caldarelli, T. Squartini and F. Saracco presents a comprehensive account of recent results in the modelling and analysis of bipartite networks. The authors focus on null models constructed as canonical ensembles of random bipartite graphs with given degree sequence on both layers of nodes. They review powerful applications of this class of bipartite configuration models to real-world ecological and economic networks, including the reconstruction of network topology from partial node-specific information and the statistical validation of empirical structural patterns such as nestedness, motifs, communities and specialization/diversification patterns.
- *Layer Communities in Multiplex Networks* by T.-C. Kao and M. Porter looks at networks where edges belong to multiple types or layers. The authors propose a method to cluster these layers in terms of the similarity between the topology of the graphs associated with each edge type. The calculation of the inter-layer similarity is based on the measured overlaps of edges between pairs of layers and is used for the subsequent detection of communities in the space of partitions of layers. The authors apply their method to both real-world and simulated multiplex networks. The results confirm that the approach is able to detect reasonable hierarchies of layers.
- *A framework for imperfectly observed networks* by D. Aldous and X. Li presents a formalism for analysing networks whose precise structure is subject to uncertainty and

is only partially observed. In particular, the authors consider a situation where there is a ‘true’ underlying weighted graph whose edge weights represent the frequency with which a direct interaction between the different pairs of nodes can be observed. Assuming that repeated observations of these interactions are generated from the true edge weights via a Poisson process, the authors derive a series of results for the assessment of the statistical reliability of hypothetical measurements of various network properties.