Extreme Value Statistics for Evolving Random Networks

Natalia Markovich [1,*] and Marijus Vaičiulis [2]

1   V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 117997 Moscow, Russia
2   Institute of Data Science and Digital Technologies, Vilnius University, Akademijos St. 4,
LT-08663 Vilnius, Lithuania; marijus.vaiciulis@mif.vu.lt
*   Correspondence: markovic@ipu.rssi.ru

Abstract: Our objective is to survey recent results concerning the evolution of random networks and related extreme value statistics, which are a subject of interest due to numerous applications. Our survey concerns the statistical methodology but not the structure of random networks. We focus on the problems arising in evolving networks mainly due to the heavy-tailed nature of node indices. Tail and extremal indices of the node influence characteristics like in-degrees, out-degrees, PageRanks, and Max-linear models arising in the evolving random networks are discussed. Related topics like preferential and clustering attachments, community detection, stationarity and dependence of graphs, information spreading, finding the most influential leading nodes and communities, and related methods are surveyed. This survey tries to propose possible solutions to unsolved problems, like testing the stationarity and dependence of random graphs using known results obtained for random sequences. We provide a discussion of unsolved or insufficiently developed problems like the distribution of triangle and circle counts in evolving networks, or the clustering attachment and the local dependence of the modularity, the impact of node or edge deletion at each step of evolution on extreme value statistics, among many others. Considering existing techniques of community detection, we pay attention to such related topics as coloring graphs and anomaly detection by machine learning algorithms based on extreme value theory. In order to understand how one can compute tail and extremal indices on random graphs, we provide a structured and comprehensive review of their estimators obtained for random sequences. Methods to calculate the PageRank and PageRank vector are shortly presented. This survey aims to provide a better understanding of the directions in which the study of random networks has been done and how extreme value analysis developed for random sequences can be applied to random networks.

Keywords: random network; evolution; PageRank; Max-linear model; tail index; extremal index; community detection; preferential attachment; clustering attachment; information spreading; leading nodes

MSC: 62G32; 90B15

1. Introduction

    The study of power-law real-world random networks attracts significant attention due to the many fields of their application, e.g., citation, phone-call, cellular, urban transport, and economic trade networks, among others [1–3]. Several monographs [4–7] have certainly contributed to basic concepts of random network theory and practice. "Real networks differ from random graphs in that often their degree distribution follows a power-law" [1]. Heavy-tailed distributions are accepted as realistic models for many phenomena in random networks and graphs. Among the recent applications, we find graphical models that are used to model extremal dependence or causality between events [8,9], including a bond percolation; see [10] among others. The distributions of various extremal characteristics of random discrete structures, such as the maximum number of common neighbors of a set of vertices in the

graph, the maximum codegree, and the maximum number of cliques sharing a given vertex in binomial random hypergraphs, are derived in [11,12].

In our survey, we focus on the standard configuration model, i.e., a model of a random graph with arbitrary degree distribution [13]. Let us denote sets of nodes (or vertices) and edges as $V(n)$ and $E(n)$, respectively, their cardinality as $\|V(n)\|$, $\|E(n)\|$, and the number of edges as n. A graph is a pair of unordered sets $G(n) = (V(n), E(n))$. Our survey concerns random graphs evolved using preferential or clustering attachment of new nodes. Considering extremes in directed graphs, recent results in [14,15] concerning the tail and extremal indices of non-stationary sums and maxima of regularly varying random variables (r.v.s) are represented in the new context of evolving directed random graphs [16]. The relation between the tail and extremal indices of the latter sums and maxima and of PageRanks and the Max-linear models is discussed in [16]. This novel approach allows us to find the tail and extremal indices of PageRanks and the Max-linear models of random graphs for attachment methods providing that the distribution tails of the latter measures are regularly varying, e.g., the preferential attachment.

Special attention is devoted to the impact of the removal of vertices and/or edges on a graph topology and the tail and extremal indices of node influence measures during evolution. An important item of the latter modeling is a partitioning of graphs into communities that constitute some node clusters. The nodes in the communities have many mutual links and a few links with other communities. Such communities may be considered independent or weakly dependent sets of nodes. The concept can be helpful for the statistical analysis of real data.

Our objective is to survey results regarding extreme value statistics and the tail and extremal indices for the graph enlargement and network evolution obtained in the literature. The tail index shows the heaviness of the distribution tail of an underlying r.v. The extremal index measures a local dependence or a cluster structure of a random sequence. It is used to determine a limit distribution of the maximum of a stationary sequence of r.v.s [17,18]. To determine the extremal index of such node influence indices as PageRank and the Max-linear model, it is proposed in [16] to consider the nodes as roots of elementary trees with the nearest neighbor nodes. Then PageRanks and the Max-linear models of the roots are calculated using sums and maxima of PageRanks of their nearest neighbors that have in-coming edges to the roots. Using Theorem 1 in [16], the tail and extremal indices of PageRanks and the Max-linear models are determined by the latter indices of the neighbor nodes belonging to the most heavy-tailed communities of the network.

The attachment of new nodes usually begins from a seed network consisting of a single node or node communities. The node indices in the communities may be heterogeneous. The nodes of the graph cannot be definitely enumerated since $V(n)$ is an unordered set. This creates an obstacle to defining the stationarity in the graphs. In [19], it is determined that "a graph is stationary if, for all finite sets of vertices with the same adjacency matrices, the joint distributions of their in- and out-degrees are the same." Tail and extremal indices of sets of injected nodes considered in time depend both on the choice of the seed network and the attachment policy. There are several open problems, namely, (1) how to define and test the stationarity on graphs; and (2) how to determine the dependence on graphs, among other problems. We survey results that may shed light on these problems.

An application related to network evolution is given with information spreading. The survey of spreading methods is also given in the paper.

We start with our methodology and contributions in Section 2. A power-law distribution and regularly varying distributions arising in random graphs, including evolving ones, are recalled in Section 3. Regularly varying distributions of PageRanks and the Max-linear models used as node influence indices are discussed, too. How to determine and test the stationarity and dependence on random graphs is discussed in Section 4. We review preferential and clustering attachment tools and distributions of triangle counts, the topics related to the clustering coefficient, and the attachment probability, as well as a short representation of other models of random networks in Section 5. Community

detection and related topics like graph coloring and anomaly detection are discussed in Section 6. In Section 7, a spreading of information in evolving graphs and its relation to a preferential attachment are considered. Conclusions and a discussion are presented in Section 8. Auxiliary materials are represented in the Appendices A–D.

2. Our Methodology and Contributions

2.1. Methodology

This survey focuses on the statistical methodology based on extreme value theory and its application to evolving random networks rather than on the structure of random graphs. We do not aim to fit graph models to the observed networks. Our objective is to find heavy-tailed phenomena on random graphs. As the node influence indices such as the in-degree, PageRank, and Max-linear model are heavy-tailed distributed, particularly, if the preferential attachment is used for evolution, and each of the latter indices can be dependent due to links between nodes, the described methods are based on the statistical theory of extremes of random sequences, an asymptotic distribution of the maximum of a finite number of r.v.s (see the Fisher–Tippet–Gnedenko's theorem [17]).

The tail and extremal indices are key measures of the extreme value theory. To obtain the latter indices of the PageRank and Max-linear model in the evolving graphs, we use the results regarding the same indices for sums and maxima of random length non-stationary random sequences derived in [14] and extended in [15]. To this end, communities of graphs may be used as a series of r.v.s. To partition graphs into communities, one needs specific algorithms. Considering existing techniques of community detection, we pay attention to such related topics as coloring graphs and anomaly detection using machine learning algorithms based on extreme value theory.

Sometimes, it is required that the node influence indices in the community have to be stationary distributed, or communities with the smallest tail index have to be independent or weakly dependent. Then tests of the stationarity and dependence are needed.

To estimate the tail index, we use bias-reduced estimators coupled with methods to select a number of the largest order statistics related to the distribution tail.

To estimate the extremal index in random graphs, we use the interval estimator [20] that is modified in [16,21]. Nonparametric estimators of the extremal index require , as a rule, the selection of a threshold and/or another declustering parameter, e.g., a block size [17]. A rather simple interval estimator needs only the choice of the threshold and demonstrates a good accuracy.

Preferential and clustering attachments are considered to model the graph evolution. The finding of leading communities for a fast information spreading presented in the survey is based on methods proposed in [22,23].

2.2. Contributions

This survey is an attempt to summarize the achievements of many authors regarding extremes in random graphs and to provide a structured overview of related topics like community detection and attachment methods for graph evolution that are needed for our approaches. Our achievements concern topics like the network classification and the choice of leading communities using tail indices of communities, the tail and extremal indices of PageRank, and the Max-linear model for evolving graphs with and without node and edge deletion.

The survey tries to propose possible solutions to unsolved problems, like a testing of stationarity and dependence of random graphs using known results obtained for random sequences. We provide a discussion of unsolved or not sufficiently developed problems like the distribution of triangle and circle counts in evolving networks, or the clustering attachment and the local dependence of the modularity, the impact of node or edge deletion at each step of the evolution on extreme value statistics, among many others. Tail index estimators derived for random sequences are proposed to be adapted to random graphs.

We also suggest machine learning methods for anomaly detection as partition methods for community detection.

3. Heavy-Tailed Distributed Node Influence Indices

3.1. Definitions

Let us consider the in- and out-degrees as well as PageRanks and the Max-linear models as node influence measures.

Let us begin with definitions. We recall that a discrete r.v. X exhibits a power-law distribution if

$$P(X = k) \sim Ck^{-(1+\iota)}, \quad k \to \infty, \tag{1}$$

holds for some positive constants C and ι ($a_n \sim b_n$, as usual, means that the sequences a_n and b_n are asymptotically equal, i.e., $a_n/b_n \to 1$, $n \to \infty$).

From (1) it follows that $\sum_{k=n+1}^{\infty} P(X = k) \sim (C/\iota)n^{-\iota}$, as $n \to \infty$; see, e.g., Ref. [24] for details. This fact, together with (1), implies that the power-law distribution satisfies a von Mises-type condition

$$\lim_{n \to \infty} \frac{nP(X = n)}{\sum_{k=n+1}^{\infty} P(X = k)} = \iota.$$

Thus, using Theorem 3 in [25], a discrete distribution satisfying (1) belongs to the class of distributions whose right tail $\bar{F}(x) := 1 - F(x)$, where $F(x)$ denotes the cumulative distribution function and is regularly varying at infinity with index $-\iota$ (the notation is $\bar{F} \in \mathrm{RV}_{-\iota}$), i.e.,

$$\lim_{t \to \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\iota}$$

for all $x > 0$. If $\bar{F} \in \mathrm{RV}_{-\iota}$, then it is always possible to represent $\bar{F}(x)$ as $\ell(x)x^{-\iota}$, where $\ell(x)$ is a slowly varying function ($\ell(x) \in \mathrm{RV}_0$), i.e., by definition $\lim_{x \to \infty} \ell(tx)/\ell(x) = 1$ for any $t > 0$, see e.g., p. 13 in [26].

Using [18], p. 67, a stationary sequence $\{X_n\}_{n \geq 1}$ has an extremal index $\theta \in [0, 1]$ if for each $0 < \tau < \infty$ there is a sequence of real numbers $u_n = u_n(\tau)$ such that

$$\lim_{n \to \infty} n(1 - F(u_n)) = \tau, \qquad \lim_{n \to \infty} P\{M_n \leq u_n\} = e^{-\tau\theta} \tag{2}$$

holds for $M_n = \max\{X_1, \ldots, X_n\}$. Independent identically distributed (i.i.d.) r.v.s $\{X_n\}$ provide $\theta = 1$. The converse may be incorrect. The reciprocal of θ approximates a mean cluster size, i.e., the mean number of exceedances over a threshold per cluster. Then, the cluster of exceedances may imply a block of data with at least one extreme observation (an exceedance) over a high threshold [17].

PageRank vector $R = (R_1, ..., R_n)$ used using search engine Google is determined as the unique solution to the system of linear equations [27]

$$R_i = c \sum_{j:(j,i) \in E_n} \frac{R_j}{D_j} + \frac{c}{n} \sum_{j \in \mathfrak{D}} R_j + (1 - c)q_i, \qquad i = 1, \ldots, n. \tag{3}$$

Here, the first sum is taken over pages j that are linked to page i (the in-degree). D_j is the number of outgoing links of page j (the out-degree). $\mathfrak{D}$ is a set of dangling nodes, i.e., nodes without outgoing edges. $c \in (0, 1)$ is a damping factor. $q = (q_1, q_2, \ldots, q_n)$ is a personalization probability vector (user's preferences). It holds $q_i \geq 0$ and $\sum_{i=1}^{n} q_i = 1$. n is the total number of pages. The World Wide Web (WWW) is a huge interconnected graph where nodes imply pages. The PageRank was designed to rank pages on the Web in such a way that a page is important if many important pages have a hyperlink to it [28].

A Max-linear model [8] is obtained using the following expression

$$R_i = \bigvee_{j:(j,i) \in E_n} A_j R_j \vee Q_i, \quad i = 1, \ldots, n, \tag{4}$$

where $\{Q_i\}$ are independent continuous r.v.s.

3.2. Power-Law of In- and Out-Degrees

Power-law distributions with different exponents of a node's in- and out-degree in evolving networks are approved by numerous theoretical and empirical studies [1]. A preferential attachment (PA) (see details in Section 5.1) is suggested to explain conjectured power-law degree distributions in real networks [2]. Tail indices of in- and out-degrees and the power-law model of their asymptotic distributions were obtained in [29,30] depending on parameters of specific linear PA schemes named $\alpha-$, $\beta-$ and $\gamma-$ schemes that are used for growing networks. Namely, these tail indices are obtained as

$$\alpha_{in} = \frac{1 + \delta_{in}(\alpha + \gamma)}{\alpha + \beta}, \quad \alpha_{out} = \frac{1 + \delta_{out}(\alpha + \gamma)}{\beta + \gamma}, \tag{5}$$

where nonnegative parameters of the PA schemes $(\alpha, \beta, \gamma, \delta_{in}, \delta_{out})$ are such that $\alpha + \beta + \gamma = 1$, see Section 5.1 and formula (2.9) in [30]. A joint distribution of the in- and out-degrees is derived to have jointly regularly varying tail [29]. As far as we know, the tail index of the PageRank is not yet obtained in the literature. In [31], the existence of the limit of a graph in a local weak convergence sense was derived for directed graphs, but a power-law tail of the PageRank of the root in the limiting graph (that is not a branching tree) is not yet proven.

The distributions of in-degree and the PageRank are found to be similar [32]. Assuming the weak degree correlations in the Web graph, the PageRank vector may be approximated using the node in-degrees sufficiently accurate [32]. A statistical analysis of Web graphs has shown that distribution tails of the in-degree and the PageRank are a power-law (1) with the same exponent $\iota \approx 1.1$, Ref. [33]. The in-degree is available statistics that are gathered using a search engine. Then it is easier to use the in-degree rather than the PageRank in practice.

The node degree exponent ι of several networks was found to satisfy $\iota \in (2,3)$ [34]. Uniform random graphs with specified degree sequences are commonly used to model (power-law) real-world networks (see, Refs. [35,36] for a survey). In [37], a graph Hamiltonian is proposed as a method to model heterogeneous clustered graphs.

3.3. Regularly Varying Distributions of PageRanks and Max-Linear Models

3.3.1. PageRank and Max-Linear Model as Solutions of Fixed-Point Problems

Heavy-tailed regularly varying limit distributions of PageRank were derived in [38,39], for static networks without any evolution in time. A similar result has been derived in [40] for the Max-linear model.

The latter papers exploit a probabilistic approach to study (3). A Web page that is randomly chosen in a Web graph is then represented as a root of a Galton–Watson tree with random in- and out-degrees. Its PageRank can be modeled as an r.v. R. The latter is the solution of the following fixed-point problem [28,38,39,41]

$$R =^D \sum_{j=1}^{N} A_j R_j + Q. \tag{6}$$

In [39], $\{R_j\}$ are assumed to be i.i.d. copies of R. A personalization value of the vertex is assumed to have a bounded expectation $E(Q) < 1$. N means the in-degree. All r.v.s in the triple $(Q, N, \{A_j\})$ are real-valued. $=^D$ denotes equality in distribution. The A_i's are random coefficients that equal c/D_i in [39]. D_i are the node out-degree such that $D_i \geq 1$ Let us recall the following Assumptions A. $\{R_j\}$ are i.i.d. and independent of $(Q, N, \{A_j\})$ with $\{A_j\}$ independent of (N, Q). N and Q can be mutually dependent. Then it is stated in [38,39] that the stationary distribution of R in (6) is stated to be regularly varying, and its tail index is determined using the most heavy-tailed distributed term in the regularly varying distributed pair (N, Q). Considering iterations of the PageRank, the

initial distribution of $R^{(0)}$ impacts on the heaviness of the tail of the kth iteration $R^{(k)}$ of the PageRank (see, Ref. [39], Theorem 3.2).

The latter results were extended in [42]. The unique solution of (6) is proven to be intermediate regularly varying. The class of intermediate regularly varying distributions such that $\lim_{\alpha \uparrow 1} \lim \sup_{x \to \infty} \overline{F}(\alpha x)/\overline{F}(x) = 1$ includes regularly varying distributions. if Q or N has an intermediate regularly varying distribution, or (Q, N) has a two-dimensional regularly varying distribution. The multivariate version of (6)

$$R(i) =^D \sum_{k=1}^{K} \sum_{m=1}^{N^{(k)}(i)} R_m(k) + Q(i),$$

where $R_m(k) =^D R(k)$ holds, is considered subject to similar independence assumptions and regularly varying statements, assuming that $N^{(k)}(i)$ is a number of type-k children of a type-i ancestor and considering a multi-type Galton–Watson tree.

Under Assumptions A the power-law tail $P\{|R| > x\} \sim Hx^{-\iota}, \iota > 0, H > 0$ as $x \to \infty$ of the so-called "minimal/endogeneous" solution of the equation

$$R =^D \left(\bigvee_{j=1}^{N} A_j R_j \right) \vee Q, \tag{7}$$

is derived in [40] without a specification of ι.

3.3.2. PageRank and the Max-Linear Model as Sums and Maxima of Non-Stationary Sequences of Random Lengths

The reciprocal of the extremal index of random sequences is the limiting mean cluster size in the point process of exceedance times over a high threshold, i.e., the mean number of exceedances per cluster. The cluster may imply a block of data with at least one extreme observation (an exceedance) over a sufficiently high threshold [17].

The random lengths sum and maximum at the right-hand sides of (6) and (7) were considered in [16] omitting the independence Assumptions A. Based on results in [14,15], it is derived in Theorem 1 of Ref. [16] that the tail and extremal indices of PageRank and the Max-linear model of newly appended nodes in evolving networks are determined using the latter indices of the most heavy-tailed community(ies), i.e., the communities with a minimum tail index. This result also allows us to classify newly appended nodes. Ranking the communities in ascending order of their tail indices, a new node is assigned to class i if it has at least one link to nodes of the community i. If the new node has links to different communities, then the i is a minimum number. In terms of citation networks, it means, for instance, that PageRank and the Max-linear model of a set of books cited by papers from the "dominating" communities having the minimum tail index inherit the same minimum tail index if the communities are independent or weakly dependent [16]. It is reasonable that the random number of "dominating" communities be bounded. There exists a "top" community among the "dominating" ones such that its maximum PageRank is the largest. Then the Max-linear models of the cited papers published earlier have the extremal index of the "top" community. If the communities with the minimum tail index are independent or weakly dependent, then PageRanks of the cited papers have the same extremal index as the "top" community. If the "dominating" community is unique, then PageRanks and the Max-linear models of the cited papers have its tail and extremal indices.

In [15,16], $\{Y_{n,i} : n, i \geq 1\}$ is a double-indexed array of nonnegative r.v.s such that the "row index" n indicates the time, and the "column index" i enumerates the series. The length N_n of "row" sequences $\{Y_{n,i} : i \geq 1\}$ for each n is random. For each i the "column" sequence $\{Y_{n,i} : n \geq 1\}$ is assumed to be strict-sense stationary with extremal index θ_i and having a regularly varying distribution tail

$$P\{Y_{n,i} > x\} = \ell_i(x)x^{-k_i}$$

with tail index $k_i > 0$ and a slowly varying function $\ell_i(x)$. There are no assumptions about the dependence structure in i. For a better understanding, we represent the matrix of $\{Y_{n,i} : n, i \geq 1\}$ as

$$
\begin{pmatrix}
cY_{1,1} & cY_{1,2} & cY_{1,N_1} & \ldots & 0 & 0 \\
cY_{2,1} & 0 & cY_{2,3} & \ldots & cY_{n,N_2} & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & cY_{n,2} & cY_{n,3} & \ldots & 0 & cY_{n,N_n}
\end{pmatrix},
\tag{8}
$$

$$
\begin{pmatrix}
k_1 & k_2 & k_3 & \ldots & \ldots & k_{N_n} \\
\theta_1 & \theta_2 & \theta_3 & \ldots & \ldots & \theta_{N_n}
\end{pmatrix},
$$

where $\{k_i\}$ and $\{\theta_i\}$ denote the tail and extremal indices of the "column" series in (8). Ref. [14] contains a simple result concerning the weighted sums and maxima over the "row" random-length sequences

$$
Y_n^*(z, N_n) = \max(z_1 Y_{n,1}, \ldots, z_{N_n} Y_{n,N_n}),
$$
$$
Y_n(z, N_n) = z_1 Y_{n,1} + \cdots + z_{N_n} Y_{n,N_n}, \qquad z_1, z_2, \ldots > 0.
$$

If the "column" series with a minimum tail index is unique, let us say, k_1, then the tail and extremal indices of both $Y_n^*(z, N_n)$ and $Y_n(z, N_n)$ are equal to k_1 and θ_1.

A random number d of the "column" series with a minimum tail index, which is plausible in random graphs, is considered in [15], Theorem 4. If the d "column" series are mutually independent, and independent of the remaining "columns," or weakly dependent (assumptions (A1) and (A2) in [15], respectively), and N_n and $\{Y_{n,i}\}$ are independent, then $Y_n^*(z, N_n)$ and $Y_n(z, N_n)$ have the same tail index k_1. Let us recall the latter assumptions for a fixed $d > 1$ proposed in [15].

(A1) The stationary sequences $\{Y_{n,i}\}_{n \geq 1}$, $i \in \{1, \ldots, d\}$ are mutually independent, and independent of the sequences $\{Y_{n,i}\}_{n \geq 1}$, $i \in \{d+1, \ldots, l_n\}$.

(A2) Assume $\{Y_{n,i}\}_{n \geq 1}$, $i \in \{1, \ldots, d\}$ satisfy the following conditions as $x \to \infty$

$$
\frac{P\{Y_{n,i} > x\}}{x^{-k_1} \ell_1(x)} \quad \to \quad c_i, \ i \in \{1, \ldots, d\},
$$

for some non-negative numbers c_i,

$$
\frac{P\{Y_{n,i} > x, Y_{n,j} > x\}}{x^{-k_1} \ell_1(x)} \quad \to \quad 0, \ i \neq j, \ i, j \in \{1, \ldots, d\}.
$$

By assumption (A4) in [15,16] there exists $i \in \{1, \ldots, d\}$ such that it holds

$$
P\{\max_{1 \leq j \leq d, j \neq i}(z_j M_n^{(j)}) > u_n, z_i M_n^{(i)} \leq u_n\} = o(1), \ n \to \infty,
\tag{9}
$$

where

$$
M_n^{(i)} \quad = \quad \max\{Y_{1,i}, Y_{2,i}, \ldots, Y_{n,i}\}, \ i \in \{1, \ldots, l_n\}, n \geq 1
$$

denotes the maximum over the ith "column" of (8). If (A4) holds, then $Y_n^*(z, N_n)$ has the extremal index θ_i. $Y_n(z, N_n)$ has the same extremal index if, in addition to (A4), (A1) or (A2) holds. The (A4) is valid for all d "column" series such that

$$
M_n^{(1)} \leq M_n^{(2)} \leq \ldots \leq M_n^{(d)}
\tag{10}
$$

holds.

In [16], the results obtained in [14,15] are applied to evolving random networks. To this end, the evolution starts from a seed network that is represented by matrix (8). The following recursions

$$Y_{i,j}^{(m)} = c \sum_{s=j}^{N_i} Y_{i,s}^{(m-1)} + Q_i, \tag{11}$$

$$X_{i,j}^{(m)} = \left(c \bigvee_{s=j}^{N_i} X_{i,s}^{(m-1)} \right) \vee Q_i, \quad \{X_{i,j}^{(0)}\} \equiv \{Y_{i,j}^{(0)}\}, \tag{12}$$

$m, i, j \geq 1$, where m is connected with the time, are considered. $Y_{i,j}^{(m)}$ and $X_{i,j}^{(m)}$ build the jth "column" series of the next generation matrix $A^{(m)}$ using $A^{(m-1)}$ starting by $A^{(0)}$. It may be, for instance, the following

$$A^{(0)} = \begin{pmatrix} Y_{1,1}^{(0)} & Y_{1,2}^{(0)} & Y_{1,3}^{(0)} & \cdots & 0 & Q_1 \\ Y_{2,1}^{(0)} & 0 & Y_{2,3}^{(0)} & \cdots & Y_{2,N_2}^{(0)} & Q_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Y_{n,1}^{(0)} & Y_{n,2}^{(0)} & Y_{n,3}^{(0)} & \cdots & Y_{n,N_n}^{(0)} & Q_n \end{pmatrix},$$

$$\begin{pmatrix} k_1^{(0)} & k_2^{(0)} & k_3^{(0)} & \cdots & k_N^{(0)} & k_{N+1}^{(0)} \\ \theta_1^{(0)} & \theta_2^{(0)} & \theta_3^{(0)} & \cdots & \theta_N^{(0)} & 1 \end{pmatrix}.$$

$\{Q_i\}$ is a sequence of i.i.d. r.v.s. Hence, its extremal index is equal to 1.

Network communities as "column" series of $A^{(0)}$

Network communities may be treated as columns of $A^{(0)}$.

Since the communities have random lengths, the columns in matrix $A^{(0)}$ may be supplemented with zeros. A zero sth element in the ith row $Y_{i,s}^{(0)}$, $s \geq 1$ of $A^{(0)}$ implies that the ith root node has no followers in the sth community or there is no link between them; see Figure 1a. For example, if a row corresponds to a set of papers citing a book, then zero implies that the book is not cited by a paper from the corresponding community. Using Theorem 4 in [15], it is provedn in Theorem 1 in [16] that sums and maxima $Y_{i,j}^{(m)}$ and $X_{i,j}^{(m)}$ at the mth step of the evolution inherit the corresponding tail and extremal indices of the most heavy-tailed "column" series of $A^{(0)}$ for any $m \geq 1$ under specific conditions. The conditions are weaker for maxima over rows of $A^{(0)}$ than for sums. The maxima require the condition (A4) (see, Formula (9)) that is valid for all d the most heavy-tailed "column" series such that (10) holds. Considering the latter communities with the same minimum tail index (that are not necessarily distributed the same way) as the "columns" and taking into account that their order does not matter, one can conclude that (10) is always fulfilled for maxima over communities.
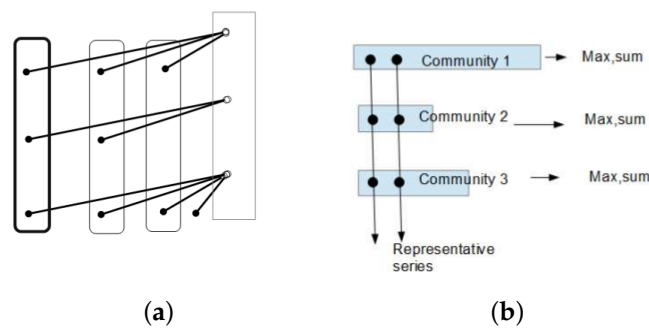
Figure 1. The root nodes and their nearest neighbors (marked by open and filled circles, respectively) where the community with the heaviest distribution tail is marked by a rectangle with a thick black line (**a**). The scheme of communities as the "row" sequences and the representative node series as the "column" series are demanded where the maxima and sums are taken over node influence indices in the communities: representative series may be formed by ranking PageRanks of nodes in the communities (reprinted from Figure 1a in [19]) (**b**).

Sums over rows of $A^{(0)}$ require the independence or weak dependence of the most heavy-tailed "column" series (see, namely, conditions (A1) and (A2) in [15,16]) additionally to (10). The number of the most heavy-tailed "column" series may be random. Theorem 1 in [16] is valid assuming that each row of $A^{(0)}$ contains at least one nonzero element in the columns with the minimum tail index, i.e., the columns with the most heavy-tailed distribution tail. Otherwise, the sums and maxima over rows will be non-stationary distributed with different tail indices. Theorem 1 in [16] states that the limit distributions of the recursions (11) and (12) are determined using the distributions of columns $Y_{n,i}^{(0)} : n, i \geq 1$. The proposed approach is valid for any attachment method that leads to PageRanks with regularly varying distribution tails, particularly, for a preferential attachment.

Example 1. *The approach in [16] may be interpreted in terms of citation networks. Namely, if newly published papers cite at least one paper from the most heavy-tailed communities published earlier, then the Max-linear models of the new papers have the same tail and extremal indices of such communities. If communities are weakly connected or disconnected, then the same result is valid for the PageRanks of the new papers.*

Network communities as "row" series of $A^{(0)}$

From another side, network communities may be interpreted as rows of $A^{(0)}$; see, Figure 1a in [19]. Such an approach has the advantage that sums and maxima over communities may be considered independent or weakly dependent r.v.s. Testing of the dependence on graphs is discussed in Section 4.3. The independent communities simplify a tail index estimation of sums and maxima over the communities. Let us recall that the communities are random length sub-graphs. The communities are allowed to be non-homogeneous. Namely, their node PageRanks may be non-stationary distributed in contrast to the previous case when the communities representing the most heavy-tailed "column" series of $A^{(0)}$ are to be stationary distributed.

Using a simulation study, it was found in [19] that the tail indices of the sums and maxima over communities are close to the minimum tail index of the representative series extracted from the communities. The representative series can be formed by taking one of the nodes of a community as a representative. To avoid numerous combinations, the *i*th representative series was chosen using the *i*th PageRank maxima within each community. Since communities have a random size, some nodes may fall into several representative series, leading to their dependence. In case of different pair-wise dependency among elements of the $d > 1$ "column" series (the representative series) with the minimum tail index, the sums and maxima calculated over rows (communities) may be non-stationary distributed [15]. The different pair-wise dependence between the members of the most

heavy-tailed communities is very restrictive and indefinite since nodes in the communities are unordered. In practice however, the representative series with a minimum tail index estimate is likely unique. Thus, the mentioned pair-wise dependency is not required. The results of the empirical study in [19] are in agreement with theoretical results in [14,15].

4. Stationarity and Dependence on Graphs

4.1. Interpretation of Stationarity on Graphs

In Section 3.3.2, the communities defined as the "column" sequences of matrices (8) (or $A^{(0)}$) have to be stationary distributed. Moreover, it is important to understand whether there are communities with the same tail index. The same issue may concern to representative series. The following questions arise. How can one interpret the stationarity on random graphs? How does one test a deviation in the tail indices?

One of the ideas to check non-stationarity is to use data blocks. This approach is usually applied to random sequences in different applications (see, e.g., Ref. [43]). To apply this idea to random graphs, one can partition communities into subcommunities (blocks) and estimate their tail indices, for instance, by methods recalled in Appendix A. Since the number of communities is random, one can apply a random graph coloring and stochastic decomposition methods of the symmetric adjacency matrix derived from the underlying undirected graph; see [44–46] among others (see, Section 6).

Another idea is to use stationary distributed Markov chains like the Metropolis algorithm (see, Refs. [47,48] among others) as sampling tools over networks and to test the stationarity of the obtained random sequences of node indices in a classical way.

For layered networks considered in [49], the homogeneity for networks implies that the replacing in the location and layer would lead to the same in-degree distribution.

4.2. Testing of a Change in the Tail Index among Communities

Here, we describe some tests applied to i.i.d. non-homogeneous random sequences containing observations with different tail indices. The latter tests can be related to random graphs within our vision.

In [43], an extreme value analysis for independent but non-identically distributed observations is considered. Assuming continuously changing extreme value indices (EVIs) (Assuming continuously changing extreme value indices (EVIs), the latter provide a non-parametric estimate for the EVI functional $\gamma(s)$, $s \in [0,1]$, based on the observations $X_1, \ldots, X_n$. The (positive) EVI constitutes the reciprocal of the tail index). these observations provide a non-parametric estimate for the EVI functional $\gamma(s)$, $s \in [0,1]$, based on the observations $X_1, \ldots, X_n$. The main idea in [43] is the following. The sample $X_1, \ldots, X_n$ is divided into blocks

$$X_{[(i-1)nh]+1}, \ldots, X_{[inh]}, \quad i = 1, 2, \ldots, [1/(2h)],$$

where each block contains $[2nh]$ observations. $[\cdot]$ denotes the integer part and $h = h(n)$ is a bandwidth such that $h \to 0$, $nh \to \infty$ as $n \to \infty$. To estimate $\gamma(s)$, $s \in [0,1]$ locally, Hill's estimator [50] is applied to each of the $[1/(2h)]$ blocks. Let $\hat{\gamma}_i$ denote the Hill estimate associated with the i-th block. Then the function $\Gamma(s) = \int_0^s \gamma(u) du$ is estimated by aggregating the local estimators $\hat{\gamma}_i$, $i = 1, 2, \ldots, [1/(2h)]$ as follows:

$$\hat{\Gamma}(s) = 2h \sum_{1 \leq i \leq 1/[2h]: \, i[2h] \leq s} \hat{\gamma}_i, \quad s \in [0,1].$$

The global estimator $\hat{\Gamma}(s)$ can be used to test a pre-specified parametric trend in the EVIs, see [43] for details.

Let

$$\mathcal{I}_n = \{I_n(v)\}, \qquad \mathcal{O}_n = \{O_n(v)\}, \qquad v \in V(n) \tag{13}$$

denote the sets of the in- and out-degrees of a directed graph $(V(n), E(n))$ at a time n. Due to the unorderedness of the elements of the sets $\mathcal{I}_n$ and $\mathcal{O}_n$, the estimation of the functional tail index $\alpha(s) = 1/\gamma(s)$, $s \in [0, 1]$ provided in [43] cannot be applied to random graphs directly. However, the idea can be useful in the statistic of directed graphs. Let $d(u, v)$ denote the distance in terms of the number of links between the nodes u and v. We define the h-neighborhood of node $v \in V(n)$ as follows

$$\mathcal{B}_n(v) = \{u \in V(n): \; d(u, v) < h \cdot N(n)\}, \qquad h \in (0, 1],$$

where $N(n)$ is the number of nodes in the network at time n. If the cardinality of a set $\mathcal{B}_n(v)$ is large enough (let us say, at least 500), we may obtain an estimate of the tail index $\alpha(v)$ based on $\{I_n(u), \; u \in \mathcal{B}_n(v)\}$. The set of estimates $\{\hat{\alpha}(v), v \in V(n)\}$ built by disjoint sets $\mathcal{B}_n(v)$ can be used for a preliminary analysis of the tail index behavior.

Assuming that the breakpoint of the tail shape behavior is known, a test for the null hypothesis that the tail index is constant over time is constructed in [51]. Considering the case with a single known breakpoint m, we have "split" observations

$$X_1, \ldots, X_m, X_{m+1}, \ldots, X_n$$

where $X_1, \ldots, X_m$ are i.i.d. r.v.s from a distribution F_1, for which $\bar{F}_1 \in \mathrm{RV}_{-\alpha_1}$, while $X_{m+1}, \ldots, X_n$ are i.i.d. r.v.s from a distribution F_2, which satisfies $\bar{F}_2 \in \mathrm{RV}_{-\alpha_2}$. Let us adopt below a test of [51] to directed random graphs. It is worth noting that this test can be adopted for undirected random graphs as well.

Let $(V(n)', E(n)')$ and $(V(n),'' E(n)'')$ be two non-intersected communities of $(V(n), E(n))$ with the corresponding sets of in-degrees $\mathcal{I}_n' = \{I_n(v), v \in V(n)'\}$ and $\mathcal{I}_n'' = \{I_n(v), v \in V(n)''\}$. In particular, the null hypothesis is

$$H_0: \quad \alpha_1 = \alpha_2 = \alpha,$$

where α_1 and α_2 are the tail indices of the in-degrees in $\mathcal{I}_n'$ and $\mathcal{I}_n''$, respectively. A modified statistic of Phillips and Loretan

$$S \;=\; \frac{k_1^*(\hat{\alpha}_2)^2(\hat{\alpha}_1/\hat{\alpha}_2 - 1)^2}{(\hat{\alpha}_1)^2 + (k_1^*/k_2^*)(\hat{\alpha}_2)^2} \tag{14}$$

is represented in [52]. Here, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are estimates of the tail index computed using Hill's estimator $\hat{\alpha}_{in}^{(i)}(k)$, $i \in \{1, 2\}$ from observations $\mathcal{I}_n'$ (with the optimal choice $k = k_1^*$ of the k largest order statistics) and $\mathcal{I}_n''$ (with the optimal choice $k = k_2^*$), respectively. The tail index estimation is described in Appendix A. Under the null hypothesis, statistic S converges in distribution to an r.v. that is chi-squared distributed with one degree of freedom. For example, at the 5% level of significance, the rejection region would be the interval $[3.841, +\infty)$.

One can calculate nodes' PageRanks (see, Appendix C) of communities for directed graphs or node degrees for undirected ones instead of the in-degrees, and then check the null hypothesis for pairs of the communities.

Note that the Hill estimator is weakly consistent for i.i.d. data; see [53]. The Hill and the ratio estimators may be applied to dependent data [54,55]. We recall that the ratio estimator introduced in [56] is a generalization of the Hill estimator (see Appendix A). The asymptotic behavior of the Hill estimator when independent observations are from non-identical distributions with comparable tails was discussed in [57]. The discussion of the tail index estimation in the case of dependent and non-stationary data can be found in [58]. There, in particular, a tail index estimator is proposed, whose construction is based on the block maxima method. The latter is weakly consistent when observations are from a non-stationary moving maxima model. One more application of the block maxima method for tail index estimation can be found in [59].

Although node indices within the communities are not i.i.d., the Hill estimation still works in practice; see [19,30]. Its consistency for undirected graphs is proven in [60], but the result regarding directed graphs remains, to the best of our knowledge, an unresolved problem.

4.3. Testing of Dependence on Graphs

Let us recall some methods which may be used to test the dependence of node influence indices of two communities.

The ρ-correlated Erdős–Rényi (ER) model was proposed to capture correlations between ER graphs in [61]. A random graph ER where an edge is sampled in i.i.d. fashion from a Bernoulli distribution with some parameter p is denoted as $ER(p)$. One can calculate a Pearson's correlation coefficient of two r.v.s belonging to two graphs. G_{ij} is a r.v. denoting whether there is an edge between nodes i and j in graph G. Two graphs G and H are called ρ-correlated $ER(p,q)$ if G is $ER(p)$, H is $ER(q)$, and the r.v.s G_{ij} and H_{ij} have Pearson's correlation

$$\rho = \frac{P\{G_{ij} = 1, H_{ij} = 1\} - pq}{\sqrt{p(1-p)q(1-q)}}$$

for all $\{i,j\} \in \binom{[n]}{2}$, where $[n]$ denotes an integer part, $\binom{[n]}{2}$ denotes a binomial coefficient. The null hypothesis of the graph independence test is $\rho = 0$, and the alternative is $\rho \neq 0$.

The Stochastic Block Model (SBM) generalizes ER graphs [4]. Using the SBM, which is parameterized by the block probability matrix $B \in [0,1]^{k \times k}$, where k is the number of blocks, one can check the dependence between two communities [62]. For $k = 1$ a SBM is ER. A block refers to a submatrix in the adjacency matrix formed by the edges connecting every node in community i to every node in community j. The entry $B_{i,j}$ gives the probability of an edge from a node in community i to a node in community j, for all $i, j \in \{1, \ldots, k\}$. The prior estimate of the number of communities can be taken as k. The choice of k is a sensitive point of the approach. One can generalize the ρ-correlated ER to the ρ-correlated SBM.

In [63], it is shown that the sample Pearson correlation coefficient fails to capture linear dependence between two random variables when their variances are infinite, which is realistic for real-world networks. In such networks, it is important to measure degree–degree dependencies for neighboring vertices. Examples where the Pearson coefficient converges to zero in a network with strong negative degree-degree dependencies and an example where this coefficient converges in distribution to an r.v. are provided. It is also shown that Spearman's ρ coefficient is able to reveal strong dependencies in large graphs.

A distance correlation is an extension of Pearson's correlation both to linear and nonlinear associations between two r.v.s or random vectors [64]. It takes values in $[0,1]$. The independence corresponds to the distance correlation equal to zero. The distance correlation is applied with a permutation test to check the dependence hypothesis because the numeration of nodes is undefined. The distance correlation is calculated first for an original pair of vectors. It is compared with those calculated with shuffles of these vectors. Under a ρ-correlated SBM, a naive permutation and Pearson's test for a conditional dependency graph model is shown to be invalid in [64]. As an alternative, a block-permutation procedure is proposed. The procedure is proved to be valid and consistent—even when the two graphs have different marginal distributions.

One can measure dependencies in data of a power-law graph using statistical inference for multivariate regular variation. To this end, the polar coordinate transform to the examined random vectors $\{X_i\}$ and $\{Y_i\}$, $i = 1, \ldots, n$ can be applied [30,55,65]. The empirical distribution function (edf) of the angular coordinates for the k largest values of the radial coordinate is calculated. The total dependence (or total independence) corresponds to the concentration of the edf to $\pi/4$ (or, to 0 or $\pi/2$). In the case of bivariate data, a Resnick–

Starica plot built with radii can be used to find a suitable value of k.

To detect the dependence between in- and out-degrees a polar coordinate transform

$$(I^a, O) \quad \rightarrow \quad \left(\sqrt{I^{2a} + O^2}, \arctan(O/I^a) \right) = (R, T)$$

is applied in [30] if the limiting random vector $(I, O) \sim F$ is jointly regularly varying. Here, $a = \tau_{in}/\tau_{out}$ is estimated in practice using the estimation of tail indices τ_{in} and τ_{out} of in- and out-degrees. The conditional distribution $P\{T \in \cdot | R > r\} \rightarrow S(\cdot)$ converges weakly to the angular measure $S(\cdot)$. The edf of $S(\cdot)$ is calculated using the sample angles $T_n(v) = \arctan(O_n(v)/I_n(v)^{\hat{a}})$ for vertices $v = 1, 2, \ldots, N(n)$ of the graph $G(n)$ at time n for which radial coordinates $R_n(v) > r$ exceed some large threshold r. In [30], the tail indices are proposed to be estimated with the Hill estimator without a rigorous justification of the latter for non-i.i.d. data. For the case of dependent data consisting of node degrees from preferential attachment models, the consistency of the Hill estimator has been shown in [60,66] but asymptotic normality of the Hill estimator in this setting remains an open problem.

A conditional dependence is often considered in graphs. Suppose cycles of nodes c_1 and c_2 may have joint vertices, but they do not have any edge in common. Then for such c_1 and c_2 the r.v.s Y_{c_1} and Y_{c_2} are conditionally independent given the node weights $W_1, \ldots, W_n$ [67]. Here, Y_c denotes the indicator that $c \in I(k)$ occurs as a cycle in the graph, taking into account two orientations and k starting points, and $I(k)$ denotes the set of potential cycles of length k.

Conditional independence is tightly linked to graphical models [8,10,68–70]. Nodes in graphical models are associated with variables, and the set of edges encodes the conditional independence relations. Let $X = (X_i)_{i \in I}$ be a random vector, and $X_A = (X_i)_{i \in A \subset I}$ denotes the sub-vector. Then $X_A \perp\!\!\!\perp X_B | X_{I \setminus (A \cup B)}$ is denoted if X_A and X_B are conditionally independent given $X_{I \setminus (A \cup B)}$. Let us consider disjoint subsets A, B, C of the set of vertices V of the undirected graph $G = (V, E)$. In terms of graphs, the conditional independence $A \perp\!\!\!\perp B | C$ implies that all paths from A to B pass through at least one vertex in C [68]. The conditional independence relates to max-stable random vectors. vectors. A random vector X is called max-stable if it satisfies the distributional equality $a_n X + b_n =^D \max(X^{(1)}, \ldots, X^{(n)})$ for independent copies $(X^{(1)}, \ldots, X^{(n)})$ of X for some normalizing sequences $a_n > 0$ and $b_n \in R$, where all operations are meant componentwise [70]. In [70], it is shown that for a max-stable random vector with positive continuous density, conditional independence implies unconditional independence.

Considering a sequence of random vectors $X(n) = (X_1(n), \ldots, X_d(n))^T \in R^d$, where $d = d(n)$ is a sequence of positive integers, the distribution of the maximal component of X and the distribution of the maximum of their independent copies are proved to be asymptotically equivalent under certain weakly sufficient dependence conditions, i.e.,

$$\left| P\{\max_{i \in [d]} X_i \leq x\} - \prod_{i \in [d]} P\{X_i \leq x\} \right| \quad \rightarrow \quad 0$$

for any fixed $x \in R$; see [11]. The notation $[d]$ is used for the set of indices $\{1, 2, \ldots, d\}$. This property is called extremal independence. The latter means, in fact ,the fulfillment of the max-stable property for the maximal component of X. As applications, distributions of various extremal characteristics of binomial random hypergraphs, such as a maximum codegree and a maximum number of cliques sharing a given vertex, are obtained in [11].

5. Network Evolution: Attachment Tools

5.1. Preferential Attachment

A PA, introduced by [71], is the most popular model of network evolution. It reflects a natural attachment of newly appended nodes likely to the most influential ones, i.e., to those with the largest node degrees. Usually, the attachment starts from a single node connected

with itself (or a seed network) to evolve a network. The evolving network without node and edge deletion can be represented as a sequence of graphs $G(n) = (V(n), E(n))$. Typically, the PA leads to the sudden appearance of a giant connected component at a certain critical point that is reflected in a heavy-tailed distribution of vertex degrees [30,72,73]. Using a linear PA, a newly appending node can be attached randomly to an existing node i with probability

$$P_{PA}(i) = d_i / \sum_{s=1}^{||V(n)||} d_s$$

proportional to the degree d_i (or a number of nearest neighbors) of node i. Further generalization of the linear PA is given in [74], where a non-linear PA with the attachment probability proportional to the function of the node degree $P_{PA}(i) \propto f(d_i)$ is proposed. $A \propto B$ means that A is directly proportional to B, i.e. there exists some constant k such that $A = kB$. Particularly, the power function $f(d_i) = d_i^\alpha$, $\alpha > 0$ is studied to generate graphs. Two models of the PA probability for non-directed graphs that depend on the preferential attachment function $f(j)$, $j \geq 1$ are proposed in [60]. Namely, a new node v_{n+1} may connect to one of the existing nodes $v_i \in V(n)$ with probability

$$\frac{f(D_i(n)) + \delta}{\sum_{i=1}^{n}(f(D_i(n)) + \delta)} \qquad \text{(Model A)},$$

where $f(j)$ is assumed to be deterministic and non-decreasing, $D_i(n)$ is the degree of v_i in the existing graph $G(n)$, or

$$\frac{f(D_i(n)) + \delta}{\sum_{i=1}^{n}(f(D_i(n)) + \delta) + f(1) + \delta} \qquad \text{(Model B)},$$

where $\delta > -f(1)$ is a parameter. A fruitful discussion of Model A can be found in [60]. In particular, three cases of the preferential attachment function f are distinguished. If $f(j) = j$, $j = 1, 2, \ldots$, then Model A is called the linear preferential attachment model. Degree frequencies have a power-law distribution in this case. The function $f(j) = j^\beta$, $\beta > 1$ corresponds to a so-called super-linear case. We refer to [75] for a comprehensive study of this model. Finally, by taking $f(j) = j^\beta$, $0 < \beta < 1$, we consider a so-called sub-linear case. Then the degree distribution is much light-tailed distributed in comparison with the linear case.

The PA provides the "rich-get-richer" mechanism since earlier appended nodes may increase the number of their edges further. The PA exhibits the scale-free and the small-world properties [76]. Spatial versions exhibit geometric clustering [77,78]. The scale-free property means that the degree distribution of such a network follows a power law for large node degrees. To model large-scale networks, a class of random graphs is defined in [79], which may show the small-world characteristics of "localized" clustering with a longer range of connectivity. Such a graph may be considered as the superposition of many subgraphs, each of whose subgraph contains only edges of a certain range k, which are presented with a power-law probability.

The small-world property [80] implies that the graph distance between different network nodes is typically short. The small-world phenomenon or "six degrees of separation" (i.e., every pair of individuals in the world is separated by only six links on average) is a fundamental issue in social networks where there is an abundance of short paths in a graph whose nodes are people with links joining pairs of people who know one another [80]. The small-world networks have two properties: small average path length and a high clustering coefficient [5]. An ER network displays a very small average path length as small-world networks, but fails in reproducing high clustering coefficients [5]. The Watts–Strogatz model is used in the small-world random network [81]. In this model, if n is equal to the number of nodes, each node is connected to the number of $k/2$ closest clockwise neighbors and also to $k/2$ closest counterclockwise neighbors. This produces

a ring that is full of triangles for $k > 2$ and, thus, has a high clustering coefficient. The geometric clustering affirms the property that a geometric vicinity of vertices results in a higher probability of establishing an edge between them.

Nodes i and j may be connected with probability $d_i d_j / \sum_{s=1}^{||V(n)||} d_s$, [73]. A kind of PA with a Poisson random number of new edges to the new vertex is proposed in [73]. The procedure works as follows. First, considering a sequence of graphs G_n with possible multiple edges and self-loops, one calculates the mean degree Λ_i over G_n of each node i which is called capacity, $i = 1, \dots, ||V(n)||$. Then, each pair of nodes is connected, independently of the other pairs, with $E(i, j)$ edges, where $E(i, j)$ has Poisson distribution with parameter $\Lambda_i \Lambda_j / \sum_{k=1}^{||V(n)||} \Lambda_k$.

The Spatial Preferential Attachment (SPA) model is introduced in [78]. This model combines the geometry and preferential attachment by introducing "spheres of influence," which volume grows with the degree of a vertex. Then, the vertex degree distribution was shown to follow a power law.

The linear $\alpha-$, $\beta-$, and $\gamma-$ PA schemes for directed graphs are proposed in [2] and developed in [29,30]. The dynamics of this model depend not only on non-negative parameters α, β, and γ, but on parameters $\delta_{in} > 0$ and $\delta_{out} > 0$ as well. A single directed edge is added to a directed graph at each step of evolution with an appended new node. A finite directed graph $G(n_0)$ is used as a seed network. It consists of at least one node v_0 and n_0 edges. A new node v is appended to the existing graph $G(n - 1), n > n_0$, by adding a single edge to $G(n - 1)$. Thus, after n steps, $G(n)$ will be a graph with $n + n_0$ edges and a number of nodes $N(n) \geq n + 1$. The edge creation is provided by flipping a three-sided coin with probabilities α, β, and γ such that $\alpha + \beta + \gamma = 1$. An i.i.d. sequence of trinomial r.v.s with values 1, 2, and 3 and the corresponding probabilities $\alpha, \beta, \gamma \geq 0$ are generated to select one of the following schemes. Furthermore, $I_n(v)$ and $O_n(v)$ denote the in- and out-degree of v.

1 ($\alpha-$scheme) The edge $v \to w \equiv (v, w) \in E(n)$ directed from the new node $v \in V(n)$ to an existing node $w \in V(n - 1)$ is created with probability α, while the existing node $w \in V(n - 1)$ is chosen with probability

$$\frac{I_{n-1}(w) + \delta_{in}}{n - 1 + \delta_{in} N(n - 1)}. \tag{15}$$

2 ($\beta-$scheme) An edge (v, w) is added to $E(n - 1)$ with probability β and the existing nodes $v, w \in V(n - 1) = V(n)$ are chosen independently from $G(n - 1)$ with probability

$$\left(\frac{I_{n-1}(w) + \delta_{in}}{n - 1 + \delta_{in} N(n - 1)} \right) \cdot \left(\frac{O_{n-1}(v) + \delta_{out}}{n - 1 + \delta_{out} N(n - 1)} \right). \tag{16}$$

3 ($\gamma-$scheme) An edge (w, v) from the existing node $w \in V(n - 1)$ to the new node v is created with probability γ. The existing node w is chosen with probability

$$\frac{O_{n-1}(w) + \delta_{out}}{n - 1 + \delta_{out} N(n - 1)}. \tag{17}$$

It is remarkable that the latter schemes allow us to construct multiple edges and loops in the graph.

5.2. Tail Indices of Node Influence Characteristics for Preferential Attachment

The tail indices of the in- and out-degrees of graphs evolved by the linear $\alpha-$, $\beta-$, and $\gamma-$ PA schemes may be calculated using formula (2.9) in [30] (see also [29]):

$$\alpha_{in} = \frac{1 + \delta_{in}(\alpha + \gamma)}{\alpha + \beta}, \quad \alpha_{out} = \frac{1 + \delta_{out}(\alpha + \gamma)}{\beta + \gamma},$$

where marginals of the in- and out-degrees satisfy a power-law behavior

$$p_i^{in} \sim C_{in}i^{-(1+\alpha_{in})} \quad \text{as} \quad i \to \infty,$$
$$p_j^{out} \sim C_{out}j^{-(1+\alpha_{out})} \quad \text{as} \quad j \to \infty$$

for some positive constants C_{in} and C_{out}. The parameters δ_{in} and δ_{out} of the PA method (see, Section 5.1) can be estimated with the semi-parametric extreme value method (EV) based on the maximum-likelihood method; see [30].

In [82], a generalized class of random graphs with common dynamics is considered. At every time step, a new vertex appears in the graph, and it connects to $m \geq 1$ existing vertices with a probability proportional to a function of the degrees of the nodes $f(d_i(n-1))$, where $d_i(n-1)$ denotes the degree of vertex i in the graph PA_{n-1} obtained using the PA with $n-1$ vertices, and $f(\cdot)$ is some PA function. Regarding the original Barabási–Albert model, which is the most popular random graph model for real-life networks, $f(k) = k$ holds. The function $f(k) = k + \delta$, for some constant $\delta > -m$, can also be found in applications. In this case, the power-law exponent of the degree distribution is given by $\tau = 3 + \delta/m$, [7]. In general, every vertex may join the existing network with a random number of edges that connects it to the existing vertices preferentially to their degrees. In [83], it is shown that such a system also exhibits the power-law degree distribution.

The formulae of the tail index of PageRank and the Max-linear model are not yet obtained to the best of our knowledge. In [19], tail indices of the PageRanks and the Max-linear models of superstar nodes in graphs evolved using the linear PA (15)–(17) are investigated with an empirical study. A superstar node within the community is assumed to have incoming links from all nodes of the community. Such nodes may be artificial, and they may not exist in the network. The sums and maxima of PageRanks over communities may serve as PageRanks and the Max-linear models of superstar nodes, respectively. It is shown that the tail indices of PageRanks and the Max-linear models of superstar nodes are close, and they may be approximated by the minimum tail index of PageRanks among representative series containing nodes taken within the communities as their representatives. The novelty of study [19] is that the evolution without node and edge deletion, with the uniform node deletion, and the uniform edge deletion has been studied. Namely, one node or one edge can be deleted at each evolution step when a new node is appended. The number of nodes in the graph is preserved if a node is deleted. Estimates in [84] suggest that the quantity of nodes in the Web graph is growing by a few percent a month. Thus, the mixed deletion policy can be plausible. It is found in [19] that the distribution tails of the PageRanks and Max-linear models of superstar nodes become heavier during the PA evolution without node and edge deletion. This means that the superstars become "richer" and rarer. In contrast, the distribution tails of the PageRanks and Max-linear models of superstar nodes become lighter by the evolution with a uniform node or edge deletion.

In a traditional directed PA, every new edge is added sequentially into the network. However, for real datasets, it can be realistic that several new edges are created at the same timestamp. Previous analyses on the evolution of social networks reveal that after reaching a stable phase, the growth of edge counts in a network follows a nonhomogeneous Poisson process with a constant rate across the day but varying rates from day to day. Taking this result into account, a new modification of the PA model with Poisson edge growth is proposed in [85], and its asymptotic behavior is studied.

5.3. Clustering Attachment

A network may grow through attachment of nodes proportionally to the clustering coefficients, or local densities of triangles of existing nodes. The clustering coefficient is proposed in [13] where the clustering is determined as "the propensity for two neighbors

of the same vertex also to be neighbors of one another, forming a triangle of connections in the network." The local clustering coefficient of vertices with degree k equals to

$$c(k) = \frac{1}{N_k}\frac{2\Delta_k}{k(k-1)} \tag{18}$$

for all k with $N_k \geq 1$, where N_k denotes the number of vertices of degree k, and Δ_k is the number of triangles attached to vertices of degree k [35]. The local clustering coefficient $c(k)$ is not defined if $N_k = 0, k \in \{0, 1\}$. The asymptotic number of triangles for specific graphs was obtained in [35,86]. The asymptotic behavior of $c(k)$ for uniform graphs is derived in [35] and for the SPA model in [87], see also Section 5.5.

Usually, one can observe only node connections, but the presence of a network geometry is not always evident. High clustering and triangle counts are not enough to indicate the presence of geometry in the network since closely located nodes (like neighbors living in the same district or working in the same office) may intensively communicate and build triangles in the same way as nodes with high degrees, Ref. [86]. The extremal index cannot reflect the network geometry by indicating a concentration measure of giant nodes as well.

Based on the clustering coefficient of an individual node, a clustering attachment (CA) for undirected graphs is proposed in [88]. The attachment to an existing node i is proceeded with a probability proportional to its clustering coefficient

$$P_{CA}(i, t) \propto c_{i,t}^{\alpha} + \epsilon, \tag{19}$$

where

$$c_{i,t} = \begin{cases} 0, & k_{i,t} = 0 \text{ or } k_{i,t} = 1, \\ 2\Delta_{i,t}/(k_{i,t}(k_{i,t}-1)), & k_{i,t} \geq 2, \end{cases}$$

where $c_{i,t}$ is the clustering coefficient, $k_{i,t}$ is the degree of node i, $\Delta_{i,t}$ is the number of triangles involving node i at a time t, ϵ is a constant probability for attachment, which may be zero, and $\alpha \geq 0$ is a parameter of the model. $x \propto y$ means there is a non-zero constant C such that $x = C \cdot y$. Non-negative values of α are considered in [88].

For brevity, we omit further t in notations. Since $k_i(k_i - 1)/2$ is the maximum number of triangles that may exist for node i, $0 \leq c_i \leq 1$ holds. In fact, c_i and hence, $P_{CA}(i)$ are r.v.s. Regarding the PA tool, the ith node degree may only increase, i.e., $k_i = k_i + 1$ upon attaching to node i. For the CA one of two cases $(k_i \to k_i + 1, \Delta_i \to \Delta_i)$ or $(k_i \to k_i + 1, \Delta_i \to \Delta_i + 1)$ is possible. With (19), one can see that even when a new triangle appears, the clustering coefficient increases only if $k_i > (1 + a)^2/(1 - a^2)$, where $a = \Delta_i/(\Delta_i + 1)$. In contrast to PA, the CA does not feature a power-law distributed k_i since attaching to a node i most likely drives down the i's probability for further attachment.

Another novelty introduced in [88] is that each new node may attach not only to one of the existing nodes for the linear PA in [30] but to a fixed number $m_0 \geq 2$ of existing nodes. This approach excludes the appearance of multiple edges. An existing node is (uniformly) removed at each step of evolution when a new node is added to preserve a constant number of nodes in the network in [88]. Removal of a node or an edge at each evolution step impacts both Δ_i and k_i and influences the attachment probability to node i.

The CA causes clusters of consecutive exceedances of the evolving modularity over a sufficiently high threshold [88] The modularity is a measure that allows us to divide graphs into communities. It shows the connectivity of nodes in the community. More precisely, regarding an undirected graph $G = (V, E)$, the modularity Q of a partition C shows how many edges exist within communities and between them [89]:

$$Q = \frac{1}{2s}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2s}\right]\mathbf{1}(i, j) \tag{20}$$

Here, A_{ij} is an element of an adjacency matrix ($A_{ij} = 0$ if the edge between vertices i and j does not exist and $A_{ij} = 1$ otherwise), $s = \frac{1}{2}\sum_{ij}A_{ij} = \|E\|/2$ is a half of the edge number

in G, $\mathbf{1}(i,j)$ is equal to 1 when nodes i and j belong to the same community, $k_i = \sum_j A_{ij}$ is the degree of node i, Refs. [90,91]. In fact, s increases linearly as $m_0 \geq 2$ or it is constant as $m_0 = 1$ if one edge is removed at each evolution step, and m may remain constant, or it decreases if one node is removed together with its edges. This evolution impacts modularity.

The community detection using modularity maximization using the definition of (20) is unable to find the community structure in networks with many small communities [92]. To overcome the problem, a generalized modularity function is proposed in [93] that can be written in the form

$$Q = \frac{1}{2s} \sum_{ij} \left[A_{ij} - \gamma \frac{k_i k_j}{2s} \right] \mathbf{1}(i,j), \tag{21}$$

where γ is the resolution parameter. A similar generalization was proposed previously on different grounds in [94]. What γ one should use is discussed in [92].

Regarding directed graphs the modularity can be easily formulated by

$$Q_d = \frac{1}{s} \sum_{ij} \left[A_{ij} - \frac{k_i^{in} k_j^{out}}{s} \right] \mathbf{1}(i,j), \tag{22}$$

where k_i^{in} and k_j^{out} denote the in-degree of node i and the out-degree of node j, respectively [91].

Considering the change of the modularity at each evolution step, the extremal index of the modularity sequence indicates the consecutive large connectivity of nodes within the evolving communities, and thus, it reflects the evolution of the communities, [95]. Fluctuations of the modularity over time for $\alpha \neq 0$ in (19) are far larger than for $\alpha = 0$; see Figure 1(b) in [88]. The modularity can become very large due to the sparsity in the network.

Denoting the graph at evolution step t as $G(t) = (V(t), E(t))$, the probabilistic measure

$$P_{CA}(i,t) = \frac{c_i^{\alpha}(t) + \epsilon}{\sum_{j \in V(t)} c_j^{\alpha}(t) + \|V(t)\|\epsilon} \tag{23}$$

is used instead of (19). The case $\epsilon = 0$ is specific. If, in addition, node i does not belong to any triangle of nodes and $c_i = 0$ holds not for all $i \in V(t)$, then $P_{CA}(i,t) = 0$ follows by (23). It implies that new nodes cannot be attached to node i.

The simulation in Figure 2 shows that a cluster structure of the time series $\psi(t) = Q(t)/\langle Q \rangle - 1$, where $\langle Q \rangle$ denotes the average over evolution steps over a time interval, against evolution steps, t is very sensitive to node removals. The creation of a new triangle leads to increasing modularity and, thus, to the appearance of its clusters of exceedances. The CA evolution without node and edge deletion causes a kind of a barbell graph (see, Figure 1 in [95]), i.e., a path of well-conducted cliques that are weakly connected with each other due to bottlenecks. A uniform node deletion at each step of the evolution leads to a large number of isolated nodes since the number of edges decreases due to node removal. The modularity is nearly constant for $\alpha = \epsilon = 0$ due to a constant attachment probability $P_{CA}(i,t)$ since per (23) it holds

$$P_{CA}(i,t) = \frac{\mathbf{1}\{c_{i,t} > 0\} + \epsilon}{\sum_{j \in V(t)} \mathbf{1}\{c_{j,t} > 0\} + \|V(t)\|\epsilon}, \quad i \in V(t), \ \alpha = 0. \tag{24}$$

The extremal index of the evolving modularity and the tail index of node degrees were estimated for a variety of parameters α and $\epsilon \in \{0,1\}$ in (23) [95]. The smaller the extremal index is, the larger the clustering (or local dependence) of a modularity sequence. Applying different parameters of the CA both to the cases without the node and edge deletion and with uniform node deletion, it is found that $\epsilon = 1$ in (23) leads to stable large values of the extremal and tail indices. This feature happens since the clustering coefficient does not

impact the CA due to a dominating ϵ. It means a weak clustering of the modularity and a light-tailed node degree distribution. Without the node and edge deletion, $\epsilon = 0$ causes decreasing extremal indices that are close to zero and increasing of the tail indices of node degree as $\alpha > 0$ increases [95]. With uniform node deletion, the extremal indices increase, and node degree tail indices increase slower than for the case without edge and node removal. It means that the node deletion leads to a weaker clustering of the modularity and a heavier tail of node degrees than without node and edge deletion.
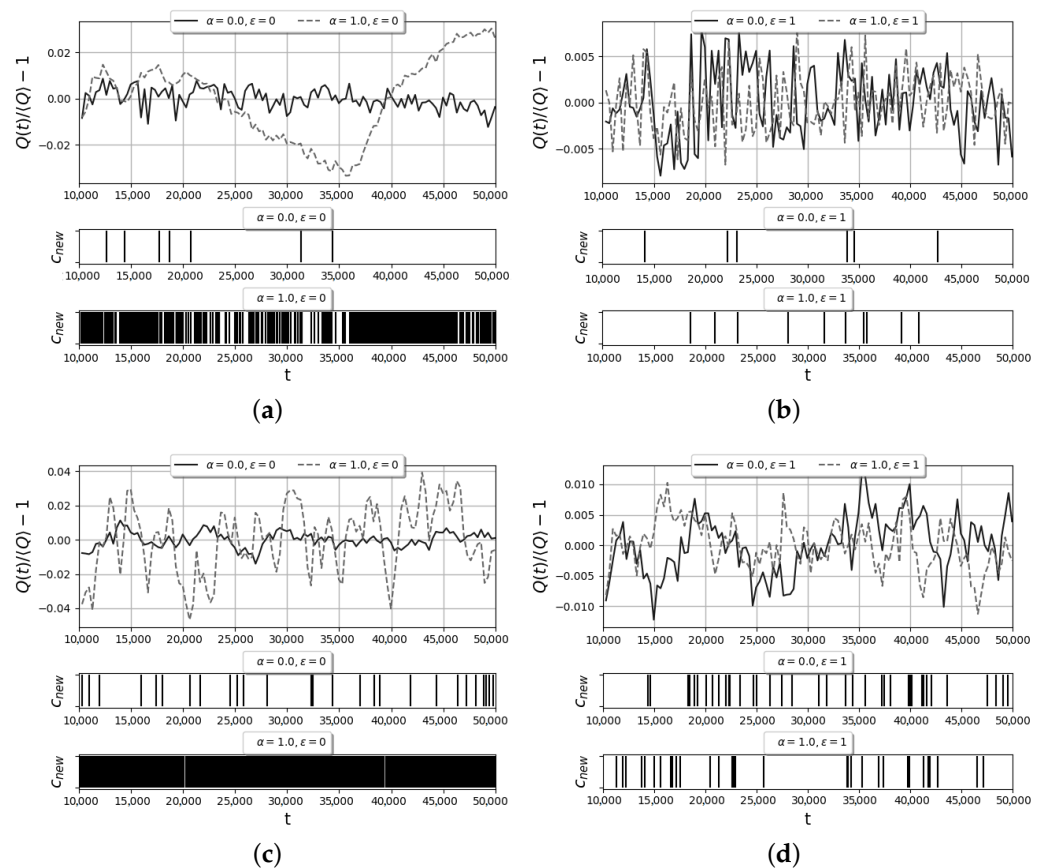

Figure 2. The evolution of the normalized graph modularity $\psi(t) = Q(t)/\langle Q \rangle - 1$, where $\langle Q \rangle$ denotes the average over evolution steps over the interval $t \in [10^4, 5 \cdot 10^4]$, against the CA evolution steps; and spike trains denoting injections of new triangles when new nodes are appended and their clustering coefficient c_{new} is equal to one: without node and edge deletion (**a,b**) and with uniform node deletion (**c,d**) for $m_0 = 2$ (reprinted from Figure 2 in [95]).

5.4. Other Models of Random Networks

Some other models to generate random graphs that may fit observed networks are represented in [4,96,97].

In [96], the model is characterized by four stochastic discrete-time processes—the creation processes for node- and edge-creation, and the deletion processes for node- and edge-deletion. It is stated that random copying of edges is a simple stochastic mechanism for creating a Zipfian degree distribution. More exactly, one adds links to a node v by picking a random (other) node u in the graph, and copying some links from u to v. A node is created or deleted independently at each step with some probabilities. With probability β one adds k edges from node v to nodes chosen independently and uniformly at random. With probability $1 - \beta$, one copies k edges from a randomly chosen node to v. The value k is the parameter to be selected. If the out-degree of u is less than k, its edges are first copied to v, and then by picking another random node u' one can copy the required number of edges from u' to u until the number of edges of u will not be enough. The processes creating the

graphs differ from traditional graph models. However, the process of copying generates complicated dependencies that make analysis very complex.

In [97] such models like a deterministic block model, the configuration random graph model, d-regular random graph, a geometric random graph model apart from the ER, and the SBM described in Section 4.3 are presented as a list of most common models. In [4], much attention is devoted to spatial networks, e.g., the Spatially Embedded Random Networks and the Waxman model.

Spectral density (i.e., the distribution of the graph adjacency matrix eigenvalues) and its kernel estimator are used in [97] to find the best graph model for the observed network in such a way that the divergence between the estimated spectral density and the known limiting spectral density (or one obtained using Monte Carlo) will be the smallest. A rich theory of the kernel estimators of probability density functions can be applied to spectral densities in a similar way. As the divergence measure, one can take the ℓ_1 distance as in [97] or the Kullback–Leibler measure as proposed in [98].

5.5. Triangle Counts and Local Clustering Coefficients

The number of triangles of connected nodes has been studied for specific models of random graphs with known node degree distributions and in the context of different problems.

Network geometry is an important feature of real-world networks since it explains their scale-invariance, high clustering, and overlapping community structures. Indeed, triangles may overlap in real networks. Intuitively, node connections can either be formed between high-weight nodes, or between close-by nodes. When the node degree distribution has a heavy tail, the triangles are formed by high weighted vertices. At the same time, in geometric graphs similar vertices (which could represent people living in the same region and having similar social interests) are likely to connect and form triangles [86]. In [86], it is shown that the triangle counts $\{\Delta_i\}$ and an average clustering coefficient

$$\overline{C}_t = \frac{1}{n} \sum_{i \in V(t)} c_{i,t} \tag{25}$$

for $\|V(t)\| = n$ nodes, where the clustering coefficient $c_{i,t}$ of node i at the time t is determined per (19), are not enough to detect the presence of geometry in the graph if the node degree distribution has a sufficiently heavy tail. To show this, the inhomogeneous random graph (IRG) and geometric inhomogeneous random graph (GIRG) were compared as benchmark models. For the GIRG model, each node i was equipped with a weight h_i and a uniformly sampled position x_i on a d-dimensional torus $[0,1]^d$ endowed with the infinity norm. Weights were sampled independently from the Pareto distribution with the density

$$\rho(h) = K_1 h^{-\tau}, \qquad \tau \in (2,3),$$

for any $h > h_0 > 0$, where K_1 is the normalized constant. Two nodes i and j are then connected independently with probability

$$p(h_i, h_j, x_i, x_j) \quad = \quad K_2 \min\left(\frac{h_i h_j}{\mu n} \|x_i - x_j\|^{-d}, 1 \right)^{\gamma}, \tag{26}$$

for some $\gamma > 1$, where μ is the average weight, K_2 is a correction factor, which ensures that the expected degree of any vertex i is proportional to its weight h_i. The IRG model is a non-geometric one and it differs from the GIRG in the sense that the nodes possess only weights h_i, sampled from the Pareto distribution. They are connected with probability

$$p(h_i, h_j) = \min\left(\frac{h_i h_j}{\mu n}, 1 \right).$$

Denoting the number of triangles in a graph G as $\Delta(G)$, the following asymptotic results are derived in [86]

$$\frac{\Delta(G)}{n^{3(3-\tau)/2}} \xrightarrow{\mathbb{P}} A_{GIRG}$$

for the GIRG, when $\tau < 7/3$, and

$$\frac{\Delta(G)}{n^{3(3-\tau)/2}} \xrightarrow{\mathbb{P}} A_{IRG}$$

for the IRG as $n \to \infty$, where A_{GIRG} and A_{IRG} are explicit constants that depend on model parameters. Regarding the GIRG, $\Delta(G)$ scales as n, when $\tau \geq 7/3$ holds. Since the model parameters, and hence, the constants A_{GIRG} and A_{IRG}, cannot be evaluated in practice, one cannot detect the geometry by distinguishing the constants. Similar conclusions concern the average clustering coefficient (25).

If the value of $\overline{C}$ does not vanish in n, then this indicates a geometry in the graph [86]. The average clustering in the GIRG is $\Omega(1)$, i.e., it does not vanish as n increases [99]. The asymptotic decay of $\overline{C}$ for non-geometric graphs depends on the tail index ι in (1), and it may be very slow for some values of ι [86]. For example, $\overline{C}$ decays asymptotically as $n^{\iota-2} \ln n$ for the IRG [7]. Then it is difficult to distinguish between the geometry and non-geometry of the graph in practice.

An important notice is that for $d = 1$, $\gamma = \infty$ (γ is the parameter in (26)) the GIRG is asymptotically equivalent to the hyperbolic random graph. Thus, the asymptotic results on the GIRG stay valid for the hyperbolic random graph.

As an alternative to the triangle counts and clustering coefficients, Ref. [86] proposes weighing triangles so that triangles with low evidence for geometry have a low weight. The statistic

$$W = \sum_{i,j,k \in V, i<j<k} \frac{1}{d_i d_j d_k} \mathbf{1}\{(i,j,k) = \Delta\},$$

where $\mathbf{1}\{(i,j,k) = \Delta\}$ is the indicator function of the event that the vertices i,j,k form a triangle, is implemented in [86]. It is significantly greater for GIRGs ($W = O(n)$) than for IRGs ($W = O(1)$), and hence, it may serve to detect the network geometry.

Another natural model for real-world networks is given using the uniform random graph (URG). Given a positive integer n and a graphical degree sequence, a sequence of n positive integers $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$, where $\sum_{i=1}^{n} d_i \equiv 0 \pmod{2}$, the uniform random graph is a simple graph, uniformly sampled from the set of all simple graphs with degree sequence $(d_i)_{i \in [n]}$, Refs. [35,36]. It is assumed that $\boldsymbol{d}$ is a realizable degree sequence, meaning that there exists a simple graph with degree sequence $\boldsymbol{d}$. Let $G(\boldsymbol{d})$ denote the ensemble of all simple graphs on degree sequence $\boldsymbol{d}$, and let $d_{\max} = \max_{i \in [n]} d_i$, $[n] = \{1, 2, \ldots, n\}$. In [35], $(d_i)_n$ follows a power-law distribution with exponent $\tau \in (2, 3)$. If, in addition, the empirical degree distribution $F_n(j) = 1/n \sum_{i \in [n]} \mathbf{1}\{d_i \leq j\}$ satisfies Assumption 1: (i) There exists a constant $K > 0$ such that for every $n \geq 1$ and every $0 \leq j \leq d_{\max}$, $1 - F_n(j) \leq K j^{1-\tau}$; (ii) There exists a constant $C > 0$ such that for all $j \in O(\sqrt{n})$, $1 - F_n(j) = C j^{1-\tau}(1 + o(1))$ holds [35], then the number of triangles in the URG follows

$$\frac{\Delta(G)}{n^{3(3-\tau)/2}} \xrightarrow{\mathbb{P}} -\frac{1}{12} \left(\frac{\pi C (\tau-1) \mu^{-(\tau-1)/2}}{\cos\left(\frac{\pi\tau}{2}\right)} \right)^3,$$

where $\mu = \mathbb{E}[\boldsymbol{d}]$, and C is a constant. The result is similar to the number of triangles in the erased configuration model, where all multiple edges of the configuration model are merged, and all self-loops are removed

$$\frac{\Delta(G)}{n^{3(3-\tau)/2}} \xrightarrow{\mathbb{P}} -\frac{1}{12} \left(C(\tau-1) \mu^{-(\tau-1)/2} \Gamma\left(\frac{1}{2} - \frac{\tau}{2}\right) \right)^3.$$

The local clustering coefficient $c(k)$ averaging over the vertices of degree k was intensively studied both theoretically and empirically. In real-world networks $c(k) \sim k^{-\phi}$ for some $\phi > 0$ is observed. The asymptotic behavior of $c(k)$ in (18) is derived for the uniform random graph in Theorem 2 in [35]. For small values of k, $c(k)$ is independent of k. Then a range of slow decay in k follows. When $k \gg \sqrt{n}$, $c(k)$ decays as a power of k.

In [100], it is shown that $c(k)$ can be well approximated using k^{-1} for four large networks. In [34], $\phi = 0.75$, and in [101], $\phi = 0.33$ were obtained. In [87], it was proven for the SPA model that both $c(k)$ and a local individual clustering coefficient c_i of a vertex i (see, (19)) with a large enough degree k behave as k^{-1}.

The convergence with rate $O\left(n^{-1/2}\right)$ in total variation between the distribution of triangle counts and a Poisson distribution in generalized random graphs (GRG) is proved using the Stein–Chen method in [67]. Let $W_i > 0$ be the weight of node i. The probability of the edge between any two nodes i and j, for $i \neq j$, is equal to

$$p_{ij} = \frac{W_i W_j}{W_i W_j + \sum_{i=1}^{n} W_i}.$$

Self-loops are prohibited. The node weights $\{W_i\}_n$ can be taken to be either deterministic or random. The ER random graph with $p_{ij} = \lambda/n$ is a special case of the GRG, if $W_i \equiv n\lambda/(n-\lambda)$ for some $0 \leq \lambda < n$. Let $\mathcal{L}(Y)$ be a distribution law of Y. For any integer-valued non-negative r.v.s Y and Z, we denote the total variation distance between their distributions $\mathcal{L}(Y)$ and $\mathcal{L}(Z)$ by

$$\|\mathcal{L}(Y) - \mathcal{L}(Z)\| \quad \equiv \quad \sup_{\|h\|=1} |Eh(Y) - Eh(Z)|,$$

where h is any real function defined on $\{0, 1, 2, \dots\}$ and $\|h\| \equiv \sup_{m \geq 0} |h(m)|$. Let W_i, $i = 1, 2, \dots, n$ be i.i.d. r.v.s distributed as a r.v. W. Let $S_n(k)$ be the number of cycles of length $k \geq 3$ (cycles are triangles for $k = 3$). The main result per [67] (Theorem 1) states that for $k \geq 3$, one has

$$\|\mathcal{L}(S_n(k)) - \mathcal{L}(Z_k)\| \quad = \quad O\left(n^{-1/2}\right), \tag{27}$$

provided that

$$\mathbb{P}(W > x) = o\left(x^{-2k-1}\right), \text{ as } x \to +\infty.$$

Z_k is a r.v. having a Poisson distribution with parameter $\lambda(k) = \left(\mathbb{E}W^2/\mathbb{E}W\right)^k/(2k)$. Relation (27) is valid, if W has a power-law distribution or the moment EW^{2k+1} is finite.

6. Community Detection Methods and Related Topics

6.1. Partition into Communities

Graph clustering attracts interest from many authors with a different understanding of what the cluster in the graph means [4]. The cluster is determined to be a smaller, denser ER graph planted within an ER graph in [102]. Generally, the community detection, or the graph clustering, consists of partitioning the vertices of a graph into clusters that are more densely connected [45]. Communities may more generally refer to groups of vertices that behave similarly [45]. Numerous existing methods for community detection relate mostly to conductance [22,23] or modularity measures, [90–92,94] as well as to information theory [103], and Bayesian generative models [104,105].

A Greedy Modularity Maximization Algorithm (GMMA) [90] is used to detect the community structure quickly. To maximize the modularity, Loivan's algorithm is applied to non-directed graphs [106]. The directed Loivan's algorithm with adapted modularity (20) is used for directed graphs [91,107].

The monitoring of the formation and evolution of communities in large social networks like Twitter is an important problem. One can study evolving subgraphs corresponding to a topical community. The conductance (see, Appendix D) quantifies how well or poorly connected a subgraph is to the rest of the graph relative to its internal connections. The paper [108] addresses tracking the conductance in real-time since the number of communities that are active at any time is very large, and the rate at which the communities evolve is very high. The minimum conductance might not be the best criterion for local graph clustering [102]. The mean field analysis is based on an aggregation of Web pages into classes according to pairs $k = (k_{in}, k_{out})$ of their in- and out-degrees and using averages of PageRanks within each k-degree class to calculate the PageRank [109].

Most of the proposed methods to discover a community structure in networks are unsuitable for very large networks because of their computational cost. The GMMA is a hierarchical agglomeration algorithm for detecting communities, which is faster than many competing algorithms: its running time on a network with n vertices and m edges is $O(md \log(n))$, where d is a depth of the "dendrogram" describing the community structure [90,110]. Since many real-world networks are sparse and hierarchical, with $m \sim n$ and $d \sim \log n$, the GMMA runs in essentially linear time, $O(n \log^2 n)$.

Some methods of community detection, including ones based on modularity, return multiple plausible partitions rather than just one. Many of them are similar to one another, differing only by a few nodes. In [111], it is proposed to cluster similar partitions into a small number of groups and then identify an archetypal partition as a representative of each group. This method allows the reduction of a number of partitions into communities.

Some cut-based methods and their relaxation in the form of spectral clustering for separating the graph into $K \geq 2$ groups—such that inside the group, the edge density is higher than between two different groups—are considered in [4]. The Laplacian is used instead of the adjacency matrix for spectral clustering. It is noted that the computation of eigenvectors required for the spectral analysis has complexity $O(n^3)$ (n is the number of nodes in the graph). In practice, when dealing with a sparse matrix whose eigenvalues are well separated, the complexity can be close to $O(Kn)$, where K is the number of eigenvectors needed.

6.2. Related Topics

6.2.1. Coloring Random Graphs

A vertex coloring of the graph relates to the community detection color problem. The related detection algorithms are based on the theory of random graph coloring and stochastic decomposition methods of the symmetric adjacency matrix derived from the underlying undirected graph; see [44–46,112] among others. Let us focus on directed graphs with possibly non-symmetric adjacency matrices. A survey of polynomial time algorithms that optimally color random k-colorable graphs (including the sparse ones) for a fixed $k \geq 3$ with high probability is given in [46]. The sparsity is governed by a parameter p that specifies the edge probability. The latter probability may be determined using the PA $\alpha-$, $\beta-$, $\gamma-$ schemes recalled in Section 5.1. In [46] a polynomial time algorithm that works for sparse random 3-colorable graphs for a specific p is proposed.

The determining or estimating of the chromatic number $\chi(G)$ of a graph G, which is the minimum number of colors in a proper vertex coloring, is also connected to the selection of the number of communities in the random graph. The probable value of $\chi(G)$ is roughly comparable with the average degree of the random graph, and it may be, therefore, much higher than a fixed number of expected colors [44].

6.2.2. Anomaly Detection Using Machine Learning

Community detection and clustering are central problems in machine learning [45]. A comprehensive review of an extreme learning machine is given in [113]. In [114,115], the machine learning algorithms based on extreme value theory are proposed to detect an abnormal class of anomaly observations. Since a training set may not include all possible

classes, one may distinguish between known normal data and unknown abnormal test data. Considering a community to be a class may be used to test a new community not observable before.

Let $x_i \in R^p$ be the training data that are labeled as a class $y_i \in \{C_1, \ldots, C_J\}$, $i = 1, \ldots, n$. $J \in N$ is the number of different classes in the training set. $p \in N$ is the dimension of the predictor space. The extreme value machine (EVM) introduced in [115] is based on the concept of margin distance of a training point x_i as half of the minimum distance between x_i and all the points belonging to a different class in the training data set:

$$M^{(i)} = \min_{j:y_j \neq y_i} D_j^{(i)} = \min_{j:y_j \neq y_i} \frac{\|x_i - x_j\|}{2}.$$

A new point is classified as normal if it is inside the marginal distribution of some point in the training set with high probability.

Using the equivalent representation $\bar{M}^{(i)} = \max_{j:y_j \neq y_i}(-D_j^{(i)})$, it is proposed to use the Fisher–Tippet–Gnedenko theorem to fit to the k largest observed $-D_j^{(i)}$ for each point x_i a Generalized Extreme Value (GEV) distribution

$$W^{(i)}(z) = \begin{cases} \exp\{-\left(-\frac{z}{\sigma_i}\right)^{\alpha_i}\}, & \text{if } z \leq 0, \\ 1, & \text{if } z \geq 0, \end{cases}$$

assuming a zero upper endpoint. $b^* = \sup\{x \in R : F(x) < 1\}$ denotes the right endpoint of the distribution $F(x)$. $b^* = 0$ since $\bar{M}^{(i)}$ is a negated distance. Here, $\alpha_i \in R$ and $\sigma_i \geq 0$ are the shape and scale parameters, respectively. The latter are to be estimated to obtain $\hat{W}^{(i)}(z)$. A new point x_0 is assigned as normal, if

$$\hat{W}^{(i)}(-\|x_0 - x_i\|) \geq \delta,$$

and as abnormal otherwise, where a threshold δ is chosen by a heuristic formula.

The drawbacks of the EVM are that it strongly relies on the distances between different classes in the training set, the endpoint is assumed to be zero, σ is not rigorously selected, and the EVM gives a non-justified premium to normal classes far from the others. In contrast, in the generalized Pareto distribution classifier (GPDC) proposed in [114], the training data are assumed to be sampled from only one class, the class of the normal data points. This is a significant simplification. The upper endpoint is taken to be $b^* < 0$. The Euclidean distances are used for simplicity. The evaluation of a new point x_0 requires computing $O(k \log n)$ distances, where k is the biggest number of negated distances $-D_{(n)}, \ldots, -D_{(n+1-k)}$, as far as the EVM needs $O(n)$ distances.

To apply the latter algorithms to irregular community detection, one can calculate the length of the shortest path between two nodes in links as the distance. To this end, Dijkstra's algorithm may be applied [116].

6.2.3. Classification of Newly Appended Nodes in Evolving Graphs

An algorithm to classify newly appended nodes in evolving graphs is proposed in Algorithm 1 in [16]. It is based on results from [15], see also Section 3.3.2.

To this end, an initial directed graph with $n > 1$ nodes used as a seed network is partitioned into communities. PageRanks of all nodes of the initial graph are calculated. The tail and extremal indices of each community are estimated by non-parametric estimators, e.g., the tail index using the Hill estimator [50] and the extremal index with an intervals estimator [20], see Appendices A and B. The communities are ranked in ascending order of their tail indices. A new node is assigned to Class 1 if it has a link to one of the most heavy-tailed ("dominating") communities with the minimum tail index k_1. If the new node has a link to communities with the next minimum tail index k_i, $i \geq 2$, and it is not linked with communities with smaller tail indices, then it is assigned to Class i.

The minimum tail index k_1 and the extremal index θ_1 of the "dominating" community are assigned to PageRanks and the Max-linear models of the new nodes from Class 1, if the "dominating" community is unique. If there is a random number of "dominating" communities, then the Max-linear models of the new nodes from Class 1 have the extremal index of the "dominating" community with the maximum PageRank. If arbitrary enumerated sequences of node PageRanks of the "dominating" communities are independent or weakly dependent, more precisely, they satisfy conditions (A1) or (A2) (see, Section 3.3.2), then the PageRanks of the new nodes from Class 1 have the same extremal index as Max-linear models. The conditions (A1) or (A2) for the "dominating" communities provide the same minimum tail index k_1 for the PageRanks and the Max-linear models of the new nodes from Class 1. Class 2 obtains the second minimum tail index corresponding to the next set of communities in the range and the respective extremal index as in the previous case, and in the same way, classes with numbers $i > 2$ obtain their tail and extremal indices.

Since the tail index is to be estimated, it is plausible to assume that the community with the minimum tail index is unique. However, the tail indices may have close values. One can apply a test statistic (14) to show that the tail indices of the communities are likely different.

7. Leading Nodes for Information Spreading in Evolving Graphs

The influence maximization (IM)—the problem of finding a relatively small optimal set of nodes that have the most influence—is central for many applications, e.g., in social and computer networks for rapid information spreading [22,23] or for studying an epidemic spreading in heterogeneous complex networks [117]. The IM term was first introduced in [118], where a greedy optimization algorithm was proposed. At each stage of the algorithm, the best spreader is chosen sequentially outside the current set of optimal spreaders, which generates the largest increment in the influence of the set of spreaders. The algorithm is computationally costly and can be applied to relatively small networks.

The information spreading, as a a delivery model of messages in the whole network (the full spreading) [23] or in its part (the partial spreading) [22], may be applied to the parallel grid calculations in computation networks. Using a SPREAD algorithm proposed in [23] for undirected stationary graphs and assuming that all nodes have asynchronous or synchronous clocks, a node i that initiates communication at time t is first selected. The node i chooses the next node u randomly among nodes of the network by the global clock tick that works according to a Poisson process. The node u sends all of the messages it has to the node i. In [119], Algorithm 1, the SPREAD algorithm is modified for directed graphs. It is assumed that a single message that has been at the disposal of an initial graph or just a single node has to be spread to a fixed number of the rest nodes. The clocks of the latter nodes are assumed to be asynchronous. The message can be sent from node i to node j if there is a directed edge from i to j. Node j is selected uniformly with a node i among its neighbors without the message. This is proposed to provide with probability $P_{ij} = 1/O_i$, where O_i denotes the out-degree of node i and sends its single message to node j.

Let $G = (V, E)$ be an undirected connected graph of order n and $(X_1, \ldots, X_n)$ be a sample of node characteristics with a marginal cumulative distribution function $F(u)$. A standard graph index used for the leader election is the closeness centrality C_x of node x

$$C_x = \frac{n-1}{\sum_{y, y \neq x} d(x, y)}, \qquad 0 < C_x \leq 1, \tag{28}$$

where d(x, y) is the shortest path $(x, \ldots, y)$ between nodes x and y [120]. If a node is closer to other nodes, then its value C_x is closer to 1. Using the closeness centrality as a measure of a node's leadership, the relation between its extremal index and the minimal spreading time is found using a simulation study in [21]. To determine the extremal index of the community S, assuming that the latter index exists, a high quantile of $F(x)$, the cumulative distribution function of C_x, is taken as the threshold u^*. The community S is assumed

to be a strict-sense stationary set with extremal index θ. The stationarity implies that the distribution of all nodes of S remains the same irrespective of the numeration of the nodes.

In [21], the selection of a leading community in a network and the comparison of the spreading time by the latter and other communities are observed using an example of homogeneous geometric undirected graphs. In geometric graphs, two nodes are connected if and only if the distance between them is less than a given radius r [4]. Such graphs cannot describe real-world networks, where there are many nodes of low degrees and a few nodes of extremely high degrees. It is found in [21] that a community with the same extremal index as the entire geometric graph is a leading one and determines the minimal spreading time of the entire graph.

In [119], the linear PA $\beta-$ and $\gamma-$ schemes proposed in [30] and recalled in Section 5.1 are applied to the information spreading purpose in directed graphs. The SPREAD algorithm is also reconsidered for directed graphs. Both the PA and the SPREAD are applied to non-homogeneous graphs. A message from one node is spreading to a given number of nodes in the network. The PA may be a better spreader than the SPREAD algorithm. This result is valid for the sets of the PA parameters with relatively small α. The latter means dominating proportions of created new edges from existing nodes to newly appending ones or between the existing nodes only. The information spreading is investigated both for simulated and real non-homogeneous directed graphs generated by the PA with different sets of parameters. Such graphs may contain cycles and multiple edges. It is found in [119] that some nodes in the community with the smallest tail index of the out-degrees and PageRanks may spread the message faster than other nodes.

In [121], it is proposed to divide the network into sectors of influence and then to select the most influential nodes within these sectors.

8. Conclusions and Discussion

In our survey, we focus on the problems arising in evolving networks mostly due to the heavy-tailed nature of node indices. Much attention is devoted to one of the authors' results devoted to the tail and extremal indices of PageRanks and the Max-linear models that are used as node influence indices in the evolving networks, as well as to the search for leading nodes and communities that may spread the information faster than other nodes in the network. The latter results are based on theoretical results regarding the tail and extremal indices of sums and maxima of random-length non-stationary stochastic sequences obtained in a series of the authors' papers. Some results are obtained using simulation studies due to their complexity and application to real-world networks.

Related topics like preferential and clustering attachments, community detection, stationarity and dependence of graphs, information spreading, finding the most influential nodes and communities, and concerning methods are surveyed.

There are many unsolved or not sufficiently developed problems yet, like the distribution of triangle and circle counts in evolving networks or the tail index of the PageRank, among many others. The clustering attachment and asymptotic behavior of its characteristics like node degrees and triangle counts, the local dependence of the modularity are not yet well investigated. An application of well-known estimators of the extreme value index (or the tail index) created for random sequences is not rigorously supported for random graphs due to the non-homogeneity and dependence in real-world networks.

A future study may concern reconsideration of the same problems for evolving networks with node and edge deletion. A deeper study of the attachment models may lead not only to power-law distributed node in- and out-degrees and heavy-tailed PageRanks like preferential attachment tools but also to light-tailed distributed ones like the clustering attachment model. More attention may be devoted to new influence measures like the Max-linear model. A new challenge may appear due to the application of machine-learning methods to community detection.

Our future research will concern a deeper study of the CA. We aim to investigate the evolution of random undirected graphs using the CA without node and edge deletion and

with uniform node or edge deletion, and to obtain an average clustering coefficient and an attachment probability for the CA.

Abbreviations

The following abbreviations are used in this manuscript:

CA	Clustering attachment
i.i.d.	independent identically distributed
IM	Influence maximization
IRG	inhomogeneous random graph
ER	Erdös-Rényi graph
EVI	Extreme value index
GEV	generalized extreme value
GIRG	geometric inhomogeneous random graphs
GRG	generalized random graph
KS	Kolmogorov-Smirnov
MDM	minimum distance method
MSE	mean squared error
RMSE	square root of MSE
r.v.	Random variable
PA	Preferential attachment
POT	Peak over Threshold
SBM	Stochastic Block Model
SPA	Spatial Preferential Attachment
URG	Uniform random graph
URL	Uniform Resource Locator
WWW	World Wide Web

Appendix A. Tail Index Estimators

Let us consider a sequence of random graphs. Regarding graph $G(n)$ at time n, where $N(n)$ is the number of nodes in $G(n)$, the in- and out-degrees (13) of nodes are available statistics for estimation. For the sake of brevity, we denote $I_v = I_v(n)$, $O_v = O_v(n)$.

Assume that the right tails of the marginal distributions of the in- and out-degrees are regularly varying with indices $-\alpha_{in}$ and $-\alpha_{out}$, respectively. Despite the in- and out-degrees are integer-valued, one can accept distribution with regularly varying tail as the relevant model for these r.v.s. This model is motivated by several papers; see [39,122,123] among them. Empirical analysis of social network data shows that the degree distributions follow power laws, and theoretically, this is true for linear preferential attachment models (see [2]).

Hill's estimator is a common way to estimate the tail index α_{in} (the estimation of α_{out} similar) [50]. Let $I_{(1)} \geq \cdots \geq I_{(N(n))}$ be the decreasing order statistics of $I_v(n)$,

$v = 1, \ldots, N(n)$. The semi-parametric statistics $M_r(k)$, $r \in \{1,2\}$, introduced in [124], are based on the k largest degrees and have the form

$$M_r(k) = \frac{1}{k} \sum_{j=1}^{k} \left(\ln \left(\frac{I_{(j)}}{I_{(k+1)}} \right) \right)^r.$$

The above mentioned Hill's estimator can be expressed via statistic $M_1(k)$ as follows $\hat{\alpha}_{in}^{(1)}(k) = (M_1(k))^{-1}$. Hill's estimator is known to be consistent and asymptotically normally distributed for i.i.d. samples. The consistency of Hill's estimator for some dependent data is proven in [125]. The ratio estimator is a generalization of Hill's estimator in the sense that an arbitrary threshold level x_n between some order statistics is used instead of an order statistics $I_{(k+1)}$ in $M_1(k)$ [56].

We note that the estimators

$$\hat{\alpha}_{in}^{(2)}(k) = \left(\frac{M_2(k)}{2} \right)^{-1/2}, \quad \hat{\alpha}_{in}^{(3)}(k) = \frac{2M_1(k)}{M_2(k)}$$

can be considered a good alternative to the Hill estimator. The estimator $\hat{\alpha}_{in}^{(2)}(k)$ has been suggested in [126]. $\hat{\alpha}_{in}^{(3)}(k)$ was introduced by de Vries; see [127].

The construction of the estimators $\hat{\alpha}_{in}^{(i)}(k)$, $i \in \{1,2,3\}$ is based on a peaks-over-threshold (POT) approach, i.e., one selects those of the initial observations that exceed the kth largest order statistic used as a certain high threshold. More POT-type estimators can be found in [128]. There it was noted that the slope of the least squares line through the plot

$$\left\{ \left(-\ln \left(\frac{j}{N(n)+1} \right), \ln \left(I_{(j)} \right) \right), \ 1 \leq j \leq k \right\}$$

is a weakly consistent estimator of α_{in} when observations are i.i.d. r.v.s.

Let us assume now that nodes of directed graph $G(n)$ are divided into non-overlapping communities by some partition algorithm. That is, nodes in the graph belong to only one community. Each community is nothing else but a subgraph of $G(n)$. It is natural to take several communities, say s, with the biggest number of nodes and to apply the Hill estimator to each of them. Then the sample mean over the s communities

$$\hat{\alpha}_{in}^{(4)}(s) = \frac{1}{s} \sum_{i=1}^{s} \hat{\alpha}_{in}^{(1,i)} \left(k_i^{(*)} \right) \tag{A1}$$

is the estimator of parameter α_{in}. Here, $k_i^{(*)}$ denotes the optimal choice (in the sense of minimal asymptotic mean squared error) of a sample fraction k in the ith community. The optimal choices $k_1^{(*)}, \ldots, k_s^{(*)}$ can be computed by using a heuristic Eye-Ball method [129]. Another heuristic rule is picking a fixed percentage of the number of nodes in each community, for instance, 5% of the upper-order statistics. These methods have a weak theoretical foundation and might, therefore, not be robust. The minimum distance method (MDM) to estimate the optimal choice k_i^* is proposed in [130]. It is based on a computation of the Kolmogorov–Smirnov distance between the empirical distribution tail of the upper k observations and the power-law distribution (see also [30]). A mathematical analysis of the MDM in a classical context, where data are assumed to come from an i.i.d. model, can be found in [131].

In the context of the tail index estimation in random graphs, the number of nodes within the communities may be different. Let us assume again that s communities with the largest amount of nodes are taken and the ith community, $1 \leq i \leq s$, contains m_i nodes. Let

$I_{(1)}^{(i)} \geq \cdots \geq I_{(m_i)}^{(i)}$ be the decreasing order statistics of $I_1^{(i)}, \ldots, I_{m_i}^{(i)}$. By using only the $r_i + 1$ largest order statistics $I_{(j)}^{(i)}$, $1 \leq j \leq r_i + 1$, the estimator of α_{in} is defined as

$$\hat{\alpha}_{in}^{(5)}(s) = \left(\left(\sum_{i=1}^{s} r_i \right)^{-1} \sum_{i=1}^{s} \sum_{j=1}^{r_i} \ln \left(\frac{I_{(j)}^{(i)}}{I_{(r_i+1)}^{(i)}} \right) \right)^{-1}$$

in [132]. Regarding i.i.d. data, assuming that $\{r_i\}$, $1 \leq i \leq s$ are fixed, $s = s_n$, $s_n \to \infty$, $n/s_n \to \infty$, as $n \to \infty$ and that the law of observations satisfies a classical second order condition of regular variation, the estimator $\hat{\alpha}_{in}^{(5)}(s)$ is asymptotically normal; see Theorem 2 in [132].

A generalization of $\hat{\alpha}_{in}^{(5)}(s)$ can be found in [133]. There, considering the case of equal-sized blocks and i.i.d. data it is shown that, in general, Hill's estimator $\hat{\alpha}_{in}^{(1)}(k)$ does not have a better performance than the estimator proposed in [133].

The reasons for using the block type estimators are the following. The block type estimators may be preferable against the POT estimators when the observations are not exactly i.i.d. However, at most, the block type estimators are introduced by considering the equal-sized blocks, and thus, their application to random graphs is quite restrictive.

The performance of Hill's estimator $\hat{\alpha}_{in}^{(1)}(k^*)$, where k^* is chosen using the MDM, was examined in [131]. There, data were simulated with a directed linear PA model, and the estimation of the tail index of the limiting in-degree distribution was analyzed. It is concluded in [131] that the estimator $\hat{\alpha}_{in}^{(1)}(k^*)$ often works well on the linear PA models under proper choices of parameters.

In [30], an empirical comparison of the EV model with two parametric approaches for the linear PA model (15)–(17), namely, the maximum likelihood and snapshot methods, proposed in [134] can be found. The EV estimates use the before-mentioned Hill's estimator $\hat{\alpha}_{in}^{(1)}(k^*)$ to estimate of the in- and out-degree tail indices $(\iota_{in}, \iota_{out})$ beforehand. The estimation of the tail indices of in- and out-degrees (see Figure 4.1 in [30] for biases comparison) gives the following conclusion. The estimator $\hat{\alpha}_{in}^{(1)}(k^*)$ tends to have a much larger variance than both the maximum likelihood and snapshot methods with slightly more bias. The situation is different when data are simulated from the directed linear PA model but corrupted by the random addition of edges. Then, the EV estimates for $(\delta_{in}, \delta_{out})$ obtained by using Hill's estimator exhibit a smaller bias than the maximum likelihood and snapshot methods. By considering the directed linear PA model with randomly deleting edges, the estimator $\hat{\alpha}_{in}^{(1)}(k^*)$ also gives reasonable results. This allows us to conclude that in the case when the model is misspecified, the semi-parametric estimators can provide an attractive and reliable alternative to the parametric estimation.

Appendix B. Extremal Index Estimation

Among well-known estimators of the extremal index are the interval estimator by [20], the K-gap estimator by [135], the sliding block estimators by [136,137], and the multilevel blocks estimator in [138], among others. The latter estimators are applied to stationary sequences $\{X_i\}$. The main parameter in these estimators requires the choice of a threshold u. One of the high quantiles (i.e., those quantiles close to 100%) of the sample $\{X_i\}_{i=1,\ldots,n}$ is taken usually as u or u is selected visually corresponding to a stability interval of the plot of some estimate $\hat{\theta}(u)$ against u. Following [135], a list of pairs (u, K) is selected using the information matrix test (IMT) in [139]. The semiparametric maxima estimators depend on the block size only [136,137]. In order to select u, a discrepancy method based on the Cramér–von Mises–Smirnov statistic ω^2 and calculated with the k largest order statistics of the underlying sample instead of the entire sample is proposed in [140]. Its asymptotic distribution as $k \to \infty$ is proved to coincide with the ω^2-distribution.

Let us recall the interval estimator. Let $T(u)$ imply the number of observations running under u between two consecutive exceedances. $L - 1$ is the random number of the inter-exceedance times $\{T(u)_i\}$. For exceedance times $1 \leq S_1 < \ldots < S_L \leq n$ it follows

$$T(u)_i = S_{i+1} - S_i, \qquad i \in \{1, \ldots, L - 1\}. \tag{A2}$$

The interval estimator is determined by [20]

$$\hat{\theta}_n(u) = \left\{ \begin{array}{l} \min(1, \hat{\theta}_n^1(u)), \text{ if } \max\{T(u)_i : 1 \leq i \leq L - 1\} \leq 2, \\ \min(1, \hat{\theta}_n^2(u)), \text{ if } \max\{T(u)_i : 1 \leq i \leq L - 1\} > 2, \end{array} \right. \tag{A3}$$

where

$$\hat{\theta}_n^1(u) = \frac{2(\sum_{i=1}^{L-1} T(u)_i)^2}{(L-1)\sum_{i=1}^{L-1}(T(u)_i)^2}, \tag{A4}$$

$$\hat{\theta}_n^2(u) = \frac{2(\sum_{i=1}^{L-1}(T(u)_i - 1))^2}{(L-1)\sum_{i=1}^{L-1}(T(u)_i - 1)(T(u)_i - 2)}, \tag{A5}$$

In [16,21], the interval estimator is modified for random graphs. To this end, we consider influence node indices, e.g., PageRanks, as our data $\{X_i\}$. A numeration of nodes in the graph does not matter. A high quantile of $\{X_i\}$ can be taken as threshold u^*. We mark nodes with exceedances, i.e., such that the event $\{X_i > u^*\}$ holds. Then one can calculate $T(u^*)$ as the length of the path expressed in edges between two nodes whose influence indices exceed the threshold u^*. All internal nodes along the path should have the influence indices less than u. Having a set of such L paths $\{T(u^*)_i, i \in \{1, 2, \ldots L\}\}$, one can calculate the intervals estimate of the graph per (A3)–(A5). Note that the numeration of the paths is not required.

Appendix C. Calculation of the Pagerank and Pagerank Vector

There are numerous approaches to calculate the PageRank R_i of a randomly chosen page $v = i \in V$ in a Web graph $G = (V, E)$. One of them is determined using the following iteration

$$\widehat{R}_i^{(n,0)} = 1, \ \widehat{R}_i^{(n,k)} = \sum_{j \to i} \frac{c}{D_j} \widehat{R}_j^{(n,k-1)} + (1 - c), \ k \in N, \tag{A6}$$

proposed in [28,141], for a given uniform personalization vector $q_i = 1/n, 1 \leq i \leq n = \|V\|$. $R_i^{(n,k)}$ denotes the scale-free version of the PageRank, i.e., $R_i^{(n,k)} = nR_i$ of a node $v = i$. $j \to i$ implies that node j is linked to node i, i.e., $(j, i) \in E$. The iteration (A6) is proceeding until the difference between two consecutive iterations $\|R_i^{(n,k)} - R_i^{(n,k-1)}\|$ in some metric will be small enough, which is sufficient for a moderate number of iterations k.

The mean field method provides another important approach [109]. Its idea is to average the PageRanks of nodes that are aggregated within classes according to their degree $\kappa \equiv (k_{in}, k_{out})$. Such a class includes nodes with the same in-degree k_{in} and the same out-degree k_{out}.

Analytical formulae for the PageRank vector were obtained in [142,143]. In [144] the PageRank is calculated by a gradient procedure. In [145], the PageRank problem is solved with a regularization method by decreasing the regularization parameter to $\sim 1/k$, which leads to a convergence rate with an order of $1/k$ for the kth step of the iterative algorithm. In [146], a robust PageRank problem is stated as a saddle-point problem for solving convex–concave optimization problems. It is solved using deterministic and stochastic Mirror Descent algorithms with convergence rate $\sim 1/\sqrt{k}$. The red light green light (RLGL)

method proposed in [147] demonstrates exponential convergence. However, the problem of the PageRank calculation is not the focus of the present survey.

Appendix D. Definition of Conductance

Let $G = (V, E)$ be the graph with the number of vertices $\|V\|$ and the number of edges $\|E\|$. The conductance measures the minimum relative connection strength between "isolated" subsets $\{S\}$ and the rest of the network [22,23,148]. In [148], the following definition is provided. Let A be the adjacency matrix of a graph G. The conductance of a set of nodes S is defined as

$$\Phi(S) \quad = \quad \frac{\sum_{i \in S, j \in \overline{S}} A_{ij}}{\min\{A(S), A(\overline{S})\}},$$

where $A(S) = \sum_{i \in S} \sum_{j \in V} A_{ij}$, or equivalently $A(S) = \sum_{i \in S} d(i)$, where $d(i)$ is a degree of node i in G. $\overline{S}$ denotes the complement of S. Then the conductance of the graph G is determined with

$$\Phi_G \quad = \quad \min_{S \in V} \Phi(S).$$

References

1. Albert, R.; Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [CrossRef]
2. Bollobás, B.; Borgs, C.; Chayes, J.; Riordan, O. Directed Scale-Free Graphs. In *SODA '03*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2003; pp. 132–139.
3. da Cruz, J.P.; Lind, P.G. The bounds of heavy-tailed return distributions in evolving complex networks. *Phys. Lett. A* **2013**, *377*, 189–194. [CrossRef]
4. Avrachenkov, K.; Dreveton, M. *Statistical Analysis of Networks*; Now Publishers: Boston, MA, USA; Delft, The Netherlands, 2022. [CrossRef]
5. Estrada, E. *The Structure of Complex Networks: Theory and Applications*, online ed.; Oxford Academic: Oxford, UK, 2013. [CrossRef]
6. Newman, M.E.J. *Networks: An Introduction*; Oxford University Press: Oxford, UK; New York, NY, USA, 2010.
7. van der Hofstad, R. *Random Graphs and Complex Networks*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 2017; Volume 1.
8. Gissibl, N.; Klüppelberg, C. Max-linear models on directed acyclic graphs. *Bernoulli* **2018**, *24*, 2693–2720. [CrossRef]
9. Segers, J.; Asenova, S. Max-linear graphical models with heavy-tailed factors on trees of transitive tournaments. *arXiv* **2022**, arXiv:2209.14938.
10. Klüppelberg, C.; Sönmez, E. Max-linear models in random environment. *J. Multivar. Anal.* **2022**, *190*, 104999. [CrossRef]
11. Isaev, M.; Rodionov, I.; Zhang, R.-R.; Zhukovskii, M. Extremal independence in discrete random systems. *arXiv* **2021**, arXiv:2105.04917.
12. Rodionov, I.V.; Zhukovskii, M. The distribution of the maximum number of common neighbors in the random graph. *Eur. J. Comb.* **2023**, *107*, 103602. [CrossRef]
13. Newman, M.E.J. Random Graphs with Clustering. *Phys. Rev. Lett.* **2009**, *103*, 058701. [CrossRef] [PubMed]
14. Markovich, N.M.; Rodionov, I.V. Maxima and sums of non-stationary random length sequences. *Extremes* **2020**, *23*, 451–464. [CrossRef]
15. Markovich, N.M. Weighted maxima and sums of non-stationary random length sequences in heavy-tailed models. *arXiv* **2022**, arXiv:2209.08485.
16. Markovich, N.M. Extremal properties of evolving networks: Local dependence and heavy tails. *Ann. Oper. Res.* **2023** . [CrossRef]
17. Beirlant, J.; Goegebeur, Y.; Teugels, J.; Segers, J. *Statistics of Extremes: Theory and Applications*; Wiley: Chichester, UK, 2004.
18. Leadbetter, M.R.; Lingren, G.; Rootzén, H. *Extremes and Related Properties of Random Sequence and Processes*; Springer: New York, NY, USA, 1983; Chapter 3.
19. Markovich N.M.; Ryzhov M.S.; Vaičiulis M. Tail Index Estimation of PageRanks in Evolving Random Graphs. *Mathematics* **2022**, *10*, 3026. [CrossRef]
20. Ferro, C.A.T.; Segers, J. Inference for Clusters of Extreme Values. *J. R. Statist. Soc. B.* **2003**, *65*, 545–556. [CrossRef]
21. Markovich, N.M.; Ryzhov, M.S. Leader Nodes in Communities for Information Spreading. In *Distributed Computer and Communication Networks. DCCN 2020. Lecture Notes in Computer Science, vol 12563*; Vishnevskiy, V.M., Samouylov, K.E., Kozyrev, D.V., Eds.; Springer: Cham, Switzerland, 2020; pp. 475–484. [CrossRef]
22. Censor-Hillel, K.; Shachnai, H. Partial Information Spreading with Application to Distributed Maximum Coverage. In Proceedings of the 29th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC '10), Zurich, Switzerland, 25–28 July 2010; ACM: New York, NY, USA, 2010; pp. 161–170. [CrossRef]

23. Mosk-Aoyama D.; Shah D. Computing separable functions via gossip. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing (PODC '06)*; ACM: New York, NY, USA, 2006; pp. 113–122.
24. Paulauskas, V. A note on linear processes with tapered innovations. *Lith. Math. J.* **2020**, *60*, 64–79. [CrossRef]
25. Anderson, C. Local limit theorems for the maxima of discrete random variables. *Math. Proc. Camb. Philos. Soc.* **1980**, *88*, 161–165. [CrossRef]
26. Resnick, S.I. *Extreme Values, Regular Variation and Point Processes*; Springer: New York, NY, USA, 1987. [CrossRef]
27. Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117. [CrossRef]
28. Chen, N.; Litvak, N.; Olvera-Cravioto, M. PageRank in Scale-Free Random Graphs. In *WAW 2014, LNCS 8882*; Bonato, A., Fan C.G., Prałat, P., Eds.; Springer: Cham, Switzerland, 2014; pp. 120–131. [CrossRef]
29. Samorodnitsky, G.; Resnick, S.; Towsley, D.; Davis, R.; Willis, A.; Wan, P. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *J. Appl. Prob.* **2016**, *53*, 146–161. [CrossRef]
30. Wan, P.; Wang, T.; Davis, R.A.; Resnick, S.I. Are extreme value estimation methods useful for network data? *Extremes* **2020**, *23*, 171–195. [CrossRef]
31. Garavaglia, A.; van der Hofstad, R.; Litvak, N. Local weak convergence for PageRank. *Ann. Appl. Prob.* **2020**, *30*, 40–79. [CrossRef]
32. Fortunato S.; Boguñá, M.; Flammini, A.; Menczer, F. Approximating PageRank from In-Degree. In *Algorithms and Models for the Web-Graph. WAW 2006. LNCS 4936*; Aiello, W., Broder, A., Janssen, J., Milios E., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 59–71. [CrossRef]
33. Litvak, N.; Scheinhardt, W.R.W.; Volkovich, Y. In-Degree and PageRank: Why Do They Follow Similar Power Laws? *Internet Math.* **2007**, *4*, 175–198. [CrossRef]
34. Vázquez, A.; Pastor-Satorras, R.; Vespignani, A. Large-scale topological and dynamical properties of the internet. *Phys. Rev. E* **2002**, *65*, 066130. [CrossRef] [PubMed]
35. Gao, P.; van der Hofstad, R.; Southwell, A.; Stegehuis, C. Counting triangles in power-law uniform random graphs. *Electron. J. Comb.* **2020**, *27*, 1–28. [CrossRef]
36. Stegehuis, C. Distinguishing Power-Law Uniform Random Graphs from Inhomogeneous Random Graphs Through Small Subgraphs. *J. Stat. Phys.* **2022**, *186*, 37. [CrossRef]
37. House, T. Heterogeneous clustered random graphs. *Europhys. Lett.* **2014**, *105*, 68006. [CrossRef]
38. Jelenkovic, P.R.; Olvera-Cravioto, M. Information ranking and power laws on trees. *Adv. Appl. Prob.* **2010**, *42*, 1057–1093. [CrossRef]
39. Volkovich, Y.V.; Litvak, N. Asymptotic analysis for personalized web search. *Adv. Appl. Probab.* **2010**, *42*, 577–604. [CrossRef]
40. Jelenkovic, P.R.; Olvera-Cravioto, M. Maximums on trees. *Stoch. Process. Appl.* **2015**, *125*, 217–232. [CrossRef]
41. Olvera-Cravioto, M. Asymptotics for weighted random sums. *Adv. Appl. Probab.* **2012**, *44*, 1142–1172. [CrossRef]
42. Asmussen, S.; Foss, S. Regular variation in a fixed-point problem for single- and multi-class banching processes and queues. Branching Processes and Applied Probability. Papers in Honour of Peter Jagers. *Adv. Appl. Probab.* **2018**, *50A*, 47–61. [CrossRef]
43. De Haan, L.; Zhou, C. Trends in Extreme Value Indices. *J. Am. Stat. Assoc.* **2021**, *116*, 1265–1279. [CrossRef]
44. Coja-Oghlan, A.; Krivelevich, M.; Vilenchik, D. Why Almost All k-Colorable Graphs Are Easy. In *STACS 2007. Lecture Notes in Computer Science*; Thomas, W., Weil, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; p. 4393. [CrossRef]
45. Abbe, E. Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **2017**, *18*, 6446–6531. [CrossRef]
46. Alon, N.; Kahale, N. A Spectral Technique for Coloring Random 3-Colorable Graphs. *SIAM J. Comput.* **1997**, *26*, 1733–1748. [CrossRef]
47. Levin, D.A.; Peres, Y. *Markov Chains and Mixing Times*, 2nd Revised ed.; AMS: Providence, RI, USA, 2010.
48. Roberts, G.O.; Rosenthal, J.S.; Segers, J. Extremal indices, geometric ergodicity of Markov chains, and MCMC. *Extremes* **2006**, *9*, 213–229. [CrossRef]
49. Heydenreich, M.; Hirsch, C. Extremal linkage networks. *Extremes* **2022**, *25*, 229–255. [CrossRef]
50. Hill, B.M. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **1975**, *3*, 1163–1174. [CrossRef]
51. Phillips, P.C.B.; Loretan, M. Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *J. Empir. Financ.* **1994**, *1*, 211–248.
52. Quintos, C.; Fan Z.; Phillips, P.C.B. Structural Change Tests in Tail Behaviour and the Asian Crisis. *Rev. Econ. Stud.* **2001**, *68*, 633–663. [CrossRef]
53. Mason, D.M. Laws of Large Numbers for Sums of Extreme Values. *Ann. Probab.* **1982**, *10*, 754–764. [CrossRef]
54. Novak, S.Y. Inference of heavy tails from dependent data. *Sib. Adv. Math.* **2002**, *12*, 73–96.
55. Resnick, S.I.; Stărică, C. Smoothing the Moment Estimate of the Extreme Value Parameter. *Extremes* **1999**, *1*, 263–294. [CrossRef]
56. Goldie, C.M.; Smith, R.L. Slow variation with remainder: Theory and applications. *Quart. J. Math. Oxf.* **1987**, *38*, 45–71. [CrossRef]
57. De Haan, L.; Zhou, C. Extreme Value Analysis with Non-Stationary Observations. Preprint. 2012. Available online: https://personal.eur.nl/ldehaan/noniid28082012.pdf (accessed on 27 August 2012).
58. Vaičiulis, M. Local-maximum-based tail index estimator. *Lith. Math. J.* **2014**, *54*, 503–526. [CrossRef]
59. Ferreira, A.; de Haan, L. On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **2015**, *43*, 276–298. [CrossRef]

60. Wang, T.; Resnick, S.I. Consistency of Hill estimators in a linear preferential attachment model. *Extremes* **2019**, *22*, 1–28. [CrossRef]
61. Fishkind, D.E.; Meng, L.; Sun, A.; Priebe, C.E.; Lyzinski, V. Alignment strength and correlation for graphs. *Pattern Recognit. Lett.* **2019**, *125*, 295–302. [CrossRef]
62. Xiong, J.; Shen, C.; Arroyo, J.; Vogelstein, J. Graph Independence Testing. *arXiv* **2019**, arXiv:1906.03661.
63. van der Hofstad,R.; Litvak, N. Degree-Degree Dependencies in Random Graphs with Heavy-Tailed Degrees. *Internet Math.* **2014**, *10*, 287–334. [CrossRef]
64. Shen, C.; Priebe, C.E.; Vogelstein, J.T. From Distance Correlation to Multiscale Graph Correlation. *J. Am. Stat. Assoc.* **2020**, *115*, 280–291. [CrossRef]
65. Volkovich, Y.; Litvak, N.; Zwart, B. Measuring extremal dependencies in Web graphs. In Proceedings of the WWW '08: 17th International Conference on World Wide Web April, Beijing, China, 21–25 April 2008; ACM: New York, NY, USA, 2008; pp. 1113–1114. [CrossRef]
66. Wang, J.; Resnick, S.I. Degree growth rates and index estimation in a directed preferential attachment model. *Stoch. Process. Their Appl.* **2020**, *130*, 878–906. [CrossRef]
67. Bobkov, S.G.; Danshina, M.A.; Ulyanov, V.V. Rate of Convergence to the Poisson Law of the Numbers of Cycles in the Generalized Random Graphs. In *Operator Theory and Harmonic Analysis. Springer Proceedings in Mathematics and Statistics*; Karapetyants, A.N., Pavlov, I.V., Shiryaev, A.N., Eds.; Springer: Cham, Switzerland, 2021; Volume 358, pp. 109–133.
68. Asenova, S.; Mazo, G.; Segers, J. Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes* **2021**, *24*, 461–500. [CrossRef]
69. Engelke, S.; Hitz, A.S. Graphical models for extremes. *J. Royal. Stat. Soc. Ser. B* **2020**, *82*, 1–38. [CrossRef]
70. Papastathopoulos, I.; Strokorb, K. Conditional independence among max-stable laws. *Stat. Probab. Lett.* **2016**, *108*, 9–15. [CrossRef]
71. Barabási, A.-L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512. [CrossRef] [PubMed]
72. Krapivsky, P.L.; Redner, S. Organization of growing random networks. *Phys. Rev.* **2001**, *E63*, 066123. [CrossRef]
73. Norros, I.; Reittu, H. On a conditionally poissonian graph process. *Adv. Appl. Probab.* **2006**, *38*, 59–75. [CrossRef]
74. Allendorf, D.; Meyer, U.; Penschuck, M.; Tran, H. Parallel and I/O-Efficient Algorithms for Non-Linear Preferential Attachment. *arXiv* **2022**, arXiv:2211.06884.
75. Oliveira, R.; Spencer, J. Connectivity transitions in networks with superlinear preferential attachment. *Internet Math.* **2005**, *2*, 121–163. [CrossRef]
76. Bollobás, B.; Riordan, O. The Diameter of a Scale-Free Random Graph. *Combinatorica* **2004**, *24*, 5–34. [CrossRef]
77. Jacob, E.; Mörters, P. A Spatial Preferential Attachment Model with Local Clustering. In *Algorithms and Models for the Web Graph. WAW 2013. Lecture Notes in Computer Science*; Bonato, A., Mitzenmacher, M., Prałat, P., Eds.; Springer: Cham, Switzerland, 2013; Volume 8305, pp. 14–25. [CrossRef]
78. Aiello, W.; Bonato, A.; Cooper, C.; Janssen, J.; Prałat, P. A spatial web graph model with local influence regions. *Internet Math.* **2009**, *5*, 175–196. [CrossRef]
79. Grindrod, P. Range-dependent random graphs and their application to modeling large small-world Proteome datasets. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2002**, *66 (6 Pt 2)*, 066702. https://.. [CrossRef]
80. Kleinberg, J. The Small-World Phenomenon and Decentralized Search. *SIAM News* **2004**, *37*, 1–2.
81. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440–442. [CrossRef]
82. Garavaglia, A.; Hazra, R.S.; van der Hofstad, R.; Ray, R. Universality of the local limit of preferential attachment models. *arXiv* **2022**, arXiv:2212.05551v1.
83. Deijfen, M.; van den Esker, H.; van der Hofstad, R.; Hooghiemstra, G. A preferential attachment model with random initial degrees. *Ark. Mat.* **2009**, *47*, 41–72. [CrossRef]
84. Bharat, K.; Broder, A. A technique for measuring the relative size and overlap of public Web search engines. *Comput. Netw. Isdn Syst.* **1998**, *30*, 379–388. [CrossRef]
85. Wang, J.; Resnick, S.I. Poisson Edge Growth and Preferential Attachment Networks. *Methodol. Comput. Appl. Probab.* **2023**, *25*. [CrossRef]
86. Michielan, R.; Litvak, N.; Stegehuis, C. Detecting hyperbolic geometry in networks: Why triangles are not enough. *Phys. Rev. E* **2022**, *106*, 054303. [CrossRef]
87. Iskhakov, L.; Kamiński, B.; Mironov, M.; Prałat, P.; Prokhorenkova, L. Clustering Properties of Spatial Preferential Attachment Model. In *Algorithms and Models for the Web Graph. WAW 2018. Lecture Notes in Computer Science*; Bonato, A., Prałat, P., Raigorodskii, A., Eds.; Springer: Cham, Switzerland, 2018; Volume 10836, pp. 30–43. [CrossRef]
88. Bagrow, J.P.; Brockmann, D. Natural Emergence of Clusters and Bursts in Network Evolution. *Phys. Rev. X* **2013**, *3*, 021016. [CrossRef]
89. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]
90. Clauset, A.; Newman, M.E.; Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **2004**, *70*, 066111. [CrossRef]
91. Dugué, N.; Perez, A. Directed Louvain: Maximizing Modularity in Directed Networks. Ph.D. Thesis, Université d'Orléans, Orléans, France, 2015. [CrossRef]
92. Newman, M.E.J. Equivalence between modularity optimization and maximum likelihood methods for community detection. *Phys. Rev. E* **2016**, *94*, 052315. [CrossRef]

93. Arenas, A.; Fernández, A.; Gómez, S. Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* **2008**, *10*, 053039. [CrossRef]
94. Reichardt, J.; Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **2006**, *74*, 016110. [CrossRef] [PubMed]
95. Markovich, N.M.; Ryzhov, M.S. Clusters of Exceedances for Evolving Random Graphs. In *Distributed Computer and Communication Networks: Control, Computation, Communications. DCCN 2022. Lecture Notes in Computer Science*; Vishnevskiy, V., Samouylov, K., Kozyrev, D., Eds.; Springer: Cham, Switzerland, 2022; Volume 13766, pp. 67–74. [CrossRef]
96. Kleinberg, J.M.; Kumar, R.; Raghavan, P.; Rajagopalan, S.; Tomkins, A.S. The Web as a Graph: Measurements, Models, and Methods. In *Computing and Combinatorics. COCOON 1999. Lecture Notes in Computer Science*; Asano, T., Imai, H., Lee, D.T., Nakano, S., Tokuyama, T., Eds.; Springer: Berlin/Heidelberg, Germany, 1999; Volume 1627. [CrossRef]
97. de Siqueira Santos, S.; Fujita, A.; Matias, C. Spectral density of random graphs: Convergence properties and application in model fitting. *J. Complex Netw.* **2021**, *9*, cnab041. [CrossRef]
98. Takahashi, D.Y.; Sato, J.R.; Ferreira, C.E.; Fujita, A. Discriminating Different Classes of Biological Networks by Analyzing the Graphs Spectra Distribution. *PLoS ONE* **2012**, *7*, e49949. [CrossRef] [PubMed]
99. Bringmann, K.; Keusch, R.; Lengler, J. Geometric inhomogeneous random graphs. *Theor. Comput. Sci.* **2019**, *760*, 35–54. [CrossRef]
100. Ravasz, E.; Barabási, A.-L. Hierarchical organization in complex networks. *Phys. Rev.* **2003**, *E67*, 026112. [CrossRef]
101. Csányi, G.; Szendrői, B. Structure of a large social network. *Phys. Rev. E* **2004**, *69*, 036131. [CrossRef]
102. Avrachenkov, K.; Kadavankandy, A.; Litvak, N. Mean Field Analysis of Personalized PageRank with Implications for Local Graph Clustering. *J. Stat. Phys.* **2018**, *173*, 895–916. [CrossRef]
103. Rosvall, M.; Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1118–1123. [CrossRef]
104. Peixoto, T.P. Bayesian stochastic blockmodeling. In *Advances in Network Clustering and Blockmodeling*; Dorelan, P., Batagelj, V., Ferligoj, A., Eds.; Wiley: New York, NY, USA, 2019; pp. 289–332.
105. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174. [CrossRef]
106. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *10*, P10008. [CrossRef]
107. Leicht, E.A.; Newman, M.E.J. Community structure in directed networks. *Phys. Rev. Lett.* **2008**, *100*, 118703. [CrossRef]
108. Galhotra, S.; Bagchi, A.; Bedathur, S.; Ramanath, M.; Jain, V. Tracking the conductance of rapidly evolving topic-subgraphs. In *Proceedings of the VLDB Endowment*; VLDB Endowment: Sydney, Australia, 2015; Volume 8, pp. 2170–2181. [CrossRef]
109. Fortunato, S.; Bofuñá, M.; Flammini, A.; Menczer, F. On Local Estimations of PageRank: A Mean Field Approach. *Internet Math.* **2007**, *4*, 245–266. [CrossRef]
110. Newman, M.E.J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **2004**, *69*, 066133. [CrossRef]
111. Kirkley, A.; Newman, M.E.J. Representative community divisions of networks. *Commun. Phys.* **2022**, *5*, 40. [CrossRef]
112. Aspvall, B.; Gilbert, J.R. Graph Coloring Using Eigenvalue Decomposition. *SIAM J. Algebr. Discret. Methods* **1984**, *5*, 526–538. [CrossRef]
113. Wang, J.; Lu, S.; Wang, S.H.; Zhang, Y.-D. A review on extreme learning machine. *Multimed. Tools Appl.* **2022**, *81*, 41611–41660. [CrossRef]
114. Vignotto, E.; Engelke, S. Extreme value theory for anomaly detection—The GPD classifier. *Extremes* **2020**, *23*, 501–520. [CrossRef]
115. Rudd, E.M.; Jain, L.P.; Scheiner, W.J.; Boult, T.E. The extreme value machine. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 762–768. [CrossRef]
116. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [CrossRef]
117. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925–979. https://link.aps.org/doi/10.1103/RevModPhys.87.925. [CrossRef]
118. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the Spread of Influence through a Social Network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; ACM: New York, NY, USA, 2003; pp. 137–146. [CrossRef]
119. Markovich, N.M.; Ryzhov, M.S. Information Spreading and Evolution of Non-Homogeneous Networks. *Adv. Syst. Sci. Appl.* **2022**, *22*, 21–33. [CrossRef]
120. Stephenson, K.; Zelen, M. Rethinking centrality: Methods and examples. *Soc. Netw.* **1989**, *11*, 1–37. [CrossRef]
121. Patwardhan, S.; Radicchi, F.; Fortunato, S. Influence Maximization: Divide and Conquer. *arXiv* **2022**, arXiv:2210.01203.
122. Jessen, A.H.; Mikosch, T. Regularly varying functions. *Publ. Inst. Math. (Beograd) (N.S.)* **2006**, *80*, 171–192. [CrossRef]
123. Robert, C.Y.; Segers, J. Tails of random sums of a heavy-tailed number of light-tailed terms. *Insur. Math. Econ.* **2008**, *43*, 85–92. [CrossRef]
124. Dekkers, A.L.M.; Einmahl, J.H.J.; de Haan L. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **1989**, *17*, 1833–1855. [CrossRef]
125. Resnick, S.I.; Stărică, C. Consistency of Hill's Estimator for Dependent Data. *J. Appl. Probab.* **1995**, *32*, 139–167. [CrossRef]
126. Draisma, G.; de Haan, L.; Peng, L.; Pereira, T.T. A Bootstrap-based Method to Achieve Optimality in Estimating the Extreme-value Index. *Extremes* **1999**, *2*, 367–404. [CrossRef]
127. De Haan, L.; Peng, L. Comparison of tail index estimators. *Statist. Ned.* **1998**, *52*, 60–70. [CrossRef]

128. Das, B.; Resnick, S.I. QQ Plots, Random Sets and Data from a Heavy Tailed Distribution. *Stoch. Model.* **2008**, *24*, 103–132. [CrossRef]
129. Resnick, S.; Stărică, C. Smoothing the Hill Estimator. *Adv. Appl. Probab.* **1997**, *29*, 271–293. [CrossRef]
130. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
131. Drees, H.; Jansen, A.; Resnick, S.I.; Wang, T. On a minimum distance procedure for threshold selection in tail analysis. *SIAM J. Math. Data Sci.* **2020**, *2*, 75–102. [CrossRef]
132. Qi, Y. On the tail index of a heavy tailed distribution. *Ann. Inst. Stat. Math.* **2010**, *62*, 277–298. [CrossRef]
133. Markovich, N.; Vaičiulis, M. Asymptotic Properties of the Block-Type Statistics. *Adv. Syst. Sci. Appl.* **2022**, *22*, 106–123.
134. Wan, P.; Wang, T.; Davis, R.A.; Resnick, S.I. Fitting the linear preferential attachment model. *Electron. J. Statist.* **2017**, *11*, 3738–3780. [CrossRef]
135. Süveges, M.; Davison, A.C. Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat.* **2010**, *4*, 203–221. [CrossRef]
136. Berghaus, B.; Bücher, A. Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Stat.* **2018**, *46*, 2307–2335. [CrossRef]
137. Northrop, P.J. An efficient semiparametric maxima estimator of the extremal index. *Extremes* **2015**, *18*, 585–603. [CrossRef]
138. Sun, J.; Samorodnitsky, G. Multiple thresholds in extremal parameter estimation. *Extremes* **2019**, *22*, 317–341. [CrossRef]
139. Fukutome, S.; Liniger, M.A.; Süveges, M. Automatic threshold and run parameter selection: A climatology for extreme hourly precipitation in Switzerland. *Theor. Appl. Climatol.* **2015**, *120*, 403–416. [CrossRef]
140. Markovich, N.M.; Rodionov, I.V. Threshold selection for extremal index estimation. *arXiv Prepr.* **2020**, arXiv:2009.02318.
141. Chen, N.; Litvak, N.; Olvera-Cravioto, M. Ranking Algorithms on Directed Configuration Networks. *arXiv* **2014**, arXiv:1409.7443v2.
142. Avrachenkov, K.; Lebedev, D. PageRank of scale-free growing networks. *Internet Math.* **2006**, *3*, 207–231. [CrossRef]
143. Markovich, N.M.; Krieger, U.R. The PageRank Vector of a Scale-Free Web Network Growing by Preferential Attachment. In *Distributed Computer and Communication Networks: Control, Computation, Communications. DCCN 2021. Lecture Notes in Computer Science*; Vishnevskiy, V., Samouylov, K., Kozyrev, D., Eds.; Springer: Cham, Switzerland, 2021; Volume 13144, pp. 24–31. [CrossRef]
144. Nazin, A.V.; Polyak, B.T. Randomized algorithm to determine the eigenvector of a stochastic matrix with application to the PageRank problem. *Autom. Remote Control* **2011**, *72*, 342–352. [CrossRef]
145. Polyak, B.T.; Tremba, A.A. Regularization-based solution of the PageRank problem for large matrices. *Autom. Remote Control* **2012**, *73*, 1877–1894. [CrossRef]
146. Nazin, A.V.; Tremba, A.A. Saddle point mirror descent algorithm for the robust PageRank problem. *Autom. Remote Control* **2016**, *77*, 1403–1418. [CrossRef]
147. Avrachenkov, K.; Brown, P.; Litvak, N. Red Light Green Light Method for Solving Large Markov Chains. *J. Sci. Comput.* **2022**, *93*, 18. [CrossRef]
148. Leskovec, J.; Lang, K.J.; Dasgupta, A.; Mahoney, M.W. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Math.* **2009**, *6*, 29–123. [CrossRef]