Extremal properties of evolving networks

Natalia Markovich*

December 9, 2022

Real-world networks display a strong heterogeneity that is reflected in a heavy-tailed distribution of node influence indices. The in-degree, the PageRank (PR) and the Max-Linear Model (MLM) may be used as node influence indices of random graphs. The indices are regularly varying heavy-tailed distributed random variables (r.v.s) [8]. We predict the tail index (TI) and extremal index (EI) of the PRs and MLMs of newly appended nodes in evolved random networks. The TI shows a heaviness of the distribution tail. The reciprocal of the EI approximates a mean cluster size of the stochastic process. The cluster of exceedances means a set of consecutive exceedances of the process over a sufficiently high threshold. A network evolution means that a directed graph grows by adding single edges at discrete time steps. At each such step a new node may be added or not. The evolution can be modeled by preferential attachment (PA) tools, see [7], [9] among others. Then a newly appended node prefers to attach to an existing node with a large node degree. For simulation we use the $\alpha-$, $\beta-$ and $\gamma-$ PA schemes proposed in [9] since they create graphs with multiple edges between nodes and self-loops.

The paper is based on results obtained in [6], where the TI and EI of the PR and MLM in PA-evolved random graphs are predicted. To this end, results of extreme value theory obtained in [3]-[5] are applied to random networks. Namely, the TI and EI of sums and maxima of weighted non-stationary random length sequences of regularly varying r.v.s are derived in [3]-[5]. It was found that the sums and maxima have the same TI and EI that are equal to ones of the most heavy-tailed term in the sum or maximum if the latter term is unique. If there are a random number of the most heavy-tailed

*V.A. Trapeznikov Institute of Control Sciences Russian Academy of Sciences, nat.markovich@gmail.com

weakly dependent terms, then the TI of sums and maxima is still the same but their EI does not exist.
Starting with a set of weakly connected stationary seed communities as a hot spot and ranking them with regard to their TIs, the TI and EI of new nodes that are appended to the network may be determined by the most heavy-tailed community. This procedure allows us to predict a temporal network evolution in terms of the TI and EI. The exposition is provided by algorithms, examples and a study of simulated and real evolved random graphs.

1 Main results

In [3] a doubly-indexed array $\{Y_{n,i} : n, i \geq 1\}$ of nonnegative r.v.s in which the "row index" n corresponds to time, and the "column index" i enumerates the series is considered. The length N_n of "row" sequences $\{Y_{n,i} : i \geq 1\}$ for each n is generally random. $\{N_n : n \geq 1\}$ is a sequence of non-negative integer-valued r.v.s. For each i the "column" sequence $\{Y_{n,i} : n \geq 1\}$ is assumed to be strict-sense stationary with EI θ_i having a regularly varying distribution tail $P\{Y_{n,i} > x\} = \ell_i(x)x^{-k_i}$ with TI $k_i > 0$ and a slowly varying function $\ell_i(x)$. There are no assumptions on the dependence structure in i. Assuming that there is a unique "column" sequence with a minimum TI k_1, and N_n has a lighter tail than $Y_{n,i}$ it was found in [3] that the TI and EIs of the weighted sums and maxima

$$\begin{aligned} Y_n^*(z, N_n) &= \max(z_1 Y_{n,1}, ..., z_{N_n} Y_{n,N_n}), \\ Y_n(z, N_n) &= z_1 Y_{n,1} + ... + z_{N_n} Y_{n,N_n} \end{aligned} \qquad (1)$$

for positive constants $z_1, z_2, ...$ are equal to k_1 and θ_1. If there is a random number of such "column" sequences with k_1, then the sums and maxima may have the same TI k_1 if the latter "column" sequences are weakly dependent, but their EI does not exist, [4]. These results are applied in [6] to random graphs.
Let us explain the main idea. A seed graph from which the evolution starts is divided into communities. The PRs of nodes in the community are calculated. The communities are ranked by their TI estimates. The communities are considered as the "column" sequences. N_n is the random number of communities. We call the community as "dominating" if its TI is minimal and thus, the distribution of its PRs is the most heavy-tailed. Let a set of new nodes be appended within a fixed time of the evolution. A new node can be

considered as a root of the tree. Then N_n is at the same time an in-degree of the root node n that is the number of its nearest neighbors with in-coming links to the root. The new nodes are divided into classes. If a new node has at least one link to the community with the minimum TI k_1, then the node relates to Class 1 with the TI k_1 and the EI θ_1 if such community is unique. The TIs may only be estimated. Hence, we may assume that there is such unique community. Those nodes which have no links to the "dominating" community may have links to the second "dominating" one that has the second minimum TI $k_2 > k_1$. The nodes relate to Class 2, etc. A theorem proved in [6] states that the TI and EI of the newly appended nodes are determined by the TI and EI of the seed graph.

Let us explain more precisely how (1) may relate to the PRs and the MLMs. The PR R of a randomly chosen Web page (a node in the Web graph) is viewed as a r.v.. It has been considered as the solution to the fixed-point problem

$$R =^D \sum_{j=1}^{N} A_j R_j + Q \tag{2}$$

in [1], [8]. $=^D$ denotes equality in distribution. In the same way, a MLM is considered as the 'minimal/endogeneous' solution of the equation [2]

$$R =^D \left(\bigvee_{j=1}^{N} A_j R_j \right) \vee Q. \tag{3}$$

One can rewrite the right-hand sides of (2) and (3) as

$$Y_i(c, N_i) \;=\; c\sum_{j=1}^{N_i} Y_{i,j} + Q_i, \qquad Y_i^*(c, N_i) = c\bigvee_{j=1}^{N_i} Y_{i,j} \vee Q_i, \quad i \in \{1, ..., n\}.$$

This relates to the definition of Google's PR with a damping factor $c > 0$. Q is a personalization value of the node [8]. We deal with recursions

$$Y_{i,j}^{(m)} \;=\; c\sum_{s=j}^{N_i} Y_{i,s}^{(m-1)} + Q_i, \tag{4}$$

$$X_{i,j}^{(m)} \;=\; \left(c\bigvee_{s=j}^{N_i} X_{i,s}^{(m-1)} \right) \vee Q_i, \quad \{X_{i,j}^{(0)}\} \equiv \{Y_{i,j}^{(0)}\}, \tag{5}$$

$m \geq 1$, $i \in \{1,2,...,n\}$, $j \in \{1,2,...,N\}$, $N = \lim_{n\to\infty} \sup_{1\leq i\leq n} N_i$. Let us consider matrices related to the scheme of series $\{Y_{n,i}^{(0)} : n, i \geq 1\}$ and corresponding TI and EI $(k_i^{(0)}, \theta_i^{(0)})$:

$$A^{(0)} = \begin{pmatrix} Y_{1,1}^{(0)} & Y_{1,2}^{(0)} & Y_{1,3}^{(0)} & ... & 0 & Q_1 \\ Y_{2,1}^{(0)} & 0 & Y_{2,3}^{(0)} & ... & Y_{2,N_2}^{(0)} & Q_2 \\ ... & ... & ... & ... & ... & ... \\ Y_{n,1}^{(0)} & Y_{n,2}^{(0)} & Y_{n,3}^{(0)} & ... & Y_{n,N_n}^{(0)} & Q_n \end{pmatrix}, \quad (6)$$

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

$\{Q_i\}$ is a sequence of independent identically distributed r.v.s. Thus, its EI is equal to 1. Network communities may be interpreted as the columns of $A^{(0)}$. A zero jth element in the ith row $Y_{i,j}^{(0)}$ of $A^{(0)}$ means that the ith root node has no followers in the jth community or there is no link between them. For instance, 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. If books are cited by papers from the "dominating" community, then their TI and EI are determined by ones of the latter papers. The matrix $A^{(0)}$ corresponds to an initial (un)directed graph (a seed network) that is used to model an evolution of the graph in time by means of some attachment tool. The jth column $\{Y_{i,j}^{(m)}\}_{i\geq 1}$ (or $\{X_{i,j}^{(m)}\}_{i\geq 1}$) of the matrix $A^{(m)}$ is defined by (4) (or (5)) using the submatrix $\{Y_{n,i}^{(m-1)} : n \geq 1, i \geq j\}$ (or $\{X_{n,i}^{(m-1)}\} : n \geq 1, i \geq j\}$) of the matrix $A^{(m-1)}$. It is derived particularly in [6] that $\{Y_{i,j}^{(m)}\}_{i\geq 1}$ and $\{X_{i,j}^{(m)}\}_{i\geq 1}$ calculated by (4) and (5) have the same TI $k_j^{(0)} < k$, where $k := \lim_{n\to\infty} \inf_{j+d_j\leq i\leq l_n} k_i^{(0)}$ and the same EI $\theta_j^{(0)}$ if $d_j = 1$ for any $1 \leq j \leq l_n - 1$ for any $m \geq 1$.

2 Discussion and open problems

Since the nodes of the graph cannot be definitely enumerated, the definition and testing of the stationarity in the graphs or their communities remain an open problem. One can determine 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.
The extremal index exists for stationary random sequences. The extremal index of the PRs and MLMs of the sequence of the root nodes is considered.

The author was supported by the Russian Science Foundation (grant No. 22-21-00177).

References

[1] Jelenkovic, P. R., Olvera-Cravioto, M. Information ranking and power laws on trees, *Adv. Appl. Prob.*, Vol. 42, No.4, 2010, pp.1057-1093.

[2] Jelenkovic, P. R., Olvera-Cravioto, M. Maximums on trees, *Stoch. Process. Appl.*, Vol. 125, 2015, pp.217-232.

[3] Markovich N.M., Rodionov I.V. Maxima and sums of non-stationary random length sequences, *Extremes*, Vol. 23, No.3, 2020, pp.451-464.

[4] Markovich N.M. Extremes of Sums and Maxima with Application to Random Networks. In: Proceedings of the 5th International Conference on Stochastic Methods 2020 ICSM5 November 23-27, 2020, 107-120 Moscow, Russia, eprint arXiv:2110.04120 October 2021

[5] Markovich N.M. Weighted maxima and sums of non-stationary random length sequences in heavy-tailed models. Working paper, 2022.

[6] Markovich N. M. Extremal properties of evolving networks: local dependence and heavy tails, *Annals of Operation Research*, Under revision.

[7] Norros I., Reittu H. On a conditionally poissonian graph process, *Adv. Appl. Prob. (SGSA)*, Vol. 38, 2006, pp.59-75.

[8] Volkovich Y., Litvak N. Asymptotic analysis for personalized web search, *Adv. Appl. Prob.*, Vol. 42, No.2, 2010, pp. 577-604.

[9] Wan, P., Wang, T., Davis, R. A., Resnick, S.I. Are extreme value estimation methods useful for network data? *Extremes*, Vol. 23, 2020, pp. 171-195.