On exponential convergence of dynamic queueing network and its applications*

Elmira Yu. Kalimulina[1[0000−0001−7158−040X]]

Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia
elmira.yu.k@gmail.com
https://www.ipu.ru/staff/elmira

Abstract. This paper is a continuation of previous research in ergodicity of some models for unreliable networks. The set of random graphs and the sequence of matrixes describing the failure and recovery process has been used instead of the fixed graph for network structure. The main results about an ergodicity and bounds for rate of convergence to stationary distribution are formulated under more general assumptions on intensity rates.

Keywords: Ergodicity · Stochastic networks · Convergence rate · spectral gap.

1 Introduction

Let's remind what the standard queueing network (Jackson's type) is. The standard queueing network is the network with following parameters (see Fig.1) [2, 9]:

- the network consists of m nodes, $M = \{1, 2, \dots, m\}$;
- each node is a multi-server system with an infinite waiting room;
- the algorithm of service is FCFS (First Come First Served);
- all customers are supposed to be indistinguishable;
- there is an external Poisson arrival flow with intensity Λ (only the open queueing network is considered in this research);
- denote the routing matrix as $R = (r_{ij}), i, j = 0, 1, \dots, m$; without loss of generality R is supposed to be regular;
- denote the traffic vector as $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_m)$;

* The publication has been prepared with the support of the Russian Foundation for Basic Research according to the research project No.20-01-00575 A.

– denote service rates as $\mu = (\mu_1(n_1), \ldots, \mu_m(n_m))$;

– the number of customers in the system is denoted as $\mathbf{n} = (n_1, \ldots, n_m)$.

The state space for stochastic process describing this system is following:

$$\mathbf{n} = (n_1, , n_2, \cdots, n_m) \in \mathbf{Z}_+^m = \mathbf{E}, \tag{1}$$

and with the following transitions:

$$T_{ij}\mathbf{n} = (n_1, \cdots, n_i - 1, \cdots, n_j + 1, \cdots, n_m),$$
$$T_{0j}\mathbf{n} = (n_1, \cdots, n_j + 1, \cdots, n_m),$$
$$T_{i0}\mathbf{n} = (n_1, \cdots, n_i - 1, \cdots, n_m).$$

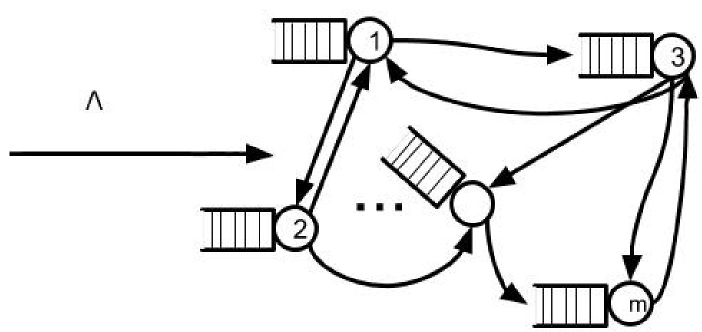

Fig. 1. Standard queueing network.

The important problem for such a network is the existence of a limit distribution and the rate of convergence to it. These problem are well studied by many researchers. One of the well-known results can be found in [9–11].

We are interested to study some modification of this standard model. The general motivation for our research is the real systems modelling such as transport networks, computer networks, telecommunication traffic models and etc. One of the key feature of these real systems is a changing structure. These systems are well described (in some approach) by queuing systems and networks models. But the changing structure (due unreliable nodes or part-time regime of operation) demands some modification of the standard approach. The problem with the standard approach is that the classical models don't include parameters specific for real systems. They are more complicated than standard queueing networks models.

So the following modification of a standard model described above (unreliable network) is considered here[3, 4]:

- each node may switch on/off (ex. break down and repair) with intensities α_i, β_i, $i = 1, ..., m$;

- a dynamic routing is being applied as a failure management mechanisms.

The principle of "dynamic routing" is in selecting the alternative node if the target node is under failure. The alternative node is selected from the nearest to the failed one. This modification make this model different from another similar ones.

There are several alternative failure management mechanisms: one of them is "blocking" (before service and after service), for details see [7, 8]. The approach suggested here is more specious for real systems, but it demands the more complicated random process to be considered.

2 Dynamic routing

The "Dynamic routing" failure management mechanisms results the extended state space by adding some component to standard state space of the process. The Fig.2 shows the initial structure of the network. The standard approach implies this graph to be fixed. Recoveries and failures form a new process that describes the transformation of this graph to another in the suggested model.

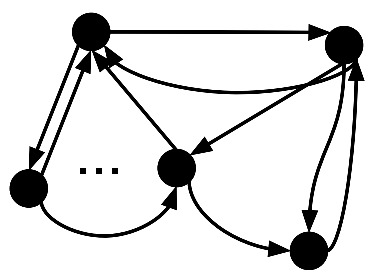

Fig. 2. Initial structure of the network.

The number of nodes is fixed, new nodes don't appear in contrast with growing networks (see [15, 16]).

In our model nodes can be blocked (by deleting/adding edges to it). This way of transformation is shown on Fig.3.

The nodes marked with red color are under failure, the transition forward and back from one graph (with working node i) to another (when the node is under failure) occurs with intensities α_i (failure rate) and β_i (recovery rate). This way of graph transformation generates the Markov process with a finite state space (because the number of nodes is finite).

We denote the state space of the graph transformation process as the set G: the node i is "removed" with some intensity α_i (failure rate for this node) or it can be restored with some intensity β_i.

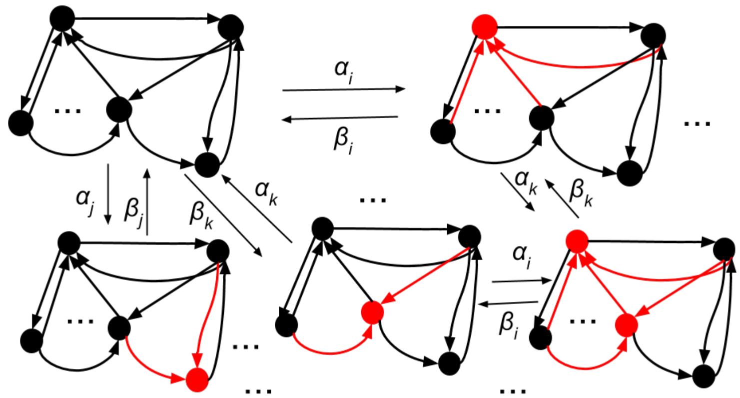

Fig. 3. The network graph evolution process.

So, the standard state space (1) for our network process is extended by adding the component G and is following:

$$\tilde{\mathbf{n}} = (G, n_1, n_2, ..., n_m) \in |G| \times \mathbf{Z}_+^m =: \mathbf{E},$$

where G is a component describing the graph (or transition matrix) transformation.

We can find a degree distribution for the process from state space G. The average number of vertices of degree k at time t: $\{E(k,t)\} = E\ P(k,t)$ can be described by the equation:

$$\begin{aligned}\{E(k,t+1)\} = \{E(k,t)\} - \frac{\alpha_k}{E\sum_k P(k)\alpha_k}\{E(k,t)\} + \\ + \frac{\alpha_{k-1}}{E\sum_{k-1} P(k-1)\alpha_{k-1}}\{E(k-1,t)\} + \\ + \frac{\alpha_{k+1}}{E\sum_{k+1} P(k+1)\alpha_{k+1}}\{E(k+1,t)\}.\end{aligned}$$

It describes the evolution of graph of our network structure in time and for the continuous time takes the form:

$$\begin{aligned}E\frac{\partial P(k,t)}{\partial t} = -\alpha_k P(k,t) + \alpha_{k-1}P(k-1,t) + \\ P(k+1,t) + \alpha_{k+1}P(k+1,t).\end{aligned} \quad (2)$$

Is easy to see for this equation that (2) is linear homogeneous equation (under assumption of constant failure and recovery rates) and has a stationary solution:

$$P(k) = \lim_{t\to\infty} P(k,t). \tag{3}$$

3 Main results

3.1 Convergence of process $X_S(t)$

The state for this network process is described by the following vector

$$\boldsymbol{n} = ((n_1, s_1), (n_2, s_2), ..., (n_m, s_m)),$$

where n_i – the number of customers at the i-th node and

$$s_i = \{ 0 \text{ , if the } i\text{th node works}, 1, \text{ otherwise}.$$

The behaviour of $\boldsymbol{n}$ is a Markov chain in continuous time. It includes an embedded homogeneous Markov chain with positive probabilities for transitions:

$$\begin{aligned} s_i &\longrightarrow (1 - s_i), \\ n_i &\longrightarrow (n_i \pm 1). \end{aligned} \tag{4}$$

Exponential convergence of reliability process $\boldsymbol{S} = (s_1, \ldots, s_m)$ converges to stationary distribution with exponential rate.

Let's consider the reliability process $X_S(t)$ of our model separately.

$$\{X_{S_i}(t+1) = X_{S_i}(t)\} = \frac{\sum\limits_{j=1}^{m} \gamma_j - \gamma_i}{\sum\limits_{j=1}^{m} \gamma_j},$$

$$\{X_{S_i}(t+1) = 1 - X_{S_i}(t)\} = \frac{\gamma_i}{\sum\limits_{j=1}^{m} \gamma_j},$$

where

$$\gamma_i = \alpha_i \mathbf{1}\{s_i = 0\} + \beta_i \mathbf{1}\{s_i = 1\}.$$

3.2 Convergence of process $X_R(t)$

The behaviour of the process $X_R(t)$ is defined by the process $X_S(t)$ with the same transition probabilities. It takes values from the finite set ($R = \|r_{ij}(t)\|$), so $X_R(t)$ has the stationary distribution and converges to it exponentially. The sequence of $R = \|r_{ij}(t)\|$ has a limit $\tilde{R} = \|\tilde{r}_{ij}\|$, where $\tilde{r}_{ij}$ are dependent random variables.

Processes $X_S(t)$ and $X_R(t)$ describe only reliability of our network. At this moment we still haven't took into consideration the service process and an input flow, that are our main interest of studying.

But they are ergodic and don't depend on the input flow and service process (in further we will apply these facts).

Definition of the main network process The process has the following state space:

$$\tilde{\mathbf{n}} = (G, n_1, n_2, \ldots, n_m) \in G \times \mathbf{Z}_+^m =: \mathbf{E}$$

The following transitions in a network are possible:

$$\begin{aligned}
T_{ij}\tilde{\mathbf{n}} &:= (G, n_1, \ldots, n_i - 1, \ldots, n_j + 1 \ldots, n_m),\\
T_{0j}\tilde{\mathbf{n}} &:= (G, n_1, \ldots, n_j + 1, \ldots, n_m),\\
T_{i0}\tilde{\mathbf{n}} &:= (G, n_1, \ldots, n_i - 1, \ldots, n_m),\\
T_f\tilde{\mathbf{n}} &:= (G^+, n_1, \ldots, n_m),\\
T_r\tilde{\mathbf{n}} &:= (G^-, n_1, \ldots, n_m).
\end{aligned}$$

Definition. We will call a "dynamic routing network" the process

$$\mathbf{X} = (X(t), t \geq 0)$$

defined by the following infinitesimal generator:

$$\begin{aligned}
&\mathbf{Q}f(\mathbf{n}) = \\
&\sum_{i=1}^{m}\sum_{j=1}^{m}(f(T_{0j}\mathbf{n}) - f(\mathbf{n}))\lambda_i r_{ij} + \\
&\sum_{i=1}^{m}\sum_{j=1}^{m}(f(T_{ij}\mathbf{n}) - f(\mathbf{n}))\mu_i(n_i) r_{ij} + \\
&\sum_{k\in G^+}(f(T_k\mathbf{n}) - f(\mathbf{n}))\alpha_k + \\
&\sum_{k\in G\setminus G^+}(f(T_k\mathbf{n}) - f(\mathbf{n}))\beta_k + \\
&\sum_{i=1}^{m}(f(T_{i0}\mathbf{n}) - f(\mathbf{n}))\mu_i(n_i) r_{i0},
\end{aligned} \tag{5}$$

where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ satisfies the balance equations.

Suppose the infinitesimal generator (5) satisfies the following **assumptions**:

1. $$\inf_{\mathbf{n},i} \sum_{i=1}^{m} \frac{\alpha_i \mu_i(\mathbf{n})}{\alpha_i + \beta_i} > \Lambda;$$
2. $\tilde{R} = \|\tilde{r}_{ij}\|$ is irreducible, so the expectation of steps visited by one customer within the network is finite.

The second condition may be checked for $R(t)$ under large t. The convergence rate of $R(t)$ may be estimated from the Markov-Doeblin condition (see, e.g. Doeblin, 1938 [14]).

The second condition guarantees the existence on non-zero values for the traffic vector $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$. It leads every customer to leave the system with non-zero probability. So the number of nodes each customer visited within the network is less than some geometrically distributed random variable and has a finite expectation.

Some notations for network process. If $\mathbf{X} = (X_t, t \geq 0)$ is a Markov process, the following notations will be used:
$Q = [q(\mathbf{e}, \mathbf{e}')]_{e,e' \in \mathbf{E}}$ – transition intensities;
π – stationary distribution;
infinitesimal generator:

$$\mathbf{Q}f(\mathbf{e}) = \sum_{\mathbf{e}' \in \mathbf{E}} (f(\mathbf{e}') - f(\mathbf{e}))q(\mathbf{e}, \mathbf{e}');$$

the scalar product for some functions f and g on $L_2(\mathbf{E}, \pi)$:

$$\langle f, g \rangle_{pi} = \sum_{\mathbf{e} \in \mathbf{E}} f(\mathbf{e})g(\mathbf{e})\pi(\mathbf{e}). \tag{6}$$

Spectral gap for $\mathbf{X}$ $[1, 6]$:

$$Gap(\mathbf{Q}) = \inf\{-\langle f, \mathbf{Q}f \rangle_\pi : \|f\|_2 = 1, \langle f, \mathbf{1} \rangle_\pi = 0\} \tag{7}$$

Theorem 1. *If* $\mathbf{X}$ *- the "dynamic routing network" process , with* $\mathbf{Q}$ *- infinitesimal generator (suppose bounded), minimal service and recovery intensities* $\mu > 0$ *and* $\beta > 0$*, and assumptions satisfy (1-2), then*

$$Gap(\mathbf{Q}) > 0$$

iff for each $i = 1, \ldots, m$*, the birth and death process with* λ_i*,* $\mu_i(n_i)$*,* α_i*,* β_i*, has* $Gap_i(\mathbf{Q}_i) > 0$*.*

Theorem 2. *If* $\mathbf{X}$ *- the "dynamic routing network" process with infinitesimal generator* $\mathbf{Q}$ *(suppose bounded), minimal service and recovery intensities* $\mu > 0$ *and* $\beta > 0$*,* $X(t)$ *satisfies the assumptions (1–2), then*

$$Gap(\mathbf{Q}) > 0$$

iff for each $i = 1, \ldots, m,$ *distribution* $\pi = (\pi_i), i \geq 0$ *is strongly light-tailed, i.e.*

$$\inf_k \frac{\pi_i(k)}{\sum\limits_{j>k} \pi_i(j)} > 0.$$

Theorem 3. *Let* $\mathbf{X}$ *- the "dynamic routing network" process with generator* $\mathbf{Q}$ *(given above) and the corresponding transition semigroup* P_t*, with minimal service and recovery intensities* $\mu > 0$ *and* $\beta > 0$*, and* $X(t)$ *satisfies the assumptions (1–2). Suppose that* G *satisfies the condition (2). If* π_i *is strongly light-tailed, for each* $i = 1, \cdots, m$*, then following statements are equivalent*

- *for all* $f \in L_2(\mathbf{E}, \pi)$

$$\|P_t f - \pi(f)\|_2 \leq e^{-Gap(\mathbf{Q})t} \|f - \pi(f)\|_2, t > 0, \tag{8}$$

- *for each* $\mathbf{e} \in \mathbf{E}$ *there exists* $C(\mathbf{e}) > 0$ *such that*

$$\|\delta_{\mathbf{e}} - \pi(f)\|_{TV} \leq C(\mathbf{e}) e^{-Gap(\mathbf{Q})t}, t > 0. \tag{9}$$

Proof. The proofs of these results are based on the standard techniques developed by T.Ligget and extended for queueing systems by other researchers [6, 12, 13]. There are two main results from Liggett[6]:

- Assume that $\mathbf{Z}$ is a birth and death process on $\mathbf{Z}_+$ with state independent birth rates $\mu > 0$, and possibly state dependent death rates $\mu(n) > 0$, and for all $i \geq 0$, and for some $b, c > 0$, we have

$$\sum_{j>i} \pi(i) \leq c\pi(i)\lambda$$

  and

$$\sum_{j>i} \pi(i) \leq b\pi(i).$$

  Then for the corresponding generator $\mathbf{Q}$

$$Gap(\mathbf{Q}) \geq \frac{(\sqrt{b+1} + \sqrt{b})^2}{c} \geq \frac{1}{2c(1+2b)}.$$

- Suppose that $\mathbf{X}$ is a Markov process with generator $\mathbf{Q}$ and stationary distribution π evolves on the product state space

$$\mathbf{E} = \mathbf{E}_1 \times \mathbf{E}_2 \times \ldots \mathbf{E}_m, \ m \geq 1,$$

having coordinates which are independent Markov processes such that i-th coordinate has generator $\mathbf{Q}_i$, denumerable state space $\mathbf{E}_i$ and invariant probability measure π_i.

Then π is the product measure of π_i and

$$Gap(\mathbf{Q}) = \inf_i Gap(\mathbf{Q}_i).$$

Consider generators $\hat{\mathbf{Q}}, \hat{\mathbf{Q}}_i, \hat{\mathbf{Q}}_0, i = 1, \ldots, m$ associated with independent processes $(\hat{\mathbf{X}}_{0t}, \hat{\mathbf{X}}_t), \hat{\mathbf{X}}_{0t}, \hat{\mathbf{X}}_t$ describing the evolution of each node of our "dynamic routing network" separately. The process $\hat{\mathbf{X}}_{0t}$ defined on the finite state space, the stationary distribution π_0 and $Gap(\hat{\mathbf{Q}}_0) > 0$. From Liggett's results we may conclude that

$$Gap(\hat{\mathbf{Q}}) = \min_i Gap(\hat{\mathbf{Q}}_i)$$

and so

$$Gap(\hat{\mathbf{Q}}) > 0.$$

3.3 Rate of convergence

The next important result relates to the convergence rate of the process $\mathbf{X}$ and is a consequence of Mu-Fa Chen results(see [5]).

Theorem 4. *Let* $\mathbf{X}$ *- the "dynamic routing network" process with generator* $\mathbf{Q}$ *(given above) and the corresponding transition semigroup* P_t*, then the classical variations formula holds*

$$Gap(Q) = \inf\{-\langle f, \mathbf{Q}\rangle_\pi : \pi(f) = 0, \|f\|_2 = 1\}$$

where

$$\langle f, g\rangle = \int f(x)g(x)\pi(dx),$$

$$\pi(f) = \int f(x)\pi(dx), Var_\pi(f) = \pi(f^2) - (\pi(f))^2,$$

and π *is invariant for* (P_t).

Let $f \in L_2(\mathbf{E}, \pi)$*, then*

$$C = Gap(\mathbf{Q})^{-1}$$

is optimal in Poincare inequality:

$$Var_\pi(f) \leq C - \langle f, \mathbf{Q}\rangle_\pi.$$

Proof. The above result for our network is a consequence from two theorems for Markov process from [5]. The first one is a Poincare inequality, the second one is the theorem about constant C existance:

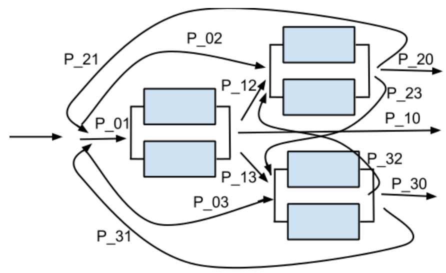

Fig. 4. Network with two-servers nodes

- Poincare inequality holds if and only if Markov process converges according to $L_2(\mathbf{E}, \pi)$-exponential convergence: for all

$$f \in L_2(\mathbf{E}, \pi)$$

$$\|P_t f - f\|_2^2 = Var(P_t f) \leq Var(f) exp(-2Gap(\mathbf{Q})t), \quad t > 0;$$

- Suppose that $\mathbf{E}$ is countable and P_t reversible.
  Then for all $f \in L_2(\mathbf{E}, \pi)$

$$\|P_t f - \pi(f)\|_2^2 = Var(P_t f) \leq Var(f) \exp(-2Gap(\mathbf{Q})t), \quad t > 0,$$

  iff for each $\mathbf{e} \in \mathbf{E}$ there exists $C(\mathbf{e}) > 0$ such that

$$\|\delta_e P_t - \pi\|_{tv} \leq C(\mathbf{e}) \exp(-2Gap(\mathbf{Q})t), t > 0.$$

So we can show that

$$C = Gap(\mathbf{Q})^{-1}$$

and from Poincare inequality for general Markov process:

$$Var_\pi(f) \leq C - \langle f, \mathbf{Q} \rangle_\pi.$$

4 The numerical example

We consider two numerical examples of network state probabilities calculation from [4]:

Example 1: The network consists of three nodes, each node is a system with two servers (see Figure 4).

Example 2: The network consists of two nodes, each node is a system with three servers (see Figure 5).

For the transition probabilities matrix (the same from [4]):

$$P_{ij} = \begin{pmatrix} 0.03 & 0.57 & 0.35 & 0.05 \\ 0.1 & 0.002 & 0.398 & 0.5 \\ 0.35 & 0.25 & 0.15 & 0.25 \\ 0.2 & 0.25 & 0.3 & 0.25, \end{pmatrix}$$

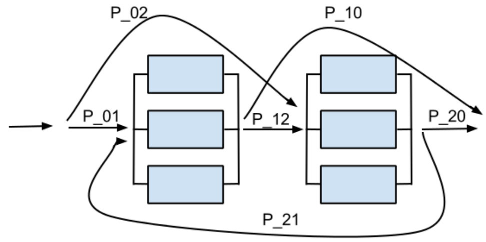

Fig. 5. Network with three-servers nodes

we suppose following failure rates α_i:

$$\alpha_0 = \alpha_1 = \alpha_2 = 3.0,$$

and recovery rates β_i:

$$\beta_0 = \beta_1 = \beta_2 = 6.0.$$

The obtained results for characteristics of this network are: - The probability of denial of service (the probability that all sites are occupied) = 0.0054; - Availability factor of the system (the system is completely free) = 0.29.

4.1 Further work and conclusion

The bounds derived above are valid only for light-tailed distribution. The convergence rates estimations for heavy-tailed distribution of service may be received via more complicated technique such as coupling method [17] and the generalized Lorden's inequality [18, 19], but only under large t. The future plan is to find polynomial bounds via this approach. This bounds will be valid only for large $t > T$, where T is computable.

References

1. Liggett, T. H. Interacting Particle Systems, 1999.
2. E. van Doorn. Many papers on the rate of convergence of birth-death processes.
3. Elmira Yu Kalimulina *Analysis of Unreliable Open Queueing Network with Dynamic Routing* DCCN, Springer, 2015.
4. Elmira Yu Kalimulina *Rate of convergence to stationary distribution for unreliable Jackson-type queueing network with dynamic routing.* DCCN, Springer, 2016.
5. Chen, M.-F. *Eigenvalues, Inequalities, and Ergodic Theory.* Springer, 2005.
6. Liggett, T. H. Exponential convergence of attractive reversible nearest particle systems *Ann. Prob*, 1989(17), 403–432.
7. Sauer, C. and Daduna, H. Availability formulae and performance measures for separable degradable networks (2003). Economic Quality Control 18(2), 165-–194. *Economic Quality Control* , 18(2), 165—194.

8. Lawler, G. F. and Sokal, A. D. Bounds on the L_2 Spectrum for Markov Chains and Markov Processes: A Generalization of Cheeger's Inequality. Trans. Amer. Math. Soc., 1988, 309(2), 557-–580.
9. E. van Doorn. Representations for the rate of convergence of birth-death processes. Theory Probab. Math. Statist, 2002, 65, 37—43.
10. Zeifman A.I.. Some Estimates of the Rate of Convergence for Birth and Death Processes, Journal of Applied Probability, 28 (1991), 268–277
11. Zeifman A.I.. Van Doorn, E.A., Zeifman, A.I. On the speed of convergence to stationarity of the Erlang loss system. Queueing Syst 63, 241 (2009).
12. Lorek P., Szekli R. Computable bounds on the spectral gap for unreliable Jackson networks. Adv. in Appl. Probab, Volume 47, Number 2 (2015), 402-424.
13. Lorek P. The exact asymptotic for the stationary distribution of some unreliable systems. arXiv:1102.4707 [math.PR].
14. Doeblin W. Mathématique de l'Union Interbalkanique, 1938
15. Albert R., Barabasi A. Statistical mechanics of complex networks. Reviews of Modern Physics. 74 (1): 47–97. arXiv:cond-mat/0106096.
16. Dorogovtsev, S. N. and Mendes, J. F. F., Evolution of Networks: From Biological Nets to the Internet and WWW (Physics), Oxford University Press, Inc.,New York, NY, USA, 2003
17. Zverkina G. Stationary coupling method for renewal process in continuous time (application to strong bounds for the convergence rate of the distribution of the regenerative process), arXiv preprint arXiv:1704.04808, 2017
18. Zverkina G. Lorden's inequality and coupling method for backward renewal process, arXiv preprint arXiv:1706.00922, 2017
19. Kalimulina E.Y., Zverkina G.A. On some generalization of Lorden's inequality for renewal processes, arXiv preprint arXiv:1910.03381