A Mean-field Approach to Some Internet-like Random Networks

Ilkka Norros
VTT Technical Research Centre of Finland

***Abstract*—A conditionally Poissonian power-law random graph with infinite degree variance is considered as a random network model. A method for elegant analytical computation of accurate approximations for various network characteristics is introduced, based on slight redefinition of the model in terms of non-homogeneous Poisson point processes and on the replacement of certain random variables by their expectations. The applications include characterization of the 'top clique' around the node of highest capacity, density of nodes falling outside of the giant component of the random graph, availability of disjoint paths and the distribution of traffic in the network, assuming a traffic matrix following a gravity rule.**

I. Introduction

This paper has its origin in the need of computing various characteristics of a power-law random graph effectively with reasonable accuracy and setting the focus at mean behaviour rather than on rare events. Such graphs have become popular as abstract structures that are 'Internet-like' in some central features (and unlike in others; see, e.g., [3], [12], [2], [1], [13], [11]). Their remarkable feature is that their main, essentially single assumed characteristic is a power-law distribution of degrees (that is, the number of neighbors) of nodes. The most interesting case of power-law graphs is that the degrees follow a distribution with infinite variance. For brevity we abbreviate 'infinite-variance power-law random graph' as IVPLRG.

These models are very useful for understanding large-scale features of large networks. For example, routing in PLRGs has been studied as a tool for looking for new solutions to Internet routing [5], [6], [7]. The approach presented here was recently invented in the work on the last mentioned paper. As a totally new possibility offered by the technique, first steps are made to modelling of large traffic flows in this kind of networks. The basic ideas may be transferable to other networks as well.

The paper is organized as follows. Properties of a conditionally Poissonian IVPLRG model are summarized in Section II (condensing [11]). The basics of the new approach are presented in Section III. The main part of the paper is Section IV, where the approach is applied to several problems. New insights are obtained in the distribution of the top clique, availability of alternate paths, and the distribution of traffic. The novel approach to large-scale traffic flow modelling might be the most interesting contribution of this paper.

II. The conditionally Poissonian IVPLRG model

Consider graphs with a set V of nodes and a set E of non-directed links. The number of nodes is denoted by $N := |V|$. Two nodes connected by a link are called neighbors. The degree of a node is the number of its neighbors. The conditionally Poissonian random

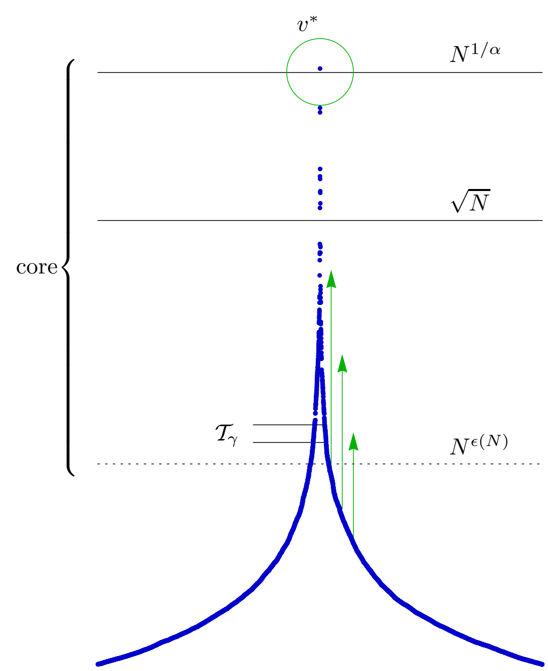

Fig. 1. The soft hierarchy structure of the core of an IVPLRG. Capacities are drawn from distribution (1) and plotted on logarithmic scale, highest in the center. The upward arrows indicate how mighty neighbors a node typically has. Nodes bigger than $\sqrt{N}$ form almost a clique. Horizontally cut 'tiers' are denoted by $\mathcal{T}_\gamma$.

graph [10] is constructed as follows. First, each node is given a 'capacity' Λ_i with a power-tail distribution. For simplicity, assume that the capacities obey the Pareto distribution

$$\mathbb{P}(\Lambda > x) = x^{-\alpha}, \quad x \in [1, \infty), \quad \alpha \in (1, 2). \tag{1}$$

The choice of the parameter α is crucial: the graph has entirely different properties if α is taken outside of $(1, 2)$, where Λ has finite mean but infinite variance. In the second step, each pair of nodes $\{i, j\}$ is independently joined by E_{ij} links, where E_{ij} is a Poisson random variable with mean

$$\frac{\Lambda_i \Lambda_j}{L_N}, \quad L_N := \sum \Lambda_k. \tag{2}$$

Loops and parallel links are accepted but mostly neglected. Note that (2) is a 'gravity rule': the expected number of links between nodes i and j is proportional to both capacities Λ_i and Λ_j. Conditioned on the capacities, the degree of node i is a random variable with distribution Poisson(Λ_i). By the addition rule of Poissonian random variables, the number of edges joining two disjoint sets of nodes is again Poissonian. From now on, we use the term IVPLRG for this particular model. Rigorous proofs of the claims below can be found in [10], [8], and a more intuitive discussion in [11]. Figure 1 helps with the presentation.

A. The core network and its soft hierarchy

The core network: Choose $\epsilon(N)$ satisfying the following three conditions as $N \to \infty$:

$$\epsilon(N) \to 0, \quad N^{\epsilon(N)} \to \infty, \quad \frac{N^{\epsilon(N)}}{\log \log N} \to 0. \tag{3}$$

We define the *core* of the IVPLRG as the set of nodes whose capacity exceeds $N^{\epsilon(N)}$. This formal definition makes of course sense only in the asymptotical behavior of the graph as $N \to \infty$. Intuitively, the core consists of 'big' nodes whose capacity 'is a power of N'. The node v^* with largest capacity satisfies $\Lambda_{v^*} \approx N^{1/\alpha}$.

Most capacity is at the bottom: Let $\gamma \in (\epsilon(N), 1/\alpha)$. The aggregated capacity of all nodes i satisfying $\Lambda_i > N^\gamma$ is proportional to $N^{1-(\alpha-1)\gamma}$ for large N. Thus, most of the capacity of any top section of the core resides near its lower boundary.

Edge density increases from low up to a clique: Consider thin horizontal sections, 'tiers' of the core, denoting

$$\mathcal{T}_\gamma = \left\{ i \in V : \Lambda_i \in (N^\gamma, N^{\gamma+\epsilon(N)}) \right\}.$$

When γ increases from $\epsilon(N)$ upwards, the internal edge density of $\mathcal{T}_\gamma$ grows from fairly low up to 1, which is obtained when γ passes the value $\frac{1}{2}$ and corresponds to a clique (a totally connected subgraph). It was recently proved in [4] that the largest clique in the IVPLRG essentially consists of nodes with capacity higher than $c\sqrt{N \log N}$, with an explicitly given constant c. Even the lower tiers of the core are almost connected internally.

The hierarchy and its height: Consider a core node v with capacity N^γ. The largest capacity found among v's neighbors is close to $\min(N^{\gamma/(\alpha-1)}, \Lambda_{v^*})$. Thus, the core network forms a 'soft' hierarchy in the sense that almost every core node has a neighbor with this much higher capacity. Iterating this observation, one obtains (see [11]) that the distance from v to v^* is with high probability at most

$$k^* := \left\lceil \frac{\log \log N}{-\log(\alpha - 1)} \right\rceil.$$

Thus, any two nodes of the core are with high probability connected by a path containing at most $2k^*$ hops and having an up-down profile in terms of capacities of the intermediate nodes.

The core is found quickly: Although the size of the core and its aggregated capacity are both asymptotically negligible parts of the whole, the latter number is sufficiently high to make the core attractive enough that a randomly chosen node is, with high probability, either connected to the core with still less than k^* hops, or not connected to it at all. Both alternatives have positive probabilities. The connected component containing the core is called the giant component.

The core is robust: If all nodes with capacity higher than N^γ are deleted, with a fixed $\gamma > 0$, the relative size of the giant component remains asymptotically unchanged, and the shortest paths are increased by a constant that does not depend on N.

III. A MEAN-FIELD APPROACH OF THE CONDITIONALLY POISSONIAN IVPLRG

The properties discussed in Section II hold asymptotically almost surely (a.a.s.) as $N \to \infty$, and they were proven with rather sophisticated arguments. In order to study and visualize various practically interesting relationships within the IVPLRG model effectively with finite N, a 'mean-field' approach was introduced in [7]. Its main ideas are the following:

- instead of a set of N nodes with capacities distributed i.i.d. like Λ, assume a non-homogeneous Poisson process on $\mathbb{R}_+$ with total mass N

- everything will be done just in terms of the intensities of Poisson processes — the nodes will not need to be realised as points except potentially
- aggregated capacities of subsets will be computed according to intensities, not by realised points — in this sense this is a mean-field approximation
- results can be visualized by plotting the various densities in appropriate scaling.

Let Λ have distribution (1). Let us start with a scaling needed for visualisation independent of N. Write

$$X = X^{(N)} = \frac{\log \Lambda}{\log N},$$

so that

$$\mathbb{P}(X > x) = N^{-\alpha x}, \quad x > 0.$$

The density of this probability measure is

$$f_0^{(1)}(x) = \alpha \log(N) N^{-\alpha x}.$$

If $f_0^{(1)}$ is used as the intensity of a Poisson point process, the realisation is a Poisson(1)-distributed number of points thrown on $\mathbb{R}_+$ independently according to the distribution $f_0^{(1)}$. A Poisson process with intensity

$$f_0 := N f_0^{(1)}$$

yields Poisson(N) similarly random points. The first mean-field element in our approach is to replace the denominator L_N in (2) by its mean value

$$\mathbb{E} L_N = N \mathbb{E} \Lambda = \frac{\alpha N}{\alpha - 1}.$$

Thus, we assume that two nodes with capacities N^x and N^y are connected with a link with probability

$$1 - \exp\left(-\frac{\alpha - 1}{\alpha} N^{x+y-1} \right)$$

For the sequel, let us call point process densities briefly as 'populations'. The examples are run with $\alpha = 1.5$.

IV. Applications

A. Top element and top clique

Let us first consider the modelling of the largest node v^*. The density of the largest among n independent copies of X is

$$\tilde{p}^{(n)}(x) = \alpha \log(n) n^{1-\alpha x} (1 - n^{-\alpha x})^{n-1}. \qquad (4)$$

In the point process model, the density of the rightmost point is obviously $p_{v^*}^{\text{Poisson}} = \mathbb{E} \tilde{p}^{(\text{Poisson}(N))}$. Now consider the following heuristic reasoning, which also illustrates the approach of this paper in general. The intensity at which there exist points larger than x is $f_0(y) 1_{\{y > x\}}$. The number of such points is Poisson distributed, so the probability that there are no such points is $\exp(-\int_x^\infty f_0(y)\,\mathrm{d}y)$. Heuristically, the intensity of points with the property that there is no larger point is obtained by thinning f_0 with this probability:

$$p_{v^*}^{\text{mf}}(x) = f_0(x) \exp\left(-\int_x^\infty f_0(y)\,\mathrm{d}y \right). \qquad (5)$$

The total mass of density (5) is exactly one, and, by the equality $e^a = \lim_{n\to\infty}(1 + a/n)^n$, (5) is a highly accurate approximation of (4) for large N. On the other hand, note that in many contexts it would be misleading to handle $p_{v^*}^{\text{mf}}$ as a point process intensity, since the number of maximal points is exactly one, not a Poisson(1) random number. We noted in section II that Λ_{v^*} be approximately $N^{1/\alpha}$. This 'typical' value is in fact close to the median, whereas the mean is considerably higher. In this paper we work mostly with mean values. If our mean-field counterpart to the largest node v^* is the intensity measure $p_{v^*}^{\text{mf}}$, the counterpart of the capacity of v^* is the integral

$$\int_0^\infty N^x p_{v^*}^{\text{mf}}(x)\,\mathrm{d}x. \qquad (6)$$

In similar fashion we can approximate the density of the k'th order statistic by

$$p_{X_{(k)}}^{\text{mf}}(x) = f_0(x) \frac{\theta(x)^k}{k!} e^{-\theta(x)}, \quad \theta(x) = \int_x^\infty f_0(y)\,\mathrm{d}y.$$

Our next question concerns the top clique. The largest clique containing v^* is not always uniquely determined, since there may be several cliques of maximal size. However, a related object can be defined uniquely:

Definition. The *quasi top clique* C_{qt} consists of nodes v having a link to every other node with capacity higher than or equal to Λ_v.

A node belonging to the quasi top clique does not necessarily belong to a maximal clique $C_{\max}$ since it may miss a link to a $C_{\max}$ member smaller than itself. On the other hand, in that case the latter node is not a member of the quasi top clique. However, it was proved in [4] that $|C_{\text{qt}}|/|C_{\max}| \to 1$ as $N \to \infty$.

Now, the distribution of quasi top clique nodes can be approximated by same type of reasoning as the derivation of (5). If $\Lambda_v = N^x$, nodes with still higher capacity exist with intensity $f_0(y) 1_{\{y > x\}}$, and the probability that such a node has no link to v is $\exp(-((\alpha-1)/\alpha) N^{x+y-1})$. Thus, the number of nodes with higher capacity than N^x but no link to v has Poisson distribution with parameter

$$u(x) = \int_x^\infty f_0(y) \exp\left(-\frac{\alpha - 1}{\alpha} N^{x+y-1} \right)\mathrm{d}y,$$

and the probability that there are no such nodes is $e^{-u(x)}$. This in turn suggests that the quasi top clique could be (approximately) modelled by a Poisson point process with intensity

$$p_{\mathrm{qt}}(x) := f_0(x)e^{-u(x)}. \tag{7}$$

In contrast to the case of v^*, the size of C_{qt} is truly random and, by the above reasoning, probably close to a Poisson random variable with parameter $\int_0^\infty p_{\mathrm{qt}}(x)\,\mathrm{d}x$ (of course not exactly, i.e. it never takes value zero). Writing $u(x)$ in terms of an incomplete Gamma function and computing the integral numerically, we obtain the following approximations of expected top clique sizes:

N	10^3	10^4	10^5	10^6	10^7	10^8	10^9
	3.1	4.4	6.4	9.6	14.7	23.0	36.5

Although the size of the top clique (from now on, we mostly suppress the word 'quasi') grows with N according to this table, in terms of capacities the share of v^* within the top clique decreases slowly — for $N = 10^9$ it is still as high as 27%.

B. Neighborhood shells

Let us next study the subsequent neighborhoods of a population g_0 as follows. The complement of this population is

$$f_1 = f_0 - g_0.$$

Let

$$c_0 = \int_0^\infty N^x g_0(x)\,\mathrm{d}x.$$

denote the mean aggregated capacity of population g_0. The expected part of population f_1 that has a link to population g_0 is, in a mean-field approximation,

$$g_1(x) = \left(1 - e^{-\frac{\alpha-1}{\alpha}c_0 N^{x-1}}\right) f_1(x).$$

Note that thanks to the Poissonian rule, we can work here with the aggregated capacity alone so that the thinning in the generation of g_1 is made just by a multiplication (in point process terms: as independent thinning at every point). The mean-field approximation lies in using a mean value c_0 instead of a random capacity obtained from a point process realization with intensity g_0.

Approximate intensities of subsequent neighborhood shells are now obtained recursively:

$$\begin{aligned} f_{k+1} &= f_k - g_k \\ g_{k+1}(x) &= \left(1 - e^{-\frac{\alpha-1}{\alpha}c_k N^{x-1}}\right) f_{k+1}(x) \\ c_{k+1} &= \int_0^\infty N^x g_{k+1}(x)\,\mathrm{d}x. \end{aligned} \tag{8}$$

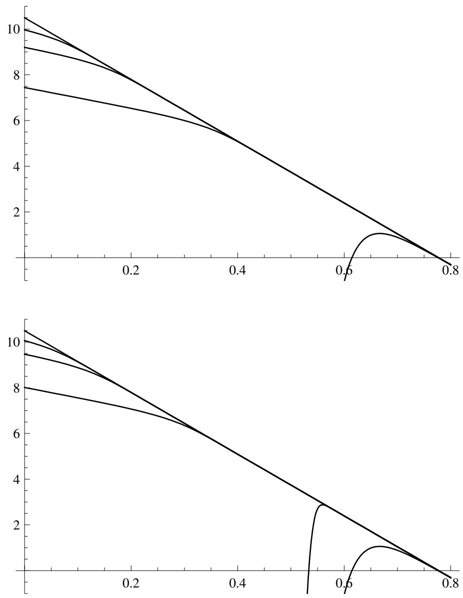

Fig. 2. Plots ($\log_{10}$-linear) of the intensity f_0 and the intensities of cumulative neighborhoods of the top. Above: $g_0 = p_{v^*}^{\mathrm{mf}}$. Below: $g_0 = p_{\mathrm{qt}}$ (for comparison, $p_{v^*}^{\mathrm{mf}}$ is shown here also). $N = 10^9$.

The intensity of nodes with capacity N^x in the k'th neighborhood shell around g_0 is given by g_k, the mean number of nodes in that shell by $d_k := \int_0^\infty g_k(x)\,\mathrm{d}x$, and their mean aggregated capacity by c_k.

Figure 2 shows the intensities of the cumulative neighborhoods g_0, $g_0 + g_1$,... starting from either v^* alone or from the top clique. The difference is non-dramatic, showing that the top clique can be considered as a kind of expanded counterpart of v^* rather than a separate layer surrounding v^* in the core hierarchy. Figure 3 provides same information as the first case in the previous figure, but now showing the proportion of the kth neighborhood of v^* among the total population f_0, $k = 0, 2, 1, 3$. The choice $N = 10^9$ was taken mainly for demonstrating the scalability of the mean-field approach — the pictures are otherwise qualitatively similar for any relevant N, but the populations are more centered for large N. To give an impression, a second picture with $N = 10^4$ is shown too.

C. Complement of the giant component

The asymptotic relative size of the giant component was given in [10] and it can in our case be computed as

$$1 - e^{\phi-1} + \frac{\phi(1-\phi)}{\alpha - 1},$$

where $\phi = \phi(\alpha)$ is determined by the conditions

$$\phi = (\alpha-1)(1-\phi)^{\alpha-1}\Gamma(1-\alpha, 1-\phi), \quad \phi \in (0,1)$$

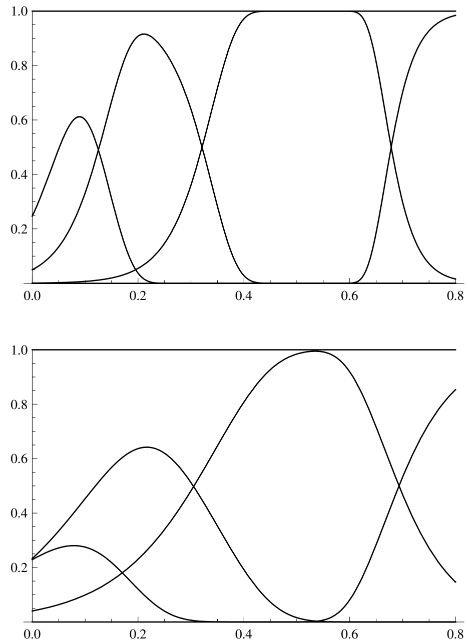

Fig. 3. Proportional sizes of neighborhood shell intensities around v^* w.r.t. the total population density. Top: $N = 10^9$. Bottom: $N = 10^4$.

with $\Gamma(a, z) = \int_z^\infty u^{a-1} e^{-u} \, \mathrm{d}u$. The number ϕ has the meaning of being the extinction probability of a Galton-Watson branching process with the number of children obeying a conditionally Poissonian distribution Poisson(Γ), where Γ has the distribution

$$\mathbb{P}(\Gamma \in \mathrm{d}\lambda) = \frac{\lambda \mathbb{P}(\Lambda \in \mathrm{d}\lambda)}{\mathbb{E}\Lambda}.$$

The probability that a node with capacity x remains outside the giant component equals the probability that a Poisson(N^x) distributed number of such branching processes all die and can be computed as

$$\sum_{k=0}^{\infty} \frac{N^{kx}}{k!} e^{-N^x} \phi^k = e^{-(1-\phi)N^x}.$$

The intensity of such nodes is then obtained as $f_0(x) e^{-(1-\phi)N^x}$, and their aggregated capacity turns out to be $N\phi\alpha/(\alpha - 1)$. Thus, the asymptotic proportion of total capacity falling outside the giant component is simply ϕ.

As a credibility check of the mean-field method, it is worth of noting that the size of the giant component (78.0% with $\alpha = 1.5$) is obtained with high accuracy as $1 - \int_0^\infty f_{k+1}(x) \, \mathrm{d}x$ when g_k has become small again, which happens in our examples with k between 5 and 7.

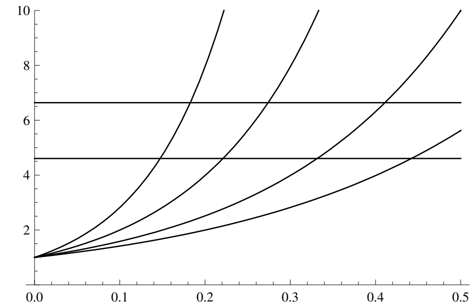

Fig. 4. The mean number of links that a node possesses to nodes with higher capacity than itself. The two horizontal lines show the levels where the probability of exceeding one (resp. two) achieves the value 0.99. Curves from below: $N = 10^3, 10^4, 10^6, 10^9$.

D. Core, soft hierarchy, and existence of disjoint paths

A practical idea of the core network, not involving the asymptotic quantity $\epsilon(N)$ of Section II, would be the set of nodes with so high capacity that the probability of having links ‘up’ is high. This is the set where the ‘soft hierarchy’ works. Let us see where the lower boundary of the core in this sense lies (of course that ‘boundary’ is ‘soft’ like everything in this model). The mean aggregated capacity of all nodes with capacity higher than N^x is

$$V(x) := \frac{\alpha}{\alpha - 1} N^{1-(\alpha-1)x}.$$

Thus, the number of ‘upward’ links from a node with capacity N^x is a Poisson random variable with mean

$$\frac{\alpha - 1}{\alpha} N^{x-1} V(x) = N^{(2-\alpha)x}.$$

This function is plotted in Figure 4 with different values of N. The lower boundary of the core can be thought as the abscissa of the intersection of the above curve with a horizontal line whose height (taken from the Poisson distribution) corresponds to the required probability of the existence of up-links. The lower boundary of the core decreases with increasing N, in harmony with the asymptotic theory summarized in Section II.

Our next question concerns the robustness of the IV-PLRG in another sense than that mentioned in Section II: on what probability does a node possess several disjoint paths to the top clique (or to v^*)? This problem was met in our recent work on routing in IVPLRGs [7], where the availability of multiple paths to the top clique was used to mitigate the inflexibility of a routing scheme. From Figure 4 it is rather obvious that multiple paths exist in the core with more or less high probability, but we leave for further studies the task of finding explicit lower bounds for the probability that two nodes with respective capacities N^x and N^y are connected

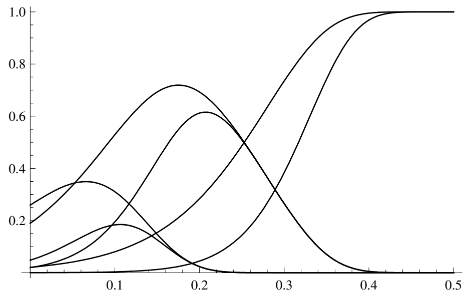

Fig. 5. The proportional intensities of the neighborhood shells 1,2,3 around the top clique together with their subpopulation possessing at least two links to the next shell. Curves coinciding at their right end correspond to same shell. $N = 10^6$.

with two disjoint paths. On the other hand, it is also obvious that the first hops of a small node's path to the core are unique with a probability bounded away from zero.

Instead, let us consider the existence of multiple *minimal* paths to the top clique (in a similar fashion, we could study paths to v^*). Denote by $g_0, g_1, \ldots$ the neighborhood shell populations of subsection IV-B, where $g_0 = p_{\rm qt}$, and by $f_1, f_2, \ldots$ the corresponding 'outsider' populations. Using the formula of the probability that a Poisson random variable is at least 2, we get that the intensity of f_{k+1}-nodes that have at least two links to the set of g_k-nodes is

$$g_{k+1}^{(2)}(x) \quad := \quad f_{k+1}(x) \times \left(1 - \left(1 + \frac{\alpha-1}{\alpha} N^{x-1} c_k\right) e^{-\frac{\alpha-1}{\alpha} N^{x-1} c_k}\right),$$

where, as before, $c_k = \int_0^\infty N^x g_k(x)\,\mathrm{d}x$. Figure 5 presents the intensities g_k, $k = 1, 2, 3$, of the three first shells around the top clique together with the intensities of their subpopulations possessing at least two links to the previous shell. It is seen that for each k there is a rather sharp capacity threshold above which the existence of multiple links to the next 'tier' has high probability. However, below that threshold there is a considerable probability to have only one link to the next tier. Thus, although the core is very well connected so that alternative routes do exist, the probability that the *minimal* route be unique is surprisingly large even for huge N.

In the computations above we did not require that the two upward links lead to different nodes. Let us consider the cost of this simplification. If the probability that a random link points to node i is p_i, then the probability of a 'double-hit', i.e. that two independent, similarly chosen random links point to the same node, is $\sum p_i^2$. If the link destinations are points of a Poisson point process X_k with finite intensity measure ν on $\mathbb{R}_+$ and the probabilities are proportional to N^{X_k}, then the double-hit probability is

$$\mathbb{E}\left\{\sum_{k=1}^{K}\left(\frac{N^{X_k}}{\sum_{j=1}^{K} N^{X_j}}\right)^2\right\},$$

where K is the total number of points. A heuristic 'default' approximation for this number would be

$$\frac{\int_0^\infty N^{2x}\nu(\,\mathrm{d}x)}{\left(\int_0^\infty N^{x}\nu(\,\mathrm{d}x)\right)^2}. \tag{9}$$

This is, however, unusable for the top clique intensity, since the numerator diverges with the choice $\nu(\,\mathrm{d}x) = p_{\rm qt}(x)\,\mathrm{d}x$. A way to obtain a computable although inaccurate estimate for top clique double-hit probability is to replace the highest order statistics as well as the top clique size by their expectations, giving the expression

$$\sum_{k=1}^{\mathbb{E}|C_{\rm qt}|}\left(\frac{\mathbb{E}N^{X_{(k)}}}{\sum_{j=1}^{\mathbb{E}|C_{\rm qt}|}\mathbb{E}N^{X_{(k)}}}\right)^2,$$

and applying then the mean-field approximations considered earlier in this paper. This computation yields the following rough estimates for double-hits to top clique:

N	10^3	10^4	10^5	10^6	10^7	10^8	10^9
	0.36	0.28	0.19	0.13	0.09	0.06	0.04

Whatever the accuracy of these estimates, they show that this probability is large for moderate N and non-negligible even for very large N. Thus, finding two different top clique nodes often requires more than two random throws into the top clique.

On the other hand, the estimate (9) should be good for the subsequent tiers $g_1, g_2, \ldots$, where the populations are large and extremely high capacities excluded. Application of this formula tells that the probability of a double-hit to g_1 is only about 1% already with $N = 1000$ and negligible for large N.

E. Traffic distribution

Our last and most novel application concerns the distribution of traffic in the IVPLRG. The question is, of course, meaningless without a specification of (i) a traffic matrix and (ii) a routing scheme. The first and maybe only study in this direction so far was reported in our paper [9], where the traffic distribution within the IVPLRG core was simulated with uniform traffic matrix and a random 'natural' routing, meaning shortest path routing with link weights $1 + \epsilon_{ij}$ for each link (i, j), with small random numbers ϵ_{ij} making the shortest paths unique. Now we consider the whole network and make the more realistic assumption that the traffic matrix follows a gravity rule. The routing is again along shortest paths, but instead of specifying unique and consistent routes, which would be impossible in general terms, we make the idealistic assumption that the traffic flows in parallel along all links connecting two neigborhood shells of the source.

Assume now that each node with capacity N^x sends to each node with capacity N^y traffic at rate

$$a_0(x,y) = N^x N^y. \tag{10}$$

This implies the rather natural rule that traffic demands are proportional to node capacities. Since the traffic unit does not matter, no constant multiplier is needed in (10). However, we want to include only traffic that is sent to nodes belonging to the giant component. According to subsection IV-C and thanks to the multiplicative form of the gravity rule, this can mostly be taken into account simply by a multiplication by $1-\phi$. In particular, the total expected output rate of a node with capacity N^x (an 'x-node') is

$$a_0(x) = \frac{(1-\phi)\alpha}{\alpha-1} N^{x+1},$$

and the aggregated source traffic in the network is

$$\frac{(1-\phi)^2\alpha^2}{(\alpha-1)^2} N^2.$$

We now model the flow of traffic originating from a single x-node as follows. The traffic is distributed to the subsequent neighborhood shells around our node, and the approach of subsection IV-B can be applied. At each shell, nodes absorb the amount of traffic intended to them, and forward the rest to the next shell. Since no routes are fixed (and cannot be fixed in this type of modelling), we assume that the forwarded traffic is distributed according to the node capacities.

Fix a node v in the giant component with $\Lambda_v = N^x$ and consider the availability of its traffic ('v-traffic') in its neighborhood shells. In phase 0, the traffic is available only at v itself, and its amount is $a_0(x)$. In the recursion (8) we now neglect the 'mass' of v as a point but recognize its capacity, choosing $g_0(x,y) \equiv 0$, $c_0 = N^x$ and, restricting to the giant component,

$$f_1(x,y) = f_0(y)\left(1 - e^{(1-\phi)N^y}\right).$$

The neighborhood shells $g_1(x,\cdot), g_2(x,\cdot), \ldots$ are now defined according to the recursion (8).

Let us denote the density of traffic absorbed in shell k as

$$b_k(x,y) \quad = \quad g_k(x,y)N^{x+y}, \quad k \geq 1.$$

Denote the total amount of v-traffic arriving to shell k by $a_k(x)$ (we don't mention v, because the quantity depends only on x). The total amount of traffic absorbed at shell k is denoted as $b_k(x) = \int_0^\infty b_k(x,y)\,\mathrm{d}y$. Obviously

$$a_k(x) = \sum_{j=k}^{\infty} b_j(x), \tag{11}$$

since the neighborhood shells cover the whole giant component.

Now, let us assume that the amount $a_k(x)$ be distributed proportionally to the capacities on shell k. That is, we model the density of v-traffic arriving to shell k by

$$a_k(x,y) = \frac{g_k(x,y)N^y}{\int_0^\infty g_k(x,z)N^z\,\mathrm{d}z} \cdot a_k(x). \tag{12}$$

The density of traffic that is not absorbed in shell k and thus remains to be forwarded is then obtained as $a_k(x,y) - b_k(x,y)$. Note that the above definitions are feasible in the sense that $a_k(x,y) - b_k(x,y) \geq 0$. Indeed, since arriving and absorbed traffic are distributed in same proportions within the shell, it is sufficient to check that $a_k(x) \geq b_k(x)$ for every k, which follows from (11). We leave, however, for further study the problem of feasibility of (12) in the more detailed sense of distributing traffic *end-to-end* with this allocation rule.

As an example, Figure 6 shows the volume of traffic originating from a node with smallest possible capacity 1 when it arrives to subsequent hops, distributed proportionally to node capacity according to (12). Note that the distributions describe the mean behaviour, not a typical behaviour.

The form of $a_1(0,y)$ reveals a non-pleasant feature of our model: in some contexts like here the possibility of super-large node capacity values dominates the mean behavior of a phenomenon. The gravity rule sets a lot of traffic mass to extremely large x-values that hardly ever appear in realizations. Nevertheless, it follows from the basic power relations, not from the possibility of over-sized capacity values, that although the population density of the first neighborhood shell is decreasing, its absorbed traffic density is indeed increasing until very high x values. This observation is quite surprising, and may be exaggerated by our continuous modelling approach.

V. Conclusions and remarks

The most interesting novelties (at least for the author) obtained in this study were (i) computation of the size of the quasi top clique, (ii) that even inside the core the shortest paths are often partly unique, (iii) that a

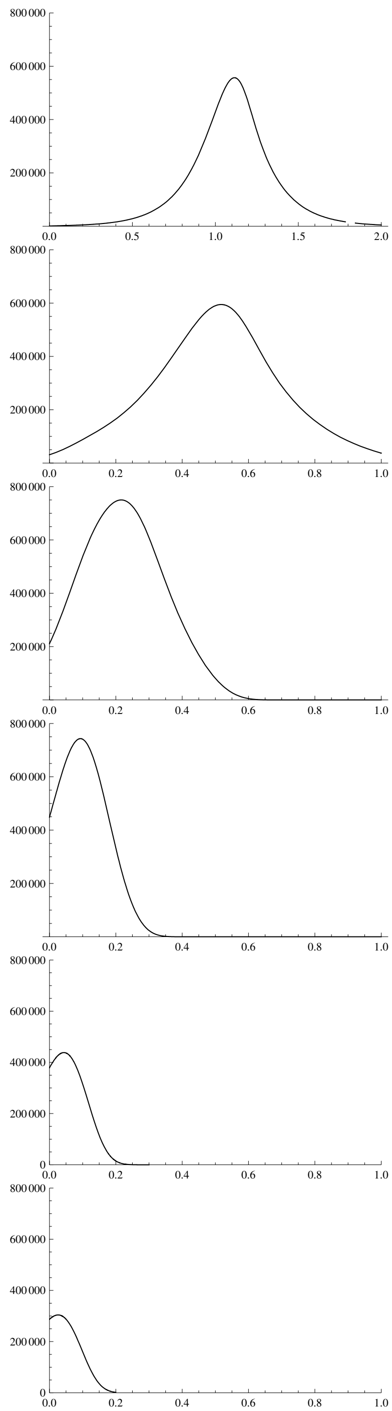

Fig. 6. The transfer of traffic from a single node with capacity 1 to everywhere else in the giant component. The arriving traffic density per node capacity is shown for the six first hops. Note the different x-range of the first plot. $N = 10^5$.

global picture of traffic distribution can be obtained in this way, and (iv) that a very large part of traffic moves to the core already in the first hop. The results on traffic had the most preliminary character and may suffer from artefacts of the continuous modelling approach. Further work is needed to understand the impact of super-large Λ_{v^*} values to the results and the possible need of modifications to the model. Finally, we wish to consider the applicability of this 'intensity approach' to other network and traffic models.

Acknowledgement. This work was supported by the GIGA program of TEKES. The author also gratefully acknowledges the hospitality of the Mittag-Leffler Institute, where this work was carried out.

REFERENCES

[1] A.-L. Barabási. *Linked: The New Science of Networks*. Perseus Publishing, 2002.

[2] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113, 2003.

[3] C. Faloutsos, P. Faloutsos, and M. Faloutsos. On power-law relationships of the Internet topology. *Computer Communications Rev.*, 29:251–262, 1999.

[4] S. Janson, T. Łuczak, and I. Norros. Large cliques in a power-law random graph. 2009. Submitted for publication. Preprint: arXiv 0905.0561.

[5] D. Krioukov, K. Fall, and X. Yang. Compact routing on Internet-like graphs. In *Infocom*, 2004.

[6] D. Krioukov, kc claffy, K. Fall, and A. Brady. On compact routing for the Internet. *ACM SIGCOMM Computer Communication Review (CCR)*, 37(3), 2007.

[7] I. Norros. Powernet: Compact routing on Internet-like random networks. In *NGI 2009*, Aveiro, July 2009.

[8] I. Norros and H. Reittu. On the robustness of power-law random graphs in the finite mean, infinite variance region. Accepted for publication in Internet Mathematics. Preprint: arXiv: 0801.1079, 2008.

[9] I. Norros and H. Reittu. Architectural features of the power-law random graph model of Internet: Notes on soft hierarchy, vulnerability and multicasting. In *ITC-18*, Berlin, September 2003.

[10] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Adv. Appl. Prob.*, 38:59–75, 2006.

[11] I. Norros and H. Reittu. Network models with a 'soft hierarchy': a random graph construction with loglog scalability. *IEEE Network*, 22(2):40–46, 2008.

[12] H. Reittu and I. Norros. On the effect of very large nodes in Internet graphs. In *Proc. GLOBECOM*, Taipei, Taiwan, 2002.

[13] H. Tangmunarunkit, R. Govindan, S. Jamin, and S. Shenker. Network topology generators: Degree-based vs. structural. In *Proc. SIGCOMM'02*, 2002.