Semantic structure and finite-size saturation in scale-free dependency networks of free software

Rajiv Nair* and G. Nagarjuna†
Homi Bhabha Centre for Science Education, Tata Institute of Fundamental Research, V. N. Purav Marg, Mankhurd, Mumbai 400088, India

Arnab K. Ray‡
Department of Physics, Jaypee University of Engineering and Technology,
A-B Road, Raghogarh, Guna 473226, Madhya Pradesh, India

A continuum model has been proposed to fit the data pertaining to the directed networks in free and open-source software. While the degree distributions of links in both the in-directed and out-directed dependency networks follow Zipf's law for the intermediate nodes, the most richly linked nodes, as well as the most poorly linked nodes, deviate from this trend and exhibit finite-size effects. The finite-size parameters make a quantitative distinction between the in-directed and out-directed networks. Dynamic evolution of free software releases shows that the finite-size properties of the in-directed and out-directed networks are opposite in nature. For the out-degree distribution, the initial condition for a dynamic evolution also corresponds to the limiting count of rich nodes that the mature out-directed network can have. The number of nodes contributing out-directed links grows with each passing generation of software release, but this growth ultimately saturates towards a finite value due to the finiteness of semantic possibilities in the network.

PACS numbers: 89.75.Hc, 89.75.Fb, 05.45.-a, 89.20.Ff
Keywords: Networks; Structures and organization in complex systems; Nonlinear dynamics; Computer science

I. INTRODUCTION

Scale-free distributions in complex networks have been very well studied by now [1–6]. The ubiquity of scale-free properties is quite remarkable, and spans across vastly diverse domains like (to name a few) the *World Wide Web* [1, 7] and the Internet [1], the social, ecological, biological and linguistic networks [5], income and wealth distributions [8, 9], trade and business networks [10], highway networks [11] and syntactic and semantic networks [12–14].

It should hardly occasion any surprise, therefore, that further developments have led to the discovery of scale-free features in electronic circuits [15] and in the architecture of computer software as well [16]. A recent work has shown that the structure of object-oriented software is a heterogeneous network characterised by a power-law distribution [17], and continuing on this theme, software fragility has been explained naturally on the basis of scale-free networks in software [18]. Keeping more closely with the objectives of the present paper, a previous work on complex networks in software engineering [19] has found evidence of power-law features in the inter-package dependency networks in free and open-source software (*FOSS*), while there have also been indications that modifications in this type of a software network follow a power-law decay as a function of time [20, 21].

It is a matter of common knowledge that when it comes to installing a software package from the *Debian GNU/Linux* repository, many other packages — the "dependencies" — are also called for as prerequisites. This leads to a dependency-based network among all the packages, and each of these packages may be treated as a node in a network of dependency relationships. Each dependency relationship connecting any two packages (nodes) is treated as a link (an edge), and every link establishes a relation between a prior package and a posterior package, whereby the functions defined in the prior package are invoked in the posterior package. This enables reuse (economy) of functions and eliminates duplicate development. As a result the whole operating system emerges as a coherent and stable semantic network, with an unambiguous flow of meaning, determined by the direction of the links.

Semantic networks have, in their own right, become a subject of major research interest, especially where small-world structures and scale-free aspects of networks are concerned [14]. Scale-free patterns of connectivity in semantic networks hold out the promise of revising conventional models of semantic organisation, which are based on hierarchical principles and arbitrarily structured networks. However, unlike other semantic networks with power-law features [14], the presently studied semantic network of nodes in the *Debian* repository is founded on one single relation spanning across all its nodes: Y depends on X; its inverse, X is required for Y. In other words, the semantic network so formed is a dependency-based network only. This is the one most crucial point for understanding all subsequent arguments in this paper, although it will not be out of place to mention here that dependency-based relations can also be viewed from a much wider perspective [22–25].

Considering any particular node in a directed network, its links (the relations with other nodes) can be found to be of two types — incoming links and outgoing links — as a result of which, there will arise two distinct kinds of directed network [5]. For the network of incoming links in the *Etch* release of *Debian*, an earlier study [26] has empirically tested the occurrence of Zipf's law in the *GNU/Linux* distribution. One may note with great curiosity that over the years there has been a widespread emergence of Zipf's law, discovered origi-

*Electronic address: rajiv@hbcse.tifr.res.in
†Electronic address: nagarjun@gnowledge.org
‡Electronic address: arnab.kumar@jiet.ac.in

nally in studies of word frequencies in natural languages [27], in various diverse phenomena [26] ranging from city size distributions [27–31] to the frequency distribution of signs in undeciphered scripts from antiquity [32].

Carrying along these very lines, the present work affirms the existence of Zipf's law as a universal feature underlying the *FOSS* network. Both the networks of incoming and outgoing links have been seen to follow Zipf's law. However, it has also been realised that simple power-law properties do not suffice to provide a complete global model for directed networks. There is a general appreciation that for any system with a finite size, the power-law trend is not manifested indefinitely [33–35], and in the context of the *FOSS* network, this is a matter that is recognised as one worthy of a more thorough investigation [26]. Deviations from the power-law trend appear for both the profusely linked and the sparsely linked nodes. The former case corresponds to the distribution of a disproportionately high number of links connected to a very few special nodes — the so-called "hubs" (alternatively termed as rich nodes or top nodes), making the importance of these nodes, a self-evident fact. The particular properties of all these outlying nodes, as well as any distinguishing characteristic of the two directed networks can only be found by modelling the finite-size effects (equivalently the saturation properties) in the respective networks , and to study how these effects are related to the underlying semantic structure in the network. These are the principal objectives of this paper.

II. A NONLINEAR CONTINUUM MODEL

The mathematical modelling of the *FOSS* network has been carried out primarily with the help of data collected from the two latest stable *Debian* releases, *Lenny* (*Debian GNU/Linux 5.0*) and *Etch* (*Debian GNU/Linux 4.0*).[1] The networks of both the incoming links and the outgoing links span about 18000 nodes (software packages) in the *Etch* release, while in the *Lenny* release, the corresponding number of nodes is about 23000. For this work, the chosen computer architecture supported by both the releases is *AMD64*. The dynamic features of the model have further been grounded on the first three generations of *Debian releases*, i.e. *Buzz* (*Debian GNU/Linux 1.1*), *Hamm* (*Debian GNU/Linux 2.0*) and *Woody* (*Debian GNU/Linux 3.0*), all of which are supported by the architecture *i386*. While the retrospective compatibility of the model with the early releases has been gratifying, moving forward in time, the model has also behaved in full consonance with the features shown by the latest unstable *Debian* release, *Squeeze* (*Debian GNU/Linux 6.0*), which is again based on the *AMD64* architecture. The graphical results presented in this paper pertain mostly to the three latest releases, *Etch*, *Lenny* and *Squeeze*, all of which have a substantial number of nodes and links, and are, therefore, well suited for the kind of quantitative analysis that is needed for an accurate model building.

[1] `http://www.debian.org/releases`

For developing the model it will be necessary first to count the actual number of software packages, ϕ, which are connected by a particular number of links, x, in either kind of network. This gives an unnormalised frequency distribution plot of $\phi \equiv \phi(x)$ versus x. Normalising this distribution in terms of the relative frequency distribution of the occurrence of packages would have yielded the usual probability density function. To bring forth a continuum model for any power-law feature in this kind of a frequency distribution, one might posit a general nonlinear logistic-type equation [36] going as

$$(x + \lambda) \frac{\mathrm{d}\phi}{\mathrm{d}x} = \alpha\phi \left(1 - \eta\phi^{\mu}\right) , \tag{1}$$

with α being a power-law exponent, μ being a nonlinear saturation exponent, η being a "tuning" parameter for nonlinearity and λ being another parameter that will be instrumental in setting a limiting scale for the poorly connected nodes. The motivation behind this mathematical prescription can be easily followed by noting that when $\eta = \lambda = 0$, there will be a global power-law distribution. However, when the distribution is finite, the power-law trend fails to hold true beyond intermediate scales of x. Such deviations from a full power-law behaviour is especially prominent for high values of x (related to the very richly connected nodes) and, therefore, it becomes possible to argue that finiteness in the distribution is connected to saturation in the system. This type of saturation behaviour is frequently modelled by a nonlinear logistic equation [36–39], which, given the stated aims of this study, continues to serve as an effective mathematical instrument here as well. And so in order to understand the saturation properties of the highly connected nodes (arguably the more important nodes) in the *Debian* network, determining the influence of nonlinearity, expressed mathematically by μ and η in the nonlinear model, assumes great significance.

It should also be pertinent here to stress that as opposed to the practice of analysing data with a cumulative distribution function, the modelling for this study has been based directly on the unprocessed (noisy) data taken from both the releases. The operative argument behind this method is simple: As regards the *Debian* network the existence of a power-law degree distribution (specifically Zipf's law) is not a matter of doubt anymore [26]. The objects worth investigating beyond this point are the saturation properties of the system, and then interpreting the related finiteness of the network as a straightforward consequence of the limit to the semantic possibilities that the network can accommodate. These questions can only be addressed by analysing the noisy data, and grounding the mathematical model self-consistently in terms of numbers that convey direct information about the network as it actually is. Quantities indicated by η and λ attend exactly to this necessity. And ultimately the most convincing justification for analysing the noisy data in this work is seen to come from the values of the parameters α, μ, η and λ, as they have been calibrated from the data. Not only do some of these quantities remain consistent with the findings of previous authors [26], but collectively they also allow a new insight to be had into the finite-size properties and dynamic aspects of the *FOSS* network.

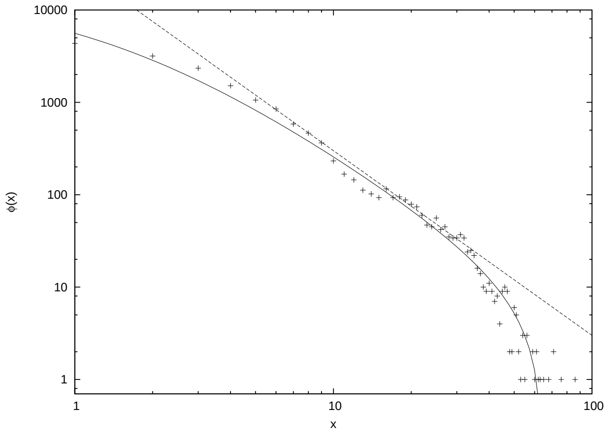

FIG. 1: For the network of incoming links in the *Etch* release, the degree distribution shows a good fit in the intermediate region with a power-law exponent, $\alpha = -2$ (as indicated by the dotted straight line), which validates Zipf's law. However, for large values of x, there is a saturation behaviour towards a limiting scale that is modelled well with the parameter, $\eta = -8$. On the other hand, when x is small, the fit is good for $\lambda = 1.5$. The global fit becomes possible only when $\mu = -1$, which turns out to be a universally valid number. For this specific plot, the data are fitted by $c \simeq 190$.

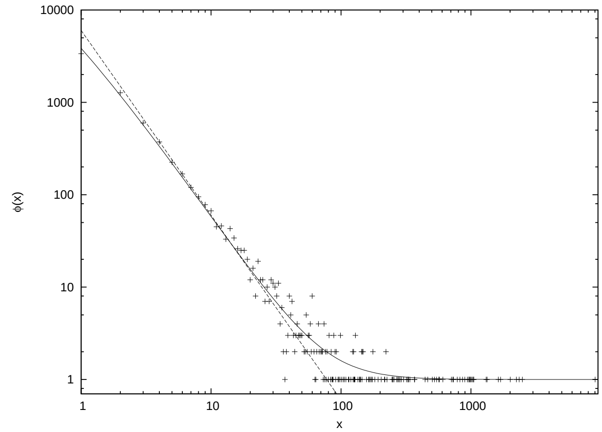

FIG. 2: For the network of outgoing links in the *Etch* release, the degree distribution of intermediate nodes are again modelled well by a power-law exponent, $\alpha = -2$, which is Zipf's law (as the dotted straight line shows). However, the saturation behaviour of the top nodes is very different from that of the network of incoming links. There is a clear convergence of ϕ towards a limit given by $\eta = 1$ (with μ remaining unchanged at -1). For the poorly linked nodes the convergence is attained for $\lambda = 0.25$. Thus, when α and μ remain the same, the value and the sign of η, as well as the value of λ, distinguish the type of a dependency network. The data are fitted for $c \simeq 80$. A solitary top node is to be seen for $x = 9025$.

Integration of Eq. (1), which is a nonlinear differential equation, can be carried out by making the appropriate substitutions, $\xi = \phi^\mu$ and $y = x + \lambda$, followed by the application of the method of partial fractions. An intermediate solution in ξ and y is then obtained as

$$\left(\frac{y}{c}\right)^{\alpha\mu} = \frac{\xi}{1 - \eta\xi}\,, \tag{2}$$

in which c is an integration constant. Resubstituting Eq. (2) back to the variables ϕ and x, will finally deliver the integral solution of Eq. (1) as (for $\mu \neq 0$)

$$\phi(x) = \left[\eta + \left(\frac{x+\lambda}{c}\right)^{-\mu\alpha}\right]^{-1/\mu}. \tag{3}$$

It is quite obvious that when $\eta = \lambda = 0$ (with the former condition implying the absence of nonlinearity), there will be a global power-law distribution for the data, going as $\phi(x) = (x/c)^\alpha$, regardless of any non-zero value of μ. The situation becomes quite different, however, when both η and λ have non-zero values. In this situation, the network will exhibit a saturation behaviour on extreme scales of x (both low and high). For the high values of x, this can be easily appreciated from Eq. (1) itself, wherefrom the limiting value of ϕ is obtained as $\phi = \eta^{-1/\mu}$.

III. MODEL FITTING OF THE *FOSS* NETWORK

The parameters α, μ, η, λ and c in the solution given by Eq. (3) can now be fixed by the distribution of links and nodes obtained from the *Debian* repository. In Fig. 1 the degree distribution for incoming links in the *Etch* release has been plotted. The dotted straight line in this log-log plot indicates the pure power-law behaviour. While this gives a satisfactory description for the distribution on intermediate scales of x, there is a clear departure from the power law both as $x \longrightarrow 0$ and $x \longrightarrow \infty$. The solution given by Eq. (3) fits the power law, as well as the departure from it, at both the small-connectivity and the high-connectivity ends, for the values $\alpha = -2$, $\mu = -1$, $\eta = -8$, $\lambda = 1.5$ and $c \simeq 190$. It would be very interesting to note here that the values of α and μ remain unchanged when it comes to giving a model fit to the degree distribution of outgoing links (once again for the *Etch* release), as shown in the plot in Fig. 2. The obvious implication of $\alpha = -2$ in both the cases is that Zipf's law universally underlies the frequency distribution of the intermediate nodes and links in both kinds of network. The only quantitative measure to distinguish between the two networks are the values of η, λ and c, set at $\eta = 1$, $\lambda = 0.25$ and $c \simeq 80$ for the distribution of outgoing links in Fig. 2.

Similarly, data from the *Lenny* release have been plotted in Figs. 3 & 4. The former plot gives the in-degree distribution of the nodes, while the latter gives the out-degree distribution. The in-degree distribution in the *Lenny* release changes from the previous release, *Etch*, in the values $\eta = -15$, $\lambda = 1.6$ and $c \simeq 210$. The saturation properties in this case, therefore, undergo a significant quantitative change at the highly connected end with a new generation of *Debian* release. In contrast, for the out-degree distribution the changes across a new generation of *Debian* release are calibrated by $\lambda = 0.35$ and $c \simeq 90$,

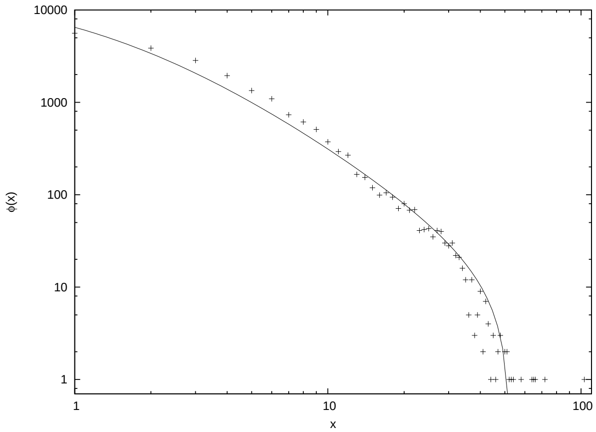

FIG. 3: For the network of incoming links from the *Lenny* release, the intermediate nodes (fitted with a power-law exponent, $\alpha = -2$) uphold Zipf's law once again. For large values of x, however, the saturation behaviour towards a limiting scale of ϕ is modelled by the value, $\eta = -15$. On the other hand, when x is small, the fit continues to be good for $\lambda = 1.6$. Once again $\mu = -1$, but for this particular plot, $c \simeq 210$. It is clear that the richly linked nodes here are generally less connected than what they are in the case of the *Etch* release, as Fig. 1 indicates.

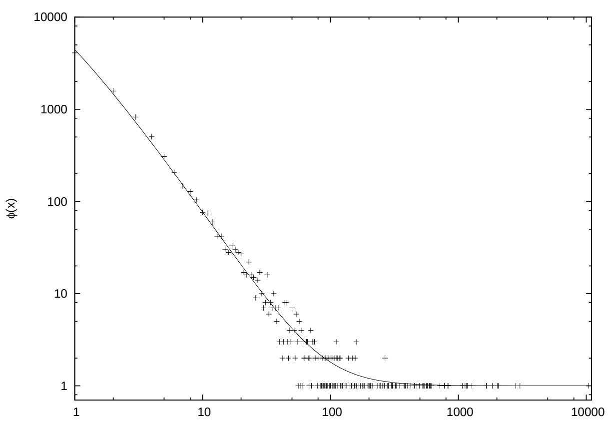

FIG. 4: For the network of outgoing links from the *Lenny* release, the distribution of intermediate nodes obeys Zipf's law, as the power-law exponent, $\alpha = -2$, shows. The saturation behaviour of the top nodes remains the same as it is for the *Etch* data. The convergence of ϕ towards a limit set by $\eta = 1$ is quite obvious, with $\mu = -1$, as usual. For the poorly linked nodes, the convergence is given by $\lambda = 0.35$. The other value that distinguishes the out-degree distribution in the *Lenny* release, from that in the *Etch* release, is $c \simeq 90$. A solitary top node is to be seen for $x = 10446$.

which implies that the saturation properties remain unchanged at the richly linked end, but changes at the poorly connected end. Changes in the value of c for a particular type of degree distribution, causes a translation of the model curve in the x–ϕ plane. And, as Figs. 1, 2, 3 & 4 indicate, Zipf's law continues to prevail in all the cases.

It should also be instructive here to make a theoretical analysis of the value of μ obtained from the data, and its accompanying consequences. Elementary algebraic manipulations on Eq. (3), followed by a power-series expansion will lead to the infinite series

$$\begin{aligned}\phi(x) \;&=\; \left(\frac{x+\lambda}{c}\right)^{\alpha} - \frac{\eta}{\mu}\left(\frac{x+\lambda}{c}\right)^{\alpha(\mu+1)} \\ &\quad + \frac{\mu+1}{2}\left(\frac{\eta}{\mu}\right)^{2}\left(\frac{x+\lambda}{c}\right)^{\alpha(2\mu+1)} + \cdots\,, \qquad (4)\end{aligned}$$

from which it is not difficult to see that a self-contained and natural truncation for this series can only be achieved when $\mu = -1$. It is remarkable that the *Debian* data conform to this fact, and in consequence of this value of μ, Eq. (1) is reduced to being a linear, first-order, nonhomogeneous equation,

$$\frac{\mathrm{d}\phi}{\mathrm{d}x} - \left(\frac{\alpha}{x+\lambda}\right)\phi = -\left(\frac{\eta\alpha}{x+\lambda}\right), \qquad (5)$$

in which η plays the role of a nonhomogeneity parameter.

With the measured values of $\alpha = -2$ and $\mu = -1$, as derived from the data for both the in-degree and out-degree distributions in the *Etch* and *Lenny* releases, the saturation properties in the network (for any value of η and λ) can, therefore, be abstracted in a compact form from Eq. (3) through the solution

$$\phi(x) = \eta + \left(\frac{c}{x+\lambda}\right)^{2}. \qquad (6)$$

The implications of the foregoing result are noteworthy. One of these is that nonhomogeneity in the system sets a firm lower bound to the number of rich nodes in the saturation regime, regardless of any arbitrarily high value of x, i.e. $\phi \longrightarrow \eta$ as $x \longrightarrow \infty$. In other words, nonhomogeneity defines a finite lower limit to the discrete count of the rich nodes. This evident deviation from the power-law model enables a few top nodes in the network of outgoing links to get disproportionately rich, as shown in Figs. 2 & 4. All the links from these top nodes are outwardly directed towards the dependent nodes, making the presence of these richly linked nodes an absolute necessity, and burdening them with the responsibility of maintaining full functional coherence in the semantic system structured by the *FOSS* network. That the value of η remains unchanged with increasing values of x across two generations of *Debian* releases, shows that the primacy of the top nodes continues to remain unshaken. A general scale for the onset of the saturation effects in the network of out-degree distributions can also be ascertained by requiring the two terms on the right hand side of Eq. (3) to be in rough equipartition with each other. This will set a scale for the saturation of the number of links in the frequency distribution as

$$x_{\mathrm{sat}} \sim |\eta|^{-1/\mu\alpha}. \qquad (7)$$

For the network of outgoing links, the *Debian* data indicate that approximately the top 1% of the nodes falls within this scale, with the package *libc6* seeming to be the most profusely connected node (having 9025 links in the *Etch* release, and 10446 links in the *Lenny* release) in the entire network.

The situation is quite the opposite for the network of incoming links, as Figs. 1 & 3 show. Here the nodes draw in links to themselves, with all links being inwardly directed towards the nodes. It is quite evident that this network of incoming links is complementary in character to the network of outgoing links. As a result, the richly linked nodes of the latter network are poorly connected in the former. In contrast to Figs. 2 & 4, which indicate that the rich nodes serve the network to an extent that is disproportionately greater than what a simple power-law behaviour would have required of them, one may discern from Figs. 1 & 3 that the most richly linked nodes in the in-degree distribution display a behaviour that falls short of what might be expected of a fully power-law trend (the top nodes here ought to have accreted more links if a pure power law were to have been followed). In fact, the value of η decreases across two generations of *Debian* releases, making it clear that the ability of the top nodes to acquire links becomes generally more enfeebled (and so it is that the deviation from the power-law behaviour becomes sharper). Beyond this qualitative observation, one may note that saturation in the network can be quantitatively determined by the parameter η, which, when $\mu = -1$, appears as a nonhomogeneity condition in Eq. (1). The value and especially the sign of η afford a precise quantitative means to differentiate between the directed networks of incoming and outgoing links. The difference in the respective degree distributions in Figs. 1 & 2 (or Figs. 3 & 4) underscores this fact.

Considering the other extreme of finite-size effects at small values of x, the very poorly linked nodes are also seen to deviate noticeably from the power-law solution. This is especially true for the in-degree distribution in Figs. 1 & 3. Apropos of this, it might be mentioned here as an aside, that the present literature in the domain of econophysics, where all relevant data distributions are nearly the same as what has been shown here, indicates that the distribution of such feeble nodes might be modelled by a Boltzmann-Gibbs or a log-normal distribution [8, 40–43], below a certain lower cut-off value of x (the lower limit of the range of the power-law regime). On the other hand, for small in-degree and out-degree distributions in the *World Wide Web* [44], an improved fit can be obtained by a simple modification in the global power-law model [1, 35]. This kind of modification can also be engineered in Eq. (6) itself to obtain a similar fit for the weakly linked nodes. In the limit of small degree distributions for both the in-directed and out-directed networks, where η ceases to have a quantitative significance, and where $x \sim 1$ (which, in the discrete count of links, is the lowest value that x can assume practically), one can find an upper bound to the number of the very sparsely linked nodes. This bound is given as

$$\phi_{\mathrm{ub}} \simeq \left(\frac{c}{1+\lambda}\right)^2 , \tag{8}$$

with the full range of ϕ, therefore, going as $\eta \leq \phi \lesssim \phi_{\mathrm{ub}}$.

While dwelling on various properties of the *FOSS* network, it will also be worthwhile to bear in mind that if this network is to operate ideally as a coherent and stable semantic system, then it will be desirable to have parsimony in the creation of the nodes (enforcing reuse of functions in packages), and elimination of duplicate development and ambiguity (multiple packages with identical purpose) — a condition suggesting absolute interpretation, whereby the operational context of each node is understood with complete disambiguity. These requirements thus enjoin that if a *FOSS* network were to be unambiguously interpreted, then no two software packages belonging to it, should be exactly alike in their functionalities and dependencies. This adheres to the conventional wisdom about the growth of semantic networks — structures which grow by semantic differentiation of existing nodes (concepts) [14]. The meaning of a new node derives from a well-defined variation on the meaning of an already existing node, and, therefore, each newly created node acquires its own very specifically determined set of links within the network, to suit its particular purpose for having come into existence. For every node, the neighbourhood consequently becomes a unique structure. Now structure always affects function [14, 45]. Taking this fact in conjunction with the semantic viewpoint that meaning is inseparable from structure [14], it is not difficult to appreciate that the meaning or the functionality of a particular node is defined uniquely only by the way in which the dependency neighbourhood of the node is structured. Equipped with a large array of such unique nodes, the network thus becomes enabled to accommodate a large variety of semantic possibilites. Extending these arguments it also becomes possible to suggest by way of an analogy that if an unequivocally determined *FOSS* network could be viewed as a single "quantum state", then the packages in the network would have to accord to an "exclusion principle" in order to maintain disambiguity of operations in the network and create a fully interpreted semantic system (partial interpretation implies ambiguity). And continuing on this "quantum" theme, the nodes might be understood to assume the character of fermions, with the full range of semantic possibilities of each package (an exclusive node in the network) being characterised as a unique set of "quantum numbers". From this perspective it can be said that in the single "quantum state" of a fully interpreted *FOSS* network, no two nodes will have an identical set of "quantum numbers", i.e. when the semantic relationship in the network is only of dependencies, no two nodes will have an identical set of dependencies.[2] The term "nodes" here implies software packages. Now it has to be noted that software packages are "coarse-grained" structures, defined by the preferences of the programmer and the exigencies arising in the operating system. The actual objective unit of the operating system is, however, the function, which lies deep within a software package. So the quantum analogy is expected to hold much more robustly if one does a

[2] In the case of the *Debian GNU/Linux* repository, it is indeed a fact that each node is identified by a unique label, and serves some unique function in the whole network. Of course, one has to be cautioned that there is no physical principle inherent in the *Debian* network that actually enforces this "exclusion principle", unlike what happens in a real quantum system with fermions.

"fine-grained" tracing of the semantic flow and the associated dependency network at the scale of the functions, rather than at the scale of packages. In passing one may also note that in the context of complex networks in general, the analogy with a quantum system is not entirely stretched, because some earlier studies [46, 47] have in fact modelled the convergence of a macroscopic fraction of links onto any particular top node in terms of Bose-Einstein condensation, with the links being interpreted as bosons, and the nodes (to which the links attach themselves preferentially) as energy levels.

A very significant difference between the degree distributions of the *World Wide Web* and the *FOSS* network is that the frequency histograms pertaining to one appear to be exactly the converse of the other. And so what looks like an in-degree distribution for one, is the out-degree distribution for the other, and vice versa [1]. An explanation for this difference can be offered. In the *FOSS* network, the dependency relationships follow precise rules, with out-degree dependencies preponderating over the in-degree ones. The existence of a lesser node is dependent on having a link directed to this node from a top node. For the *World Wide Web*, however, the situation is qualitatively different. To be relevant in this network, a weak node contributes a directed link from itself to a top node (a case of preferential linking) in a manner that becomes the dominant mode of establishing links. In any case, the network in the *World Wide Web* is not based on dependency relationships, unlike what it is in the *FOSS* network.

It happens not very infrequently that in a functioning *FOSS* network taken from the *Debian* repository, there are some packages which are not compatible with one another. The relationship among these packages is, therefore, not of dependencies, but of what is technically known as "conflicts". One could collect data on these nodes (the conflicted packages), and unearth any possible degree distribution that might govern the network. Their frequency distribution (for both the *Etch* and *Lenny* releases) looks like what has been shown in Fig. 5. The intermediate scale-free part of this distribution has been fitted well by a power law exponent of $\alpha = -4$. This makes the frequency distribution here much steeper than what it is for the dependency distributions in Figs. 1 & 2 (or Figs. 3 & 4), both of which have been fitted by Zipf's law. Consequently the approach towards the saturation state at the highly linked end (set by $\eta = 1$) is much more rapid in this case. And as one might expect of any conflict-ridden system, this "network" of dysfunctional relationships is sparsely populated. The stability of the distribution in this case is also a matter of concern, because it is well known [34, 48–50] that power-law distributions can only be stable over the Lévy-stable regime of $0 < -\alpha \leq 2$.

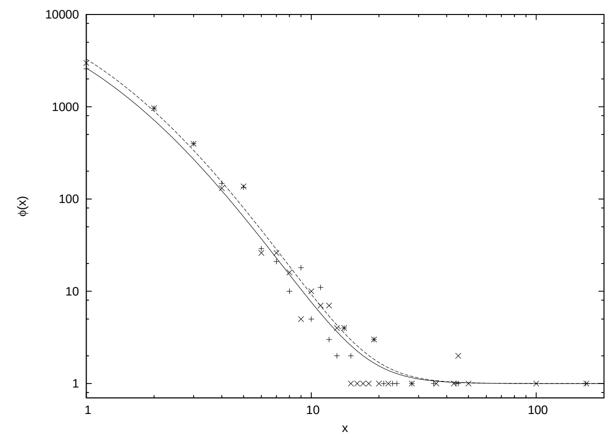

FIG. 5: For the sparsely populated network of conflicts, the power-law exponent of $\alpha = -4$ makes for a much steeper approach to the saturated state at the highly linked end, calibrated by $\eta = 1$. The value of $\mu = -1$ continues to hold true here. The continuous curve gives the theoretical fit to the *Etch* data (indicated by the $+$ points) for $c \simeq 19$, while the dotted curve gives the theoretical fit to the *Lenny* data (indicated by the $\times$ points) for $c \simeq 20$. For small values of x, the shift away from the power-law solution is fixed by $\lambda = 1.6$ for both the *Etch* and *Lenny* releases.

IV. MODELLING EVOLUTION AND SATURATION

This study, based principally on two generations (*Etch* and *Lenny*) of a standard *FOSS* network (*Debian*), has shown that the saturation properties in the in-degree and the out-degree distributions are differently affected through the passage of time (marked by new releases of *Debian*). The degree distribution of the network of outgoing links shows no change at all when it comes to model fitting for the top nodes (η continues to have the same value). This is in fact to be expected entirely of these nodes. They are the foundation of the entire network, and their prime status continues to hold. In a semantic sense, meaning flows from these nodes to the derivative nodes. The very poorly linked nodes in the outgoing network, on the other hand, are fitted by changing values of λ (as shown in Figs. 2 & 4). Again this is expected. In a mature and robust network, the possibility of semantic variations is much more open in the weakly linked derivative nodes, as opposed to the primordial nodes (which are like parent nodes, where genesis of meaning takes place, and in which all the axioms are founded).

For the in-degree distributions, the situation is nearly the opposite. What Figs. 1 & 3 show is that the model fitting can be achieved properly by changing the value of η significantly. Further, with a new release of *Debian*, η actually decreases, a fact whose import can only be that the most richly linked nodes in the in-degree distribution (which are also the most dependent nodes) acquire less links than what they might have done, if the power-law trend were to have been adhered to indefinitely. So, from the dynamic perspective, there is a terminal character to the extent upto which these dependent nodes continue to be linked.

In view of these observations, it becomes pragmatic to appreciate that the *FOSS* network is not a static entity. Rather it is a dynamically evolving network, as any standard software network is known to be [51, 52], undergoing continuous additions (even deletions) and modifications across several generations of *Debian* releases, contributed by the community of free-software developers. So any realistic model should account for this evolutionary aspect of the network distribution.

And indeed, by now many theoretical models [53–56] have afforded varied insight into the general question of dynamic evolution of networks. It has also been demonstrated conclusively that scale-free networks can only emerge through the simultaneous operation of dynamic growth and preferential attachment [53, 57]. The limiting features in such a scale-free distribution, regardless of the abundance of the links that these nodes may acquire, ought also to manifest itself naturally through the long-time dynamics.

It has been reasoned that the number of the top nodes in the out-degree distribution form the irreducible nucleus of the *FOSS* network. These nodes are by far the most influential in the network, when it comes to determining the future destiny of the network. It will be useful, therefore, to take a closer look at the dynamic features and the saturation properties of the out-degree distribution. From the perspective of a continuum model, one could envisage the frequency distribution of the nodes in the network of outgoing links, as a field, $\phi(x,t)$, evolving continuously through time, t, with the saturation in the number of nodes for high values of x, emerging of its own accord from the dynamics. In keeping with this need, an ansatz with a general power-law feature inherent in it, may be framed as

$$\phi(x,t) = \left(\frac{x+\lambda}{c}\right)^{\alpha} + \varphi(x,t)\,, \tag{9}$$

in which $\varphi \longrightarrow \eta$, as $t \longrightarrow \infty$. This prescription would be compatible with what Eq. (3) indicates when $\mu = -1$. So, under this requirement, one may describe the temporal evolution of the network by a first-order, linear, nonhomogeneous model equation, going as,

$$\tau\frac{\partial\phi}{\partial t} = \frac{\partial\phi}{\partial x} - \frac{\alpha}{c^{\alpha}}\left(x+\lambda\right)^{\alpha-1}\,, \tag{10}$$

in which τ is a parameter that indicates a representative time scale on which the *FOSS* network evolves appreciably. It should be important to emphasise that Eq. (10) already has an explicit presence of a power-law property built in it, and is expected, upon being integrated under suitable initial conditions, to make the saturation features of the top nodes appear because of nonhomogeneity. This is the exact reverse of Eq. (5), which has nonhomogeneity explicitly stated in it, and upon being integrated, leads to a power-law behaviour. The general solution of Eq. (10) can be obtained by the method of characteristics [58, 59], for which the pertinent equations are

$$-\frac{\mathrm{d}t}{\tau} = \frac{\mathrm{d}x}{1} = \frac{\mathrm{d}\phi}{\alpha(x+\lambda)^{\alpha-1}c^{-\alpha}}\,. \tag{11}$$

The solution of the $\mathrm{d}\phi/\mathrm{d}x$ equation is

$$\phi - \left(\frac{x+\lambda}{c}\right)^{\alpha} = a\,, \tag{12}$$

while the solution of the $\mathrm{d}x/\mathrm{d}t$ equation is

$$x + \frac{t}{\tau} = b\,, \tag{13}$$

with both a and b being integration constants. The general solution is to be found under the condition that one characteristic solution of Eq. (11) is an arbitrary function of the other, i.e. $a = f(b)$, with f having to be determined from the initial conditions [58, 59]. So, going by the integral solutions given by Eqs. (12) and (13), the general solution of $\phi(x,t)$ will be

$$f\left(x+\frac{t}{\tau}\right) = \phi - \left(\frac{x+\lambda}{c}\right)^{\alpha}\,, \tag{14}$$

which, under the initial condition that $\phi = \eta$ at $t = 0$ for any value of x, will characterise the profile of the arbitrary function, f, as

$$f(z) = \eta - \left(\frac{z+\lambda}{c}\right)^{\alpha}\,. \tag{15}$$

Hence, the specific solution can be obtained from Eq. (14) as

$$\phi(x,t) = \eta + \left(\frac{x+\lambda}{c}\right)^{\alpha} - \left[\frac{1}{c}\left(x+\lambda+\frac{t}{\tau}\right)\right]^{\alpha}\,, \tag{16}$$

and this, under the condition that $\alpha = -2$, will converge to the distribution given by Eq. (6), for $t \longrightarrow \infty$. In this regard the initial condition and its import are worth stressing. For any given value of x, the evolution starts at $t = 0$ with an initial node count of $\phi = \eta$, which, under all practical circumstances, will be set at $\eta = 1$. This is tantamount to saying that a node appears in the network with x number of links, where, previously, there existed no node with this particular number of links. As the network evolves, two things continue to happen: First, new nodes are added to the network, and secondly, already existing nodes accrete greater number of links and strives to gain a status of greater importance. The most richly linked among the latter kind of nodes are primordial in nature, and at $t = 0$, their number defines the minimum number of independent packages that are absolutely necessary for the *FOSS* network to evolve subsequently (for $t > 0$) into a robust semantic system. So the initial condition can be argued to have an axiomatic character, and the mature network burgeons from it on later time scales. And during the evolution, the entire network gets dynamically self-organised in such a manner, that the eventual static out-degree distribution has its saturation properties at the highly connected end (arguably the more important end) determined by what the initial field was like at $t = 0$.

At this stage it should be instructive to examine the asymptotic properties of Eq. (16), both in the limit of $t \longrightarrow 0$ and in the limit of $t \longrightarrow \infty$. In the former case, the evolution of ϕ will be linear in t for a given value of x, and will go as

$$\phi(x,t) \simeq \eta - \alpha\frac{(x+\lambda)^{\alpha-1}}{c^{\alpha}}\left(\frac{t}{\tau}\right)\,, \tag{17}$$

in which growth is assured only when $\alpha < 0$. This linearity of early growth is perfectly in consonance with the standard assumption of the linear growth of the number of nodes with time [60], with η giving the initial number of nodes.

While the temporal evolution obeys linearity on early time scales, in the opposite limit of $t \longrightarrow \infty$, the evolution shifts asymptotically to a power-law trend going as

$$\phi(x,t) - \eta - \left(\frac{x+\lambda}{c}\right)^{\alpha} \simeq -\frac{1}{c^{\alpha}}\left(\frac{t}{\tau}\right)^{\alpha} . \qquad (18)$$

Naturally, convergence towards a steady state, as it has been given by the condition on the left hand side of the foregoing relation, will be possible only when $\alpha < 0$, a requirement that is eminently satisfied by Zipf's law ($\alpha = -2$). It is certainly intriguing that the model should indicate that the power-law convergence towards a steady state solution should follow an exponent given in particular by Zipf's law, although in a general sense, open-source software has been known to have its dynamic processes driven by power laws [20, 21], which is a clear sign that long memory prevails in this kind of a system.

Now examining the steady state form of the degree distribution, as it has been suggested by Eq. (3), one can set down, for $\mu = -1$, a similar kind of a relation for the time-dependent field, $\phi \equiv \phi(x,t)$, as,

$$\phi(x,t) = \eta + \left(\frac{x+\tilde{\lambda}}{\tilde{c}}\right)^{\alpha} , \qquad (19)$$

where $\tilde{\lambda}$ and $\tilde{c}$ are "dressed" parameters, defined as $\tilde{\lambda} = \lambda\nu(x,t)$ and $\tilde{c} = c\zeta(x,t)$, respectively. The scaling form of the two functions ν and ζ can be determined by equating the right hand sides of Eqs. (16) and (19). This will lead to

$$\left(\frac{x+\tilde{\lambda}}{\tilde{c}}\right)^{\alpha} = \left(\frac{x+\lambda}{c}\right)^{\alpha} - \left[\frac{1}{c}\left(x+\lambda+\frac{t}{\tau}\right)\right]^{\alpha} . \quad (20)$$

For scales of $x \gg \lambda$ (typically $x \gtrsim 10$), a converging power-series expansion of increasingly higher orders of λ/x can be carried out with the help of Eq. (20). The zeroth-order condition will deliver the scaling profile of ζ as

$$\zeta(x,t) = \left[1 - \left(1+\frac{t}{x\tau}\right)^{\alpha}\right]^{-1/\alpha} . \qquad (21)$$

This function bears time-translational properties, and at a given scale of x, it causes the degree distribution to shift across the x–ϕ plot through the passage of time. But it is also not difficult to see that when $\alpha = -2$, there is a convergence towards $\zeta = 1$ (the steady state limit) as $t \longrightarrow \infty$. On the other hand, when $x \longrightarrow \infty$, for any finite time scale, $\zeta \longrightarrow 0$. This explains why the count of the most heavily connected nodes (for which x has a high value) stays nearly the same ($\phi = \eta$) at all times, a fact that is quite evidently borne out by the out-degree distributions in Figs. 2 & 4. The saturation scale of x for such behaviour is given by Eq. (7). A related fact that also emerges is that time-translation of the degree distribution becomes steadily more pronounced as one moves away from $x \sim x_{\mathrm{sat}}$ towards the limiting value of $x = 1$ (the lowest possible number of links that a node can possess). Consequently, as the temporal evolution progresses, the out-degree distribution assumes a slanted appearance with a negative slope in the x–ϕ plane, something that has been shown very clearly once again in Figs. 2 & 4. The model fitting in these two plots indicates that the value of c increases with time. This is exactly how it should be, going by the form of the scaling function $\zeta(x,t)$, if one is careful enough to observe that c in both the plots is to be viewed rather as $\tilde{c}$, to comprehend fully its time-dependent variation.

Gaining a clear understanding of the time-translational properties of the poorly connected nodes is not quite as straightforward. Information regarding this matter is contained in the scaling function $\nu(x,t)$. However, a look at the left hand side of Eq. (20) will reveal that ν is coupled to ζ, and it is this nonlinear coupling that causes complications. Going back to the power-series expansion in λ/x, as it can be obtained from Eq. (20), one may be tempted to think that just as the zeroth-order in the series has yielded a proper scaling form for ζ, the higher orders in the series will bring forth a similar form for ν. And indeed one does obtain such a solution, going as $\nu^k = \zeta^\alpha \left[1-(1+t/x\tau)^{\alpha-k}\right]$, with k being the order of the expansion in the power series. But this result is misleading because the parameter λ, and in connection with it, the scaling function $\nu(x,t)$, make an effective imprint only when $\lambda \gtrsim x$, with x assuming arbitrarily small values in the continuum model. Therefore, the correct approach here is not to take a series expansion in λ/x, but rather in x/λ, with a proper convergence of the series taking place for higher orders in x/λ. The zeroth-order term of this series gives the scaling form $\nu^\alpha = \zeta^\alpha \left[1-(1+t/\lambda\tau)^{\alpha}\right]$. The primary difficulty with this result is that the true functional dependence of ζ in this case is not known. This is certainly not going to be the function that is implied by Eq. (21), because this form of ζ is valid only on scales where $x \gg \lambda$ (where the scale-free trend in the degree distribution holds, or when $x \sim x_{\mathrm{sat}}$).

For all that, the unequivocal message that is conveyed by the common pattern exhibited by the two generations of out-degree distributions is that the value of λ does have a significant bearing on the number of the preponderant but sparsely-connected nodes, a fact that is described by Eq. (8). In the continuum picture of the degree distribution, λ is the theoretical lower bound of the number of links that the most weakly linked nodes may possess (which saves ϕ from suffering a divergence as $x \longrightarrow 0$, going by what Eq. (6) states). Through the evolutionary growth of the network, an increase in the value of λ suggests that these poorly linked nodes become incrementally relevant to the system by contributing more links in the out-directed network. Now these very poorly connected nodes in the out-degree distribution are concomitantly the most richly linked nodes in the in-directed network. A look at Figs. 1 & 3 will show that for these nodes the value of η decreases with the evolution of the *FOSS* network. So, as regards these nodes it stands to reason that while they become progressively more relevant as members of the out-directed network (a condition quantified by increasing values of λ), as members of the in-directed network they become incrementally less dependent (quantified by decreasing values of η). Analysing the data gathered from all the six generations of *Debian*, what one sees is that in the out-directed network, the value of λ hovers around 0.25 for the first four releases

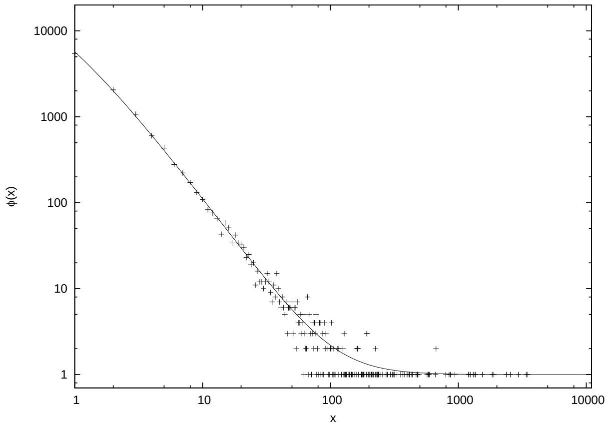

FIG. 6: The out-degree distribution of the latest unstable release, *Squeeze*, is in agreement with what the dynamic model predicts. The values of λ and c increase, as expected, to $\lambda = 0.45$ and $c \simeq 110$. The richest node in this distribution has 12524 links.

(upto *Etch*), but grows noticeably thereafter for the next two releases, *Lenny* ($\lambda = 0.35$) and *Squeeze* ($\lambda = 0.45$), of which, the latter has its out-degree distribution shown in Fig. 6. On the other hand, in the in-directed network, the value of λ grows quickly for the early releases (starting with $\lambda \simeq 0.5$) and then saturates towards $\lambda = 1.6$ for the *Lenny* and *Squeeze* releases. Remembering that in the in-directed network the most poorly linked nodes are actually the parent nodes of the entire network (both the in-directed and the out-directed), one can conclude that even these nodes become dependent on other nodes to a small extent. Taken as a whole, with the unfolding of time, the interdependency character of the entire network becomes more firmly established, with even the relatively unimportant nodes showing a tendency to contribute an increasingly greater number of out-directed links.

Quantitative support in favour of this claim comes from a different quarter. In the out-directed network (by far the more important network, when it comes to understanding the semantic flow taking place through the evolution of a *FOSS* network), the total number of nodes, $N_{\rm out}(t)$, at any given point of time, t, can be obtained by evaluating the integral

$$N_{\rm out}(t) = \int_1^{x_{\rm m}} \phi(x,t)\,{\rm d}x\,. \qquad (22)$$

The limits of this integral are decided by the limits on the number of links that the nodes may possibly possess — 1 being the lower limit and $x_{\rm m}$ being the upper (maximum) limit. The integral in Eq. (22) can be solved by taking the profile of $\phi(x,t)$ given by Eq. (16), for $\alpha = -2$. Noting that $x_{\rm m} \gg 1$ (typically $x_{\rm m} \sim 10^4$) for the out-directed network, the total number of nodes at any given point of time can, therefore, be estimated as,

$$N_{\rm out}(t) \simeq \eta x_{\rm m} + \frac{c^2}{1+\lambda} - c^2\left(1+\lambda+\frac{t}{\tau}\right)^{-1}. \qquad (23)$$

On moderate time scales, the last two terms on the right hand side in Eq. (23) are roughly equal. So the dominant contribution comes from the first term (interestingly enough, this is actually the saturation term), as a consequence of which, one could set down $N_{\rm out} \sim x_{\rm m}$ (since $\eta = 1$ according to the model fit). This argument becomes progressively more correct for large values of $x_{\rm m}$, i.e. for later releases of *Debian*.

For the out-degree distribution in the *Etch* release, $x_{\rm m} \simeq 9000$, while in the *Lenny* release, the corresponding number is about 10000. Using these values from both the releases of *Debian*, the respective count of $N_{\rm out}$ can be made very easily for the two successive generations. Both of these values of $N_{\rm out}$ represent the number of nodes that contribute at least one link in the out-directed network. These estimates compare very favourably with what the actual *Debian* data indicate. In the case of the *Etch* release, the number of software packages contributing to the out-directed network is counted to be about 7000 (which is closely comparable to the theoretical value of $N_{\rm out} \simeq 9000$), and in the case of the *Lenny* release, the total count of the out-directed nodes is about 9000 (which can be favourably compared once again to $N_{\rm out} \simeq 10000$). As a fraction of the total number of nodes, these actual counts indicate that the number of nodes in the out-directed network increases by 1% from the *Etch* release to the *Lenny* release. This certainly validates the quantitative contention that with each passing generation, the network becomes incrementally more robust in terms of out-degree contributions coming from an increasingly greater number of nodes. The values pertaining to the latest unstable release *Squeeze*, also support these findings handsomely. In this case the actual count of the out-directed nodes is about 11500, a number that is in fact much closer to the theoretical estimate of $N_{\rm out} \simeq 12500$, when one considers the corresponding match in the two earlier releases, *Etch* and *Lenny*. One may also note with curiosity that in these last three *Debian* releases — *Etch*, *Lenny* and *Squeeze* — the total number of software packages, spanning both the in-directed and out-directed networks, is roughly twice the value of $x_{\rm m}$ in the out-directed network.

The overall growth of the network, however, slowly grinds to a halt on long time scales. This is a conclusion that cannot be missed in Eq. (23), which suggests that the total number of nodes increases with time, but approaches a finite stationary value when $t \longrightarrow \infty$, unless $x_{\rm m}$ becomes infinitely large (something that is not very likely). This inclination of the network to saturate towards a finite-sized end can be explained in terms of the finite semantic possibilities associated with each of the nodes. The extent of making creative use of the existing semantic possibilities of even the most intensely linked of the top nodes is limited. Since most of the nodes in the network depend on such top nodes, there must be a saturation scale in the network. Unless novel creative elements in semantic terms are continuously added to the top nodes, the value of $x_{\rm m}$ will remain finite, and saturation cannot be prevented. Therefore, saturation in this semantic network is the inevitable consequence of the limit to which original axioms (functions deriving from the top nodes) can be made available. An emphatic illustration of this line of argument is to be seen in Fig. 7, which makes use of actual values taken from all the *Debian* releases. The lower curve in this plot tracks

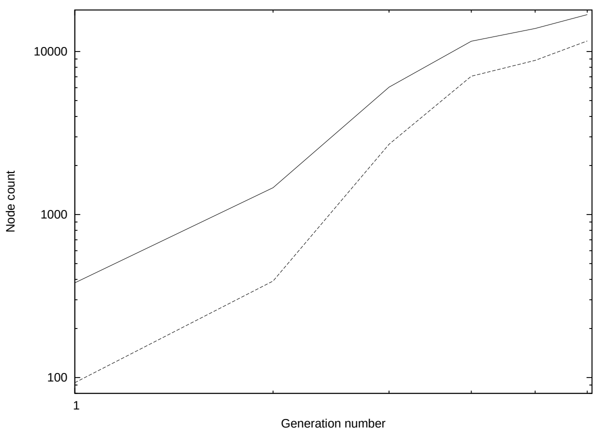

FIG. 7: The upper curve plots the growth of the number of "terminal" nodes (which do not contribute any out-directed link) in the in-directed network, while the lower curve plots the growth of the total number of nodes, $N_{\rm out}$, in the out-directed network. Both curves indicate a very closely correlated saturation behaviour with every passing generation of *Debian* release (marked along the horizontal axis). Data for this plot have been taken from all the six releases of *Debian*.

the growth of the total number of nodes in the out-directed network. All the members of this network contribute at least one out-directed link, and so meaning (the semantic context) is seen to flow out of these nodes. Therefore, to a greater (especially as regards the top nodes) or lesser extent, these nodes are the bearers of original axioms. That the growth of this entire out-directed network saturates towards a limiting value for the later releases of *Debian* (marking increasingly longer time scales) is quite obvious from the trend indicated by the lower curve in Fig. 7. A corollary of the saturation of the network of these axiom-bearing nodes (which contribute out-directed links) is that the network of in-directed nodes will also saturate in tandem. This is shown by the upper curve in Fig. 7, which plots the growth of the number of nodes which contribute no out-directed links, but only draws in links towards themselves. The semantic flow in the network terminates at these nodes, and as such the network of these "terminal" nodes is an appropriate indicator of the effect of saturation. One can see from the two curves that there is a very strong correlation between the saturation in the out-directed network, and the saturation in the number of terminal nodes.

Finally, it is worth a passing thought that the network of conflicts, whose degree distribution, taken from the *Etch* and *Lenny* releases, as Fig. 5 shows it, bears a qualitative resemblance to the out-degree distribution of dependencies, appears so sparsely populated because, with $\alpha = -4$, the temporal drive towards the static end is very rapid, leaving not much occasion for the network to flourish.

V. CONCLUDING REMARKS

The significance of finite-size effects and saturation at the extremities of the degree distribution has been cogently argued for. The mathematical model that has been developed here gives a clear quantitative characterisation of the incoming and outgoing distribution in the *Debian GNU/Linux* network. Similar features are known to exist in the degree distribution of other scale-free networks, and it should become quite possible to study the saturation properties and the specific directional characteristics of such general systems of scale-free networks, with the mathematical model that has been used here. In the context of semantic systems in particular, the scope of this study holds out interesting possibilities. Given that the entire network of functional packages in a free-software operating system can be construed to be a cognitive (albeit non-autonomous) system, its general character can help construct a model that can shed light on much more complex but realistic autonomous cognitive systems, such as the human society or even the human mind (with all their creative possibilities that may unfold eventually). Of course, it is easy to appreciate that there can be no single approach to these highly intricate issues, but this very diversity of modelling, as it frequently happens while studying the development of complex structures [61], can indicate a precise direction to be adopted and allow for a clear view to be obtained.

Acknowledgments

The authors thank A. Basu, J. K. Bhattacharjee, C. Gershenson, A. Kumar, P. Majumdar, M. Marsili, S. Sinha, S. Spaeth, R. Stinchcombe and V. M. Yakovenko for some useful remarks. Discussions with P. Pathak, H. C. Pradhan, V. A. Singh and M. N. Vahia are also acknowledged.

[1] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).
[2] M. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton and Oxford, 2006).
[3] A. Barrat, M. Barthélemy, and A. Vespigniani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
[4] S. N. Dorogovtsev and J. F. F. Mendes, cond-mat/0106144.
[5] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[6] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[7] R. Albert, H. Jeong, and A.-L. Barabási, Nature **401**, 130 (1999).
[8] P. Richmond, S. Hutzler, R. Coelho, and P. Repetowicz, *Econophysics and Sociophysics,* Pg.131 (Eds. B. K. Chakrabarti, A. Chakraborti & A. Chatterjee) (WILEY-VCH Verlag GmBH & Co. KGaA, Weinheim, 2006).
[9] A. K. Gupta, *Econophysics and Sociophysics,* Pg.161 (Eds. B. K. Chakrabarti, A. Chakraborti & A. Chatterjee) (WILEY-VCH Verlag GmBH & Co. KGaA, Weinheim, 2006).

[10] A. Chatterjee and B. K. Chakrabarti(Eds.), *Econophysics of Markets and Business Networks* (Springer-Verlag Italia, Milano, 2007).
[11] P. R. V. Boas, F. A. Rodrigues, and L. da Fontoura Costa, arXiv:0903.3010.
[12] R. F. i Cancho, R. V. Solé, and R. Köhler, Phys. Rev. E **69**, 051915 (2004).
[13] R. F. i Cancho, Phys. Rev. E **70**, 056135 (2004).
[14] M. Steyvers and J. B. Tenenbaum, Cognitive Science: A Multidisciplinary Journal **29(1)**, 41 (2005).
[15] R. F. i Cancho, C. Janssen, and R. V. Solé, Phys. Rev. E **64**, 046119 (2001).
[16] S. Valverde, R. F. Cancho, and R. V. Solé, Europhys. Lett. **60**, 512 (2002).
[17] S. Valverde and R. V. Solé, cond-mat/0307278.
[18] D. Challet and A. Lombardoni, Phys. Rev. E **70**, 046109 (2004).
[19] N. LaBelle and E. Wallingford, cs/0411096.
[20] D. Challet and Y. L. Du, International Journal of Reliability, Quality and Safety Engineering **12**, 521 (2005).
[21] D. Challet and S. Valverde, arXiv:0802.3170.
[22] J. Pearl and T. Verma, AAAI-87 Proceedings (1987).
[23] A. Keller and G. Kar, 2000 International Conference on Parallel and Distributed Processing Techniques and Applications (2000).
[24] A. Keller, U. Blumenthal, and G. Kar, Proceedings of the Fifth IEEE Symposium on Computers and Communications (2000).
[25] L. Cox, H. Delugach, and D. Skipper, Proceedings of the 9th. International Conference on Conceptual Structures (2001).
[26] T. Maillart, D. Sornette, S. Spaeth, and G. von Krogh, Phys. Rev. Lett. **101**, 218701 (2008).
[27] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley Press Inc., Cambridge, Massachusetts, 1949).
[28] M. Marsili and Y.-C. Zhang, Phys. Rev. Lett. **80**, 2741 (1998).
[29] X. Gabaix, The Quarterly Journal of Economics **114**, 739 (1999).
[30] Y. Dover, Physica A **334**, 591 (2004).
[31] K. Gangopadhyay and B. Basu, Physica A **388**, 2682 (2009).
[32] N. Yadav, H. Joglekar, R. P. N. Rao, M. N. Vahia, I. Mahadevan, and R. Adhikari, arXiv:0901.3017.
[33] A.-L. Barabási and H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).
[34] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics* (Cambridge University Press, Cambridge, 2000).
[35] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[36] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Addison-Wesley Publishing Company, Reading, MA, 1994).
[37] E. W. Montroll, Proceedings of the National Academy of Science of the USA **75**, 4633 (1978).
[38] M. Braun, *Differential Equations and Their Applications* (Springer Verlag, New York, 1983).
[39] D. W. Jordan and P. Smith, *Nonlinear Ordinary Differential Equations* (Oxford University Press, Oxford, 1999).
[40] A. Drăgulescu and V. Yakovenko, Eur. Phys. J. B **17**, 723 (2000).
[41] A. Drăgulescu and V. Yakovenko, Eur. Phys. J. B **20**, 585 (2001).
[42] A. C. Silva and V. Yakovenko, Europhys. Lett. **69(2)**, 304 (2005).
[43] A. Chatterjee and B. K. Chakrabarti, Eur. Phys. J. B **60**, 135 (2007).
[44] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).
[45] S. H. Strogatz, Nature **410**, 268 (2001).
[46] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54(4)**, 436 (2001).
[47] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[48] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, and H. E. Stanley, Eur. Phys. J. B **3**, 139 (1998).
[49] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 5305 (1999).
[50] R. K. Pan and S. Sinha, Physica A **387**, 2055 (2008).
[51] C. R. Myers, Phys. Rev. E **68**, 046116 (2003).
[52] A. A. Gorshenev and Y. M. Pis'mak, Phys. Rev. E **70**, 067103 (2004).
[53] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[54] P. L. Krapivsky, S. Redner, and F. Levyraz, Phys. Rev. Lett. **85**, 4629 (2000).
[55] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[56] P. L. Krapivsky and S. Redner, Phys. Rev. E **71**, 036118 (2005).
[57] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[58] W. E. Williams, *Partial Differential Equations* (Clarendon Press, Oxford, 1980).
[59] L. Debnath, *Nonlinear Partial Differential Equations for Scientists and Engineers* (Birkhäuser, Boston, 1997).
[60] S. Valverde and R. V. Solé, Europhys. Lett. **72(5)**, 858 (2005).
[61] A. J. Koch and H. Meinhardt, Rev. Mod. Phys. **66**, 1481 (1994).