SCIENTIFIC REPORTS

natureresearch

OPEN

Generalised thresholding of hidden variable network models with scale-free property

Sámuel G. Balogh[1], Péter Pollner[2] & Gergely Palla[2]

The hidden variable formalism (based on the assumption of some intrinsic node parameters) turned out to be a remarkably efficient and powerful approach in describing and analyzing the topology of complex networks. Owing to one of its most advantageous property - namely proven to be able to reproduce a wide range of different degree distribution forms - it has become a standard tool for generating networks having the scale-free property. One of the most intensively studied version of this model is based on a thresholding mechanism of the exponentially distributed hidden variables associated to the nodes (intrinsic vertex weights), which give rise to the emergence of a scale-free network where the degree distribution $p(k) \sim k^{-\gamma}$ is decaying with an exponent of $\gamma = 2$. Here we propose a generalization and modification of this model by extending the set of connection probabilities and hidden variable distributions that lead to the aforementioned degree distribution, and analyze the conditions leading to the above behavior analytically. In addition, we propose a relaxation of the hard threshold in the connection probabilities, which opens up the possibility for obtaining sparse scale free networks with arbitrary scaling exponent.

Network theory provides an ubiquitous and sophisticated approach for the characterization of complex systems, possibly composed of many interacting units[1,2]. One of the most widely studied features of complex networks is given by the scale-free (SF) property, manifesting in the strong inhomogeneity of the degree distribution, $p(k)$, accompanied by a power-law like decay of $p(k)$ in the large degree regime[3–7]. On the modelling ground, several growing mechanisms have been proposed for generating networks without a characteristic degree scale, including the fundamental concept of the Barabási-Albert model together with its modifications and generalizations[3,8,9]. Meanwhile it became also evident that not all networks emerge from a growing mechanism[10–12], and there are numerous examples where new connections can easily occur between already existing nodes in the system[13,14]. In some cases we can also reasonably assume that the rewiring of the network leads towards a more optimal configuration[15–17], and that the propensity for creating links is encoded in each node as an intrinsic parameter. Assumptions of this type naturally gave rise to the development of the hidden variable formalism.

Inspired by the nature of protein interaction networks, the hidden variable model was originally introduced by Caldarelli et al. for explaining the emergence of non-growing scale-free networks, where link creation might be related to some intrinsic features of the nodes[10]. In a following work, Boguñá et al. proposed a systematic analytic framework for generally characterizing classes of random graphs generated with hidden variables[18]. Based on this framework, later on a general method was implemented, capable of producing SF networks with a tunable γ scaling exponent[6]. Since then, the applicability of the hidden variable formalism has been confirmed on a large scale[19–26], and due to its very general nature, a large variety of further network models ranging from stochastic block models[27–31] through multifractal graph generators[32] and evolving, fitness based network models combined with preferential attachment[33,34] to networks defined over hidden metric spaces[35–39] can be alternatively interpreted as special forms of this approach.

Although many different variations of this model has been proposed, the basic idea of the concept is to first associate a parameter (hidden variable) to the nodes, usually drawn from a prescribed distribution, and then connect the pairs of nodes according to a probability given by a fixed linking function taking the node variables as arguments. A part of these models can be referred to as geographical, where besides the hidden variables, node also have coordinates in a d dimensional Euclidean space, and the connection function is depending on both the

[1]Department of Biological Physics, Eötvös University, H-1117, Budapest, Hungary. [2]MTA-ELTE Statistical and Biological Physics Research Group, Hungarian Academy of Sciences, H-1117, Budapest, Hungary. Correspondence and requests for materials should be addressed to S.G.B. (email: balogh@hal.elte.hu)

distance and the hidden variables[40]. A very closely related class of models is where the coordinates are distributed in a hyperbolic space instead of a Euclidean one[36,38], which provide a very interesting direction for research on their own, due to that they can generate SF networks with a high clustering coefficient in a natural way and due to their relevance in routing problems[38].

A widely studied version of the hidden variable approach is where the linking function acts as a threshold, giving a connection probability 1 when the value of the two variables fulfill some criteria, and 0 otherwise. Non-geographical threshold models of this kind have gained considerable attention[10,21,22,38,41–43], later on being extended even to the geographical space[40]. An interesting phenomenon observed in the non-geographical thresholded model is that in the infinitely large system size limit the degree distribution seems to be universally characterized by a $\propto k^{-2}$ decay for various fitness distributions[21,22]. In the present paper we extend the previously studied families of hidden variable distributions and linking functions that fall into this class, and study the mathematical conditions leading to this specific degree decay exponent analytically. In addition, we also introduce a simple and intuitive relaxation of the former 'hard' threshold functions, that allows the modification of the exponent according to numerical simulations.

The Hidden Variable Model

When generating a network with N nodes in this approach, first we need to assign a hidden variable $\{x_i \mid x_i \geq 0\ \forall i \in \{1, ..., N\}\}$ to each node i, where x_i is drawn from an arbitrary (but normalized) $\rho(x)$ probability distribution. For simplicity, x_i is often referred to as the fitness of node i. After distributing these intrinsic parameters we also have to define a linking function $0 \leq f(x, y) \leq 1$, based on which the connection probability between nodes i and j can be simply expressed as $p(\text{link between } i \text{ and } j) = f(x_i, x_j)$. Thus, in this model all of the information and the properties of the emerging network is completely encoded in the pre-defined form of $\rho(x)$ and $f(x, y)$. Following the continuous approximation introduced in[6,21], the expected degree of nodes with fitness x can be expressed as

$$k(x) = N \int_0^\infty f(x, y)\rho(y)\mathrm{d}y \tag{1}$$

Assuming that $k(x)$ is a monotonically increasing and invertible function of x, according to the rule of transformation of random variables[6] the degree distribution can be written as

$$p(k) = \int \rho(x)\ \delta(k - k(x))\mathrm{d}x = \rho(x)\frac{\mathrm{d}x}{\mathrm{d}k(x)}\bigg|_{\tilde{x}(k)}, \tag{2}$$

where $\tilde{x}(k)$ denotes the inverse function of $k(x)$.

The simplest choice for the connection function is $f(x, y) = \text{const.}$, where the connection probability is uniform and independent from the hidden variables, leading to the emergence of an Erdős-Rényi random graph. A more interesting example was shown in[10] with a linking function $f(x, y) = \frac{xy}{x_M^2}$ and $x_M = \max_i\{x_i | i = 1, ..., N\}$ where (apart from multiplicative constants) the degree distribution inherits the form of the fitness distribution $p(k) \sim \rho\left(\frac{x_M^2}{N\langle x\rangle}k\right)$. Thus, by choosing a fitness distribution of $\rho(x) \sim x^{-\gamma}$, the obtained network certainly displays the SF property with the same γ exponent. Later on it turned out that for a more general class of linking functions where $f(x, y)$ can be decomposed into a product such as $f(x, y) = g(x)g(y)$, the hidden variable formalism can be regarded as analytically well-treatable, and it is also able to reproduce fat-tailed degree distributions with arbitrary γ exponents[6]. The product form also implies that for randomly chosen links the degrees of the endpoints are un-correlated, thus, from the point of view of degree assortativity the obtained networks are neutral. Another surprising result in ref.[10] is connected to the case where an exponential fitness distribution $\rho(x) \sim e^{-x}$ is chosen together with an $f(x, y)$ corresponding to a threshold function

$$f(x, y) = \Theta(x + y - \Delta), \tag{3}$$

where $\Theta(x)$ refers to the Heaviside step function, and Δ is a constant with a logarithmic dependency on the system size N. Under these settings a power law decay of the degree distribution was detected with a $\gamma = 2$ scaling exponent, providing the first evidence for that SF networks can be generated in this approach even with non power-law like fitness distributions. As we already mentioned in the Introduction, in later studies it was observed that the inverse square decay of the degree distribution is actually a quite general feature, that holds for various other fitness distributions as well[21,22]. Further interesting occurrences of the inverse square decay is briefly discussed in[44].

Generalized Classes of Non-Geographical Thresholded SF Hidden Variable Models with $\gamma = 2$

Model class definitions.
In this section we introduce a broad set of hidden variable models where the degree decay exponent is equal to $\gamma = 2$. A common feature of these models is that they are thresholded in the sense that the linking function $f(x, y)$ has a lower cut-off, controlled by a parameter Δ. The example in (3) is a special case of this, where $f(x, y)$ immediately becomes 1 above the threshold. Here we use a much weaker assumption, namely that $f(x, y)$ is 0 for a certain range of x and y values, and is non-zero (but not necessarily 1) outside this range. This is a far more general way of thresholding the connection probabilities, which allows a very broad range of connection functions to be used in the model, as shall be shown later. The linking functions having this property are denoted as $f_\Delta(x, y)$ throughout the paper. Here we introduce sub-classes of hidden variable models, the first one is to which we refer as exponential-like and where

- $\rho(x)$ can be written as

$$\rho(x) = H'(x)\exp[-H(x)], \tag{4}$$

where $H(x)$ is a differentiable, monotonously increasing function and $H'(x)$ denotes its derivative,
- while the thresholded $f_\Delta(x, y)$ shows an additive dependency on $H(x)$ and $H(y)$,

$$f_\Delta(x, y) = \begin{cases} 0, & \text{if } H(x) + H(y) \leq \Delta, \\ \widetilde{f}\,(H(x) + H(y)) \in [0, 1], & \text{if } H(x) + H(y) > \Delta, \end{cases} \tag{5}$$

where $\widetilde{f}\,(x, y)$ is assumed to be a general function taking values in the [0, 1] interval. We refer to the second sub-class as power-like, where
- $\rho(x)$ can be written as

$$\rho(x) = G'(x)G^{-\alpha}(x), \tag{6}$$

where $G(x)$ has the same properties as $H(x)$ in (4),
- while the thresholded $f_\Delta(x, y)$ shows a multiplicative dependency on $G(x)$ and $G(y)$,

$$f_\Delta(x, y) = \begin{cases} 0, & \text{if } G(x)G(y) \leq \Delta, \\ \widetilde{f}\,(G(x)G(y)) \in [0, 1] & \text{if } G(x)G(y) > \Delta. \end{cases} \tag{7}$$

And last, if both additive and multiplicative dependency are present (mixed class):

- $\rho(x)$ can be written as

$$\rho(x) = \frac{M'(x)}{(1 + M(x))^\alpha}, \tag{8}$$

where $M(x)$ has the same properties as $H(x)$ in (4),
- while $f_\Delta(x, y)$ can be expressed as:

$$f_\Delta(x, y) = \begin{cases} 0, & \text{if } (1 + M(x))(1 + M(y)) \leq \Delta, \\ \widetilde{f}\,(1 + M(x) + M(y) + M(x)M(y)) \in [0, 1] & \text{if } (1 + M(x))(1 + M(y)) > \Delta. \end{cases} \tag{9}$$

These generalized sub-classes can be derived from two simple observations. First, it can be shown in general that when replacing the step-like function in (3) by an arbitrary $f_\Delta(x + y)$ having a lower cut-off at Δ, the degree distribution of the emerging networks is not affected. Second, the form of the hidden variable distributions and accompanying connection functions given in (4–9) are also in very close relation with the transformations of the (ρ, f) pair that leave the degree distribution of the generated network invariant. To see that, let us assume an arbitrary $\rho(x)$ and $f(x, y)$ yielding a network with a degree distribution of $p(k)$. By transforming the hidden variables using a monotonous function H as $x_i = H(z_i)$ and $z_i = H^{-1}(x_i)$ for all nodes i, according to the rule of transforming random variables the density of the original variable x can be also written as $\rho_x(x) = \rho_z(z)/H'(z) = \rho_z(H^{-1}(x))/H'(H^{-1}(x))$. Based on that, the expected degree for nodes with variable x given in (2) can be also expressed as

$$k(x) = k(H(z)) = N \int_0^\infty f(x, y)\rho_x(y)\mathrm{d}y = N \int_0^\infty f(H(z), H(z'))\rho_z(z')\mathrm{d}z', \tag{10}$$

where we have changed the integration variable y to $z' = H^{-1}(y)$. By combining the expression of (10) with the transformation rule of degrees given in (2), we obtain that a transformed model where $\rho_z(z) = \rho_x(H(z))H'(z)$ and the linking function is given by $f(H(x), H(y))$, will essentially lead to the same degree distribution as the original model. This conservation law of the degree distribution is also closely related to a general transformation rule of the fitness values written as

$$p(k) = \frac{\mathrm{d}H(z)}{\mathrm{d}z}\rho(H(z))\frac{\mathrm{d}x}{\mathrm{d}\int_0^\infty f(H(z), H(z'))\frac{\mathrm{d}H(z')}{\mathrm{d}z'}\rho(H(z'))\mathrm{d}z'}\Bigg|_{\widetilde{z}_H(k)}, \tag{11}$$

which is analogous to (2). The sub-classes of models we defined in this paper exploit this property, where the 'original' model is corresponding to the simple model introduced in[10], following the inverse square decay law. Nevertheless, this invariance of the degree distribution under appropriate simultaneous transformation of $\rho(x)$ and $f(x, y)$ is valid for the hidden variable approach in general, also in the case of geographical models. We also note that any model in one of the above defined three sub-classes can generally be mapped into a model in the other sub-class via simple transformations between $H(x)$, $G(x)$ and $M(x)$ given by

$$H(x) = \alpha \ln G(\widetilde{x}) \quad H(x) = \alpha \ln(1 + M(\widetilde{x})) \quad G(x) = 1 + M(\widetilde{x}). \tag{12}$$

$\rho(x)$	Name of the distribution	$f_\Delta(x, y)$
Exponential-like $\rho(x)$, additive $f_\Delta(x, y)$		
e^{-x}	Exponential	$f_\Delta(x+y)$
$cx^{c-1}e^{-x^c}$	Weibull	$f_\Delta(x^c+y^c)$
$b\eta e^{bx}\exp(-\eta e^{bx})$	Gompertz	$f_\Delta(\eta e^{bx}+\eta e^{by})$
Power-like $\rho(x)$, multiplicative $f_\Delta(x, y)$		
$Cx^{-\alpha}$	Power	$f_\Delta(xy)$
$cx^{c-1}e^{-x^c}$	Weibull	$f_\Delta(\exp(x^c)\exp(y^c))$
Mixed $\rho(x)$, mixed $f_\Delta(x, y)$		
$\dfrac{e^{\frac{x-\mu}{s}}}{\left(1+e^{\frac{x-\mu}{s}}\right)^2}$	Logistic	$f_\Delta\left(\left(1+e^{\frac{x-\mu}{s}}\right)\left(1+e^{\frac{y-\mu}{s}}\right)\right)$
$\dfrac{\frac{\eta}{c}\left(\frac{x}{c}\right)^{\eta-1}}{\left(1+\left(\frac{x}{c}\right)^\eta\right)^2}$	Fisk	$f_\Delta\left(\left(1+\left(\frac{x}{c}\right)^\eta\right)\left(1+\left(\frac{x}{c}\right)^\eta\right)\right)$

Table 1. Examples for fitness distributions with the accompanying thresholded fitness functions $f_\Delta(x, y)$ for each sub-class. The listed models all lead to the emergence of SF networks satisfying the inverse square decay law.
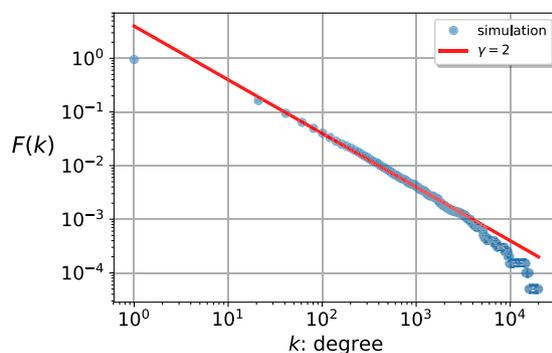
Figure 1. Complementary cumulative distribution of the node degrees $F(k)$ in a network of size $N = 20000$, obtained from simulations for the Weibull fitness distribution in Table 1, at a scale parameter $c = 2$, and a linking function defined in (13), shown on logarithmic scale. The solid line is decreasing as k^{-1}, which is corresponding to the decay characteristics of $F(k)$ in SF networks with $\gamma = 2$.

However, when the goal is to obtain a size independent SF degree distribution, the dependency of $\Delta = \Delta(N)$ on the number of nodes N can be different in each class. In summary, the previous observations clarify in a simple manner why seemingly different realizations of (ρ, f) can give rise to the emergence of networks with the same degree distributions. In addition, we also gained simple rules for mapping the different realisations of (ρ, f) into each other. A remarkable consequence of the above is that for any $\widetilde{f}_\Delta$ showing either additive, multiplicative or mixed dependency on its arguments, we can now construct a fitness distribution for which it is guaranteed that the degree distribution of the emerging network will display the $p(k) \sim k^{-2}$ behaviour.

For illustration, a few examples from each sub-classes are listed in Table 1, all generating SF networks with a degree decay exponent $\gamma = 2$. In addition, in Fig. 1. we show the simulation results for the Weibull fitness distribution from the class of exponential-like distributions, where the corresponding linking function was chosen as

$$f_\Delta(x, y) = \begin{cases} 0 & \text{if } x^2 + y^2 \leq \Delta, \\ 1 - \dfrac{\exp[a - (x^2 + y^2)]}{(x^2 + y^2)} & \text{if } x^2 + y^2 > \Delta, \end{cases} \tag{13}$$

where a is a constant and Δ is defined via the transcendent equation $\Delta = \exp(a - \Delta)$. The fact that the complementary cumulative distribution $F(k) = \int_{k'=k}^{\infty} p(k')\mathrm{d}k'$ of the degrees behaves as k^{-1} in the large degree regime in Fig. 1. is in full consistency with the inverse square decay law.

The inverse square decay law. Here we show in details that for thresholded hidden variable models falling into the class described in the previous subsection, the degree distribution of the emerging SF network will always have a degree decay exponent of $\gamma = 2$. Let us assume that we are dealing with an exponential-like model, where the fitness distribution is given in (4), and the linking function follows (5). Starting from the expression for the average degree given in (2), and multiplying both sides by $\exp[-H(x)]$ we can write

$$k(x) \exp[-H(x)] = N \int_{H^{-1}(\Delta - H(x))}^{\infty} \widetilde{f}(H(x) + H(y))H'(y) \exp[-H(x) - H(y)]\mathrm{d}y, \tag{14}$$

where we used that $f_\Delta(x, y) = 0$ if $H(x) + H(y) \leq \Delta$. By a change of variable $z = H(x) + H(y)$ we arrive to an equation where the right hand side is independent of x,

$$k(x) \exp[-H(x)] = N \int_{\Delta}^{\infty} \widetilde{f}(z)\mathrm{e}^{-z}\mathrm{d}z. \tag{15}$$

Based on (15) we define the integral

$$L_\Delta(\widetilde{f}) \equiv \int_{\Delta}^{\infty} \widetilde{f}(z)\mathrm{e}^{-z}\mathrm{d}z, \tag{16}$$

that depends on the form of the actually chosen $\widetilde{f}(z)$ appearing in (5). Assuming that this integral exists, using (15–16) we can express the average degree of nodes with fitness x and the derivative of $k(x)$ as

$$k(x) = NL_\Delta(\widetilde{f}) \exp[H(x)], \tag{17}$$

$$\frac{\mathrm{d}k(x)}{\mathrm{d}x} = NL_\Delta(\widetilde{f}) \exp[H(x)]H'(x). \tag{18}$$

In the thermodynamic limit of $N \to \infty$, by substituting (17–18) into (2) we obtain

$$p(k) = \frac{H'(x) \exp[-H(x)]}{NL_\Delta(\widetilde{f})H'(x) \exp[H(x)]}\bigg|_{\widetilde{x}_H(k)} = \frac{NL_\Delta(\widetilde{f})}{[NL_\Delta(\widetilde{f}) \exp[H(x)]]^2}\bigg|_{\widetilde{x}_H(k)} = \frac{NL_\Delta(\widetilde{f})}{k^2}, \tag{19}$$

showing that $p(k)$ is proportional to k^{-2}, since $L_\Delta(\widetilde{f})$ is a constant. Moreover, according to (19) we can formulate a very simple condition under which $p(k)$ becomes independent of the system size in the form of

$$L_\Delta(\widetilde{f}) = L(\Delta, \widetilde{f}) = \int_{\Delta}^{\infty} \widetilde{f}(z)\mathrm{e}^{-z}\mathrm{d}z = \frac{1}{N}, \tag{20}$$

which is reducing to $L(\Delta, \Theta) = \mathrm{e}^{-\Delta} = \frac{1}{N}$ in the case of the simple step-function like connection function given in (3), in complete agreement with the results in[6,18].

Similarly to the case of exponential-like distributions, if we choose the fitness distribution to be power-like as in (6), together with a thresholded connection function given in (7), the expected degree for a node having a hidden variable x can be written as

$$k(x) = N \int_{1}^{\infty} f(x, y)\rho(y)\mathrm{d}y = N \int_{G^{-1}(\Delta/G(x))}^{\infty} \widetilde{f}(G(x)G(y))G^{-\alpha}(y)G'(y)\mathrm{d}y. \tag{21}$$

By changing to the integration variable $z = G(x)G(y)$ we obtain

$$k(x)G^{-\alpha+1}(x) = N \int_{\Delta}^{\infty} \widetilde{f}(G(x)G(y))G^{-\alpha}(x)G^{-\alpha}(y)\mathrm{d}[G(x)G(y)], \tag{22}$$

leading to

$$k(x) = NG^{\alpha-1}(x) \int_{\Delta}^{\infty} \widetilde{f}(z)z^{-\alpha}\mathrm{d}z, \tag{23}$$

where $\alpha > 1$. The integral

$$K_\alpha(\widetilde{f}) \equiv \int_{\Delta}^{\infty} \widetilde{f}(z)z^{-\alpha}\mathrm{d}z \tag{24}$$

appearing in (23) depends only on the chosen form of $\widetilde{f}(z)$, thus, by assuming that $K_\alpha(\widetilde{f})$ exists we can express $k(x)$ simply as

$$k(x) = NK_\alpha(\widetilde{f})G^{\alpha-1}(x). \tag{25}$$

By substituting this into the general formula for the degree distribution given in (2) we gain

$$p(k) = \rho(x)\frac{dx}{dk(x)}\bigg|_{\widetilde{x}_G(k)} = \frac{G'(x)G^{-\alpha}(x)}{N(\alpha - 1)K_\alpha(\widetilde{f})G'(x)G^{\alpha-2}(x)}\bigg|_{\widetilde{x}_G(k)} \sim \frac{1}{G^{2(\alpha-1)}(x)}\bigg|_{\widetilde{x}_G(k)} \sim k^{-2}, \tag{26}$$

proving that the power-like sub-class also leads to the emergence of a SF network with $\gamma = 2$. However, in this case the condition for a size independent degree distribution with the $p(k) \sim k^{-2}$ property is given by a different formula compared to (20), written as

$$K_\alpha(\widetilde{f}) \equiv \int_\Delta^\infty \widetilde{f}(z)z^{-\alpha}\mathrm{d}z = \frac{1}{N}. \tag{27}$$

By following similar mathematical arguments as we did in the exponential- and power-like cases the same behaviour can be obtained for the third sub-class defined through (8–9). Nevertheless, the condition for a size independent SF degree distribution with $\gamma = 2$ has yet again, a different form from (20) or (27), given by

$$J_\alpha(\widetilde{f}) \equiv \int_\Delta^\infty \frac{\widetilde{f}(z)}{(1+z)^\alpha}\mathrm{d}z = \frac{1}{N}. \tag{28}$$

An important related remark is that for all fitness distributions and linking functions given in the forms of (4), (5) or (6), (7) or (8), (9) the emerging networks always display the inverse square decay law independent from the specific form of $\widetilde{f}$, and thus, the appropriate form of $\widetilde{f}$ only determines the condition for obtaining size-independent degree distribution.

Soft Thresholding with Tunable Degree Decay Exponent

The degree decay exponent of SF complex networks characterizing real systems is usually between $\gamma = 2$ and $\gamma = 3$[1,2]. Motivated by that, we extend the models defined in the previous section to allow the emergence of hidden variable networks with a $\gamma > 2$ exponent as well in the same framework.

The basic idea is to relax the 'hard' threshold in the connection function, controlled by the variable Δ in (3) and (5). Let us start with the Heaviside step function given in (3), which we can intuitively replace by a 'reversed' Fermi-Dirac function

$$f(x, y) = f(x + y) = \frac{1}{1 + e^{-\beta(x+y-\Delta)}}, \tag{29}$$

where β is a parameter taking positive values. Naturally, at $\beta \to \infty$ we recover the original step-like connection function (3), whereas for finite β values we obtain a 'soft' threshold function. The form of $f(x, y)$ in (29) is similar to that of the linking probability in temperature dependent graph ensembles[45,46]. The interpretation of β in this respect is analogous to the inverse temperature, with $\beta \to \infty$ corresponding to the zero temperature 'ground' state, while networks generated with finite β values can be interpreted as states at higher temperatures[36,45]. Regarding the case where we can not assume the additive but rather the multiplicative dependency on x and y in (29), a simple approach to relax the step function is to define $f(x, y)$ as

$$f(x, y) = f(xy) = \frac{1}{1 + \left(\frac{xy}{\Delta}\right)^{-\beta}}, \tag{30}$$

converging to $f(x, y) = \Theta(xy - \Delta)$ in the $\beta \to \infty$ limit.

In a similar fashion to (29), for the general exponential-like models defined in (4–5) we can replace (5) by

$$f_{\beta,\Delta}(x, y) = \frac{1}{1 + e^{-\beta(H(x)+H(y)-\Delta)}}, \tag{31}$$

and for the power-like models given in (6–7) we can change (7) to

$$f(x, y) = f_{\beta,\Delta}(x, y) = \frac{1}{1 + \left(\frac{G^\alpha(x)G^\alpha(y)}{\Delta}\right)^{-\beta}}. \tag{32}$$

Similar form of connection function can be established for the third sub-class given by (8–9) based on (12). For all β dependent connection functions defined above, at $\beta \to 0$ we obtain a linking kernel that becomes independent of the hidden variables, and thus, the generated network is an Erdös-Rényi random graph. However, in the opposite case, when $\beta \to \infty$, the connection functions converge to the original 'hard' thresholded forms, and the generated networks are scale-free with a decay exponent of $\gamma = 2$. Therefore, by tuning the parameter β from 0 to ∞ we can scan through a series of networks starting from the classical random graph, and arriving to a SF network obeying the inverse square decay law at the other end of the spectrum. Presumably, during this transition we may find a finite β range in which the degree distribution is already power law like instead of the Poisson distribution, but the decay exponent γ has not yet reached the $\gamma = 2$ limit value.

Our simulation results shown in Fig. 2. provide a strong support for the assumption above. In the four panels we display the complementary cumulative degree distribution for an exponential-like model with $\rho(x) = 3x^2e^{-x^3}$ at different β parameters.

Characterizing the $\gamma(\beta)$ transition. In this section we aim to characterize the main features of the above mentioned transition of γ as a function of the effective temperature for the exponential case. In order to do so, we first define the general β dependent integral for $k(x)$ and perform a transformation of variables ($z = H(x) + H(y) - \Delta$):
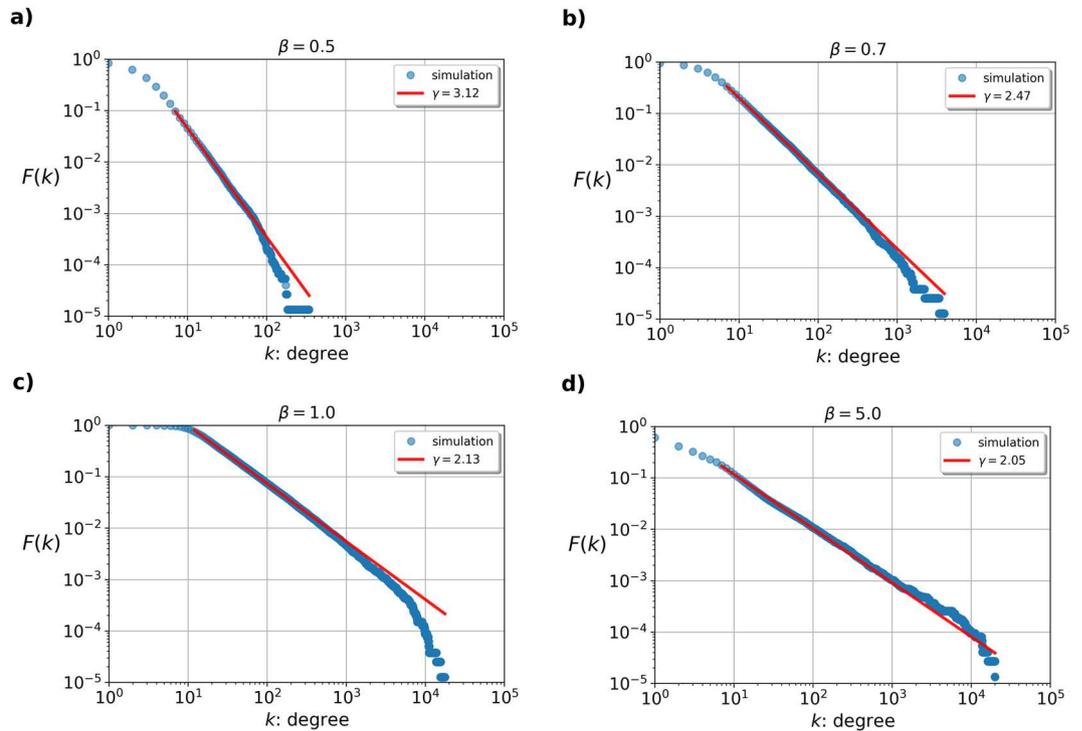
Figure 2. The complementary cumulative distribution of the node degrees $F(k)$ for four different networks, generated with the same $\rho(x)$ and $f(x, y)$, but with different β parameters. The fitness distribution was chosen to be $\rho(x) = 3x^2\exp(-x^3)$ and corresponding linking function $f(x, y)$ is given in (31). In panel (a) we used $\beta = 0.5$, and the decay characteristics of the resulting $F(k)$ seem to be close to that of SF networks with $\gamma \simeq 3.12$ (shown by the solid line). The parameter β was increased to $\beta = 0.7$ in panel (b), where the decay of $F(k)$ suggests a γ value of $\gamma \simeq 2.47$. In panels (c,d) we increased β further to $\beta = 1.0$ and $\beta = 5.0$, reducing the γ exponent to $\gamma = 2.13$ and $\gamma = 2.05$ respectively. In panel (a,b) $\Delta = \frac{1}{\beta}\ln N$, while in panel (c,d) we used $\Delta = \ln N$ for obtaining sparse networks.

$$k(x) = N\int_0^\infty \frac{H'(y)e^{-H(y)}dy}{1 + e^{-\beta(H(x)+H(y)-\Delta)}} = Ne^{H(x)-\Delta}\int_{H(x)-\Delta}^\infty \frac{e^{-z}}{1 + e^{-\beta z}}dz = Ne^{H(x)-\Delta}F_{\beta,\Delta}(x) \tag{33}$$

offering a non-invertible form for the degree variable in general. However, approximate results can still be obtained possibly with logarithmic or sub-power corrections. If $H(x) \ll \Delta$ and $\beta \in (0, 1)$, the main contribution to the integral comes from the $z \ll 0$ range, where the value of the denominator is large. Based on that, the above expression can be approximated by

$$k(x) = Ne^{H(x)-\Delta}\int_{H(x)-\Delta}^\infty \frac{e^{-z}}{1 + e^{-\beta z}}dz \approx Ne^{H(x)-\Delta}\int_{H(x)-\Delta}^\infty e^{(\beta-1)z}dz = \frac{N}{1-\beta}e^{\beta(H(x)-\Delta)}. \tag{34}$$

Even though this approximation tends to be less and less accurate as the value of β converges to zero, its advantage is that it provides an analytic and invertible expression for the degree distribution written as

$$p_\beta(k) \sim k^{-(1+\frac{1}{\beta})} + \text{subpower corrections.} \tag{35}$$

We note that similar forms have already been established in refs[36,45], however, not for the reversed Fermi-Dirac function. In addition to that, based on (34) we can also formulate an approximate condition for having a size independent degree distribution given by

$$\Delta_\beta(N) \simeq \frac{1}{\beta}\ln N. \tag{36}$$

For $\beta > 1$ the integrand appearing in (33) becomes negligible for $z = H(x) - \Delta < 0$, hence the lower bound of the integral can approximately be replaced by zero as

$$k(x) \approx Ne^{H(x)-\Delta}\int_0^\infty \frac{e^{-z}}{1 + e^{-\beta z}}dz, \tag{37}$$
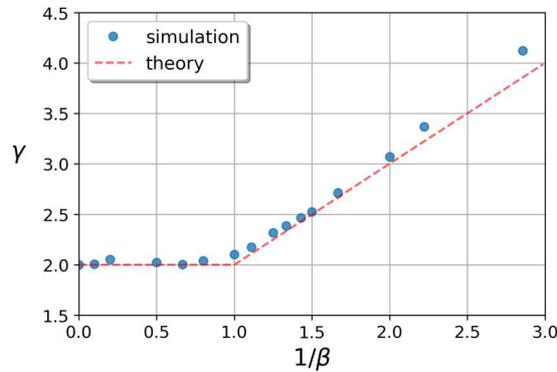
Figure 3. Scaling exponent γ of the degree distribution as a function of the effective temperature $1/\beta$ in the model with soft thresholding. The data points correspond to simulation results on networks of size $N = 20000$, where the fitness distribution was chosen to be $\rho(x) = 3x^2 \exp(-x^3)$ and the linking function $f(x, y)$ was given by (31). The dashed line shows the (approximate) analytic results given by (35) and (38).

which along similar arguments is yielding

$$p_\beta(k) \sim k^{-2} + \text{subpower corrections} \quad \text{and} \quad \Delta_\beta(N) \simeq \ln N. \tag{38}$$

The results in (38) become exact in the limit of $\beta \to \infty$. Analogously, the above analysis applies to the power sub-class as well, where the average degree as a function of fitness is written as

$$k(x) = N \int_1^\infty \frac{G'(y)G^{-\alpha}(y)}{1 + \left(\frac{G^\alpha(x)G^\alpha(y)}{\Delta}\right)^{-\beta}} dy = N G^{\alpha-1}(x)\Delta^{-\alpha+1} \int_{G(x)/\Delta}^\infty \frac{z^{-\alpha}}{1 + z^{-\alpha\beta}} dz, \tag{39}$$

with $z = \frac{G(x)G(y)}{\Delta}$. Despite the same behaviour of the degree distribution, the formula above suggests that the condition for obtaining a size independent degree distribution requires $\Delta^{\alpha\beta}$ to be proportional to N (for $\beta \in (0, 1)$). Furthermore, it also implies that the accurate characterization of the $\gamma(\beta)$ transition for the three sub-classes requires similar considerations.

Our simulation results together with the approximation discussed above are shown in Fig. 3., depicting the transition of the scaling exponent as a function of the effective temperature $1/\beta$. We kept control over the average degree of the generated networks by relying on (36) and (38), ensuring the size independence of the degree distribution.

Discussion

We have revisited the inverse square decay law of non-geographical thresholded hidden variable models, which was already studied in refs[10,21,22] from different perspectives. We provided a far more general framework for thresholding linking mechanisms, where the form of the connection function $f(x, y)$ is not restricted to the usual Heaviside function, but instead can correspond to any general function with values in the [0,1] interval, as long as $f(x, y) = 0$ for a certain range of x and y values, and is non-zero outside this range. According to our results, this considerably weaker assumption on the form of the thresholded $f(x, y)$ allows a very broad range of connection functions, that combined with properly chosen fitness distributions $\rho(x)$ result in the inverse square decay law, similarly to the models discussed in ref.[10]. Along this line we provided three general sub-classes of hidden variable distributions and accompanying connection functions (i.e., the exponential, the power-like and the mixed class) that all generate SF networks with a degree decay exponent of $\gamma = 2$, and we also discussed how these different sub-classes are interrelated to each other.

Despite the invariance of the degree distribution obtained by using different $f(x, y)$, the generated networks might show different properties at the level of local network quantities such as degree correlations or the clustering coefficient. For illustration in Fig. 4. we provide simulation results for the clustering coefficient as a function of k, when replacing the Heaviside step function with other possible forms of $f_\Delta(x + y)$. According to Fig. 4a, the average clustering coefficient $\overline{C}$ is more or less constant below a characteristic degree and is decaying as a power law above this characteristic degree for the Heaviside step function (red symbols), and also for a connection function $f_\Delta(x + y)$ converging to the Heaviside step function in an exponential manner (green symbols). In contrast, for connection functions having a peak at $x + y = z = \Delta(N)$ with a decaying tail for larger z values (green and yellow symbols), we can observe a peak in $\overline{C}(k)$, together with a faster (but still power law like) decay in the large degree regime.

We also proposed a relaxation of the 'hard' threshold for each sub-class imposed by the Δ controlled boundary in the connection functions. The basic idea was to use a 'reversed' Fermi-Dirac function providing a sigmoid transition between low and high linking probability values, where the sharpness of the transition (or in other words, the width of the intermediate linking probability values) is controlled by a parameter β. Based on analogy with temperature dependent graph ensembles[45], β can be interpreted as a sort of inverse temperature, where the original 'hard' thresholded models are recovered in the zero temperature limit of $\beta \to \infty$. The great advantage of
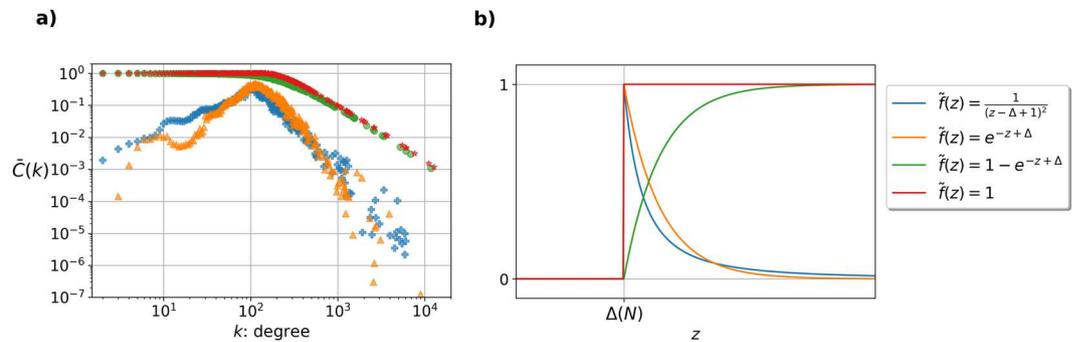
Figure 4. (**a**) Clustering coefficient as a function of node degrees $\overline{C}(k)$ for four different networks each of them containing $N = 20000$ nodes. All networks were generated by using the same exponential fitness distribution but with different connection functions $f_\Delta(x+y) = f_\Delta(z)$ corresponding to specific forms of (5) displayed by different colours and indicated in panel (b).

the higher temperature models is that according to numerical simulations, the degree decay exponent becomes larger than $\gamma = 2$, and by changing β, it can be tuned to any preferred value in the range of typical γ values measured in real systems. We also generally discussed the criteria in multiple different cases of how to generate networks having degree distribution independent of the size. Hence, the models with the relaxed threshold at finite β values offer a flexible fitness-based approach being adjustable to complicated fitness distributions for generating sparse SF networks with realistic degree decay exponent.

In conclusion, our analysis showed that linking kernels with a general lower-cutoff and having either additive, multiplicative or mixed dependence on their arguments can always generate SF networks together with the appropriately chosen fitness distributions. A further remarkable consequence of the above is that a general mapping can be established between different ρ fitness distributions and possible f linking functions. I.e., for any fitness distribution ρ^* in general there exists a family of thresholded linking functions f_Δ^* that together give rise to scale-free networks with a $\gamma = 2$, and vice versa, for any thresholded linking function f_Δ^* we can find the corresponding fitness distribution ρ^* together which they display the same property. This might provide an alternative way of understanding how those fitness/activity driven systems exhibit SF behaviour where the distributions of the hidden variables follow non-trivial, complicated forms.

References

1. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
2. Mendes, J. F. F. & Dorogovtsev, S. N. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, Oxford, 2003).
3. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Sci.* **286**, 509–512 (1999).
4. Newman, M. E. J. Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E* **64**, 016131 (2001).
5. He, B. J. Scale-free brain activity: past, present, and future. *Trends Cogn. Sci.* **18**, 480–487 (2014).
6. Servedio, V. D. P., Caldarelli, G. & Buttà, P. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Phys. Rev. E* **70**, 056126 (2004).
7. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005).
8. Bianconi, G. & Barabási, A.-L. Competition and multiscaling in evolving networks. *Eur. Lett.* **54**, 436 (2001).
9. Dorogovtsev, S., Mendes, J. F. & Samukhin, N. A. Structure of growing networks with preferential linking. *Phys. Rev. Lett.* **85**, 4633–6 (2000).
10. Caldarelli, G., Capocci, A., De Los Rios, P. & Muñoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702 (2002).
11. Timár, G., Dorogovtsev, S. N. & Mendes, J. F. F. Scale-free networks with exponent one. *Phys. Rev. E* **94**, 022302 (2016).
12. Seyed-allaei, H., Bianconi, G. & Marsili, M. Scale-free networks with an exponent less than two. *Phys. Rev. E* **73**, 046113 (2006).
13. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'. networks. *Nat.* **393**, 440–442 (1998).
14. Bedogne', C. & Rodgers, G. J. Complex growing networks with intrinsic vertex fitness. *Phys. Rev. E* **74**, 046115 (2006).
15. Guimerà, R., Díaz-Guilera, A., Vega-Redondo, F., Cabrales, A. & Arenas, A. Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* **89**, 248701 (2002).
16. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Statistical mechanics of topological phase transitions in networks. *Phys. Rev. E* **69**, 046117 (2004).
17. Kim, J., Kim, I., Han, S. K., Bowie, J. U. & Kim, S. Network rewiring is an important mechanism of gene essentiality change. *Sci. Reports* **2**, 900 (2012).
18. Boguñá, M. & Pastor-Satorras, R. Class of correlated random networks with hidden variables. *Phys. Rev. E* **68**, 036112 (2003).
19. Masuda, N. & Konno, N. Vip-club phenomenon: Emergence of elites and masterminds in social networks. *Soc. Networks* **28**, 297–309 (2006).
20. van der Hofstad, R., Janssen, A. J. E. M., van Leeuwaarden, J. S. H. & Stegehuis, C. Local clustering in scale-free networks with hidden variables. *Phys. Rev. E* **95**, 022307 (2017).
21. Masuda, N., Miwa, H. & Konno, N. Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights. *Phys. Rev. E* **70**, 036124 (2004).
22. Fujihara, A., Uchida, M. & Miwa, H. Universal power laws in the threshold network model: A theoretical analysis based on extreme value theory. *Phys. A: Stat. Mech. its Appl.* **389**, 1124–1130 (2010).
23. Petermann, T. & Rios, D. L. P. Exploration of scale-free networks. *The Eur. Phys. J. B* **38**, 201–204 (2004).
24. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. Models of social networks based on social distance attachment. *Phys. Rev. E* **70**, 056122 (2004).

25. Starnini, M. & Pastor-Satorras, R. Topological properties of a time-integrated activity-driven network. *Phys. Rev. E* **87**, 062807 (2013).
26. Garlaschelli, D., Capocci, A. & Caldarelli, G. Self-organized network evolution coupled to extremal dynamics. *Nat. Phys* **3** (2006).
27. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Networks* **5**, 109–137 (1983).
28. Fienberg, S. E., Meyer, M. M. & Wasserman, S. S. Statistical analysis of multiple sociometric relations. *J. Am. Stat. Assoc.* **80**, 51–67 (1985).
29. Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.* **107**, 065701 (2011).
30. Aicher, C., Jacobs, A. Z. & Clauset, A. Learning latent block structure in weighted networks. *J. Complex Networks* **3**, 221–248 (2015).
31. Peixoto, T. P. Nonparametric weighted stochastic block models. *Phys. Rev. E* **97**, 012306 (2018).
32. Palla, G., Lovász, L. & Vicsek, T. Multifractal network generator. *Proc. Natl. Acad. Sci. USA* (2010).
33. Eom, Y.-H. & Fortunato, S. Characterizing and modeling citation dynamics. *Plos One* **6**, 1–7 (2011).
34. Medo, M. C. V., Cimini, G. & Gualdi, S. Temporal effects in the growth of networks. *Phys. Rev. Lett.* **107**, 238701 (2011).
35. Serrano, M. A., Krioukov, D. & Boguñá, M. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.* **100**, 078701 (2008).
36. Krioukov, D., Papadopoulos, F., Vahdat, A. & Boguñá, M. Curvature and temperature of complex networks. *Phys. Rev. E* **80**, 035101 (2009).
37. Boguñá, M., Krioukov, D. & Claffy, K. C. Navigability of complex networks. *Nat. Phys.* **5**, 74–80 (2009).
38. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguñá, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, 036106 (2010).
39. Boguñá, M., Papadopoulos, F. & Krioukov, D. Sustaining the internet with hyperbolic mapping. *Nat. Commun.* **1**, 62 (2010).
40. Masuda, N., Miwa, H. & Konno, N. Geographical threshold graphs with small-world and scale-free properties. *Phys. Rev. E* **71**, 036108 (2005).
41. Konno, N., Masuda, N., Roy, R. & Sarkar, A. Rigorous results on the threshold network model. *J. Phys. A: Math. Gen.* **38**, 6277 (2005).
42. Fujihara, A. *et al.* Limit theorems for the average distance and the degree distribution of the threshold network model. *Interdiscip. Inf. Sci.* **15**, 361–366 (2009).
43. Ide, Y., Konno, N. & Masuda, N. Statistical properties of a generalized threshold network model. *Methodol. Comput. Appl. Probab.* **12**, 361–377 (2010).
44. Caldarelli, G., Caretta Cartozo, C., De Los Rios, P. & Servedio, V. D. P. Widespread occurrence of the inverse square distribution in social sciences and taxonomy. *Phys. Rev. E* **69**, 035101 (2004).
45. Garlaschelli, D., Ahnert, S. E., Fink, T. M. A. & Caldarelli, G. Low-temperature behaviour of social and economic networks. *Entropy* **15**, 3148–3169 (2013).
46. Park, J. & Newman, M. E. J. Origin of degree correlations in the internet and other networks. *Phys. Rev. E* **68**, 026112 (2003).

Acknowledgements

Author Contributions

S.G.B., P.P. and G.P. developed the concept of the study, S.G.B. derived the equations, S.G.B., P.P., G.P. contributed to the interpretation of the results, S.G.B. prepared the table and the figures, S.G.B., P.P. and G.P. wrote the paper.

Additional Information