Complex networks as an emerging property of hierarchical preferential attachment

Laurent Hébert-Dufresne,[1,2] Edward Laurence,[1] Antoine Allard,[3] Jean-Gabriel Young,[1] and Louis J. Dubé[1]

[1]*Département de Physique, de Génie Physique, et d'Optique, Université Laval, Québec (Québec), Canada G1V 0A6*

[2]*Santa Fe Institute, Santa Fe, New Mexico 87501, USA*

[3]*Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

Real complex systems are not rigidly structured; no clear rules or blueprints exist for their construction. Yet, amidst their apparent randomness, complex structural properties universally emerge. We propose that an important class of complex systems can be modeled as an organization of many embedded levels (potentially infinite in number), all of them following the same universal growth principle known as preferential attachment. We give examples of such hierarchy in real systems, for instance, in the pyramid of production entities of the film industry. More importantly, we show how real complex networks can be interpreted as a projection of our model, from which their scale independence, their clustering, their hierarchy, their fractality, and their navigability naturally emerge. Our results suggest that complex networks, viewed as growing systems, can be quite simple, and that the apparent complexity of their structure is largely a reflection of their unobserved hierarchical nature.

I. INTRODUCTION

The science of complexity is concerned with systems displaying emerging properties: systems where the properties of the whole do not directly follow from the properties of the parts [1]. However, we intend to show how one property of the whole, *hierarchy*, can alone be the origin of more complex features. We will describe hierarchical systems through a general model of *colored balls in embedded bins* which itself explains the emergence of other features through the projection of these hierarchical systems onto complex networks.

Most real networks tend to feature properties not found in most classic models of sparse random networks: *scale independence*, fat-tailed degree distribution [2,3]; *modularity*, the grouping of nodes in denser groups [4–6]; *hierarchy*, the embedding of multiple levels of organization [7,8]; *fractality*, the self-similarity between levels of organization [9,10]; and *navigability*, the possibility of efficient communication through a hidden metric space [11–13].

Sophisticated algorithms can be designed to reproduce most of these features, often based upon a multiplicative process to force their emergence by reproducing a basic unit on multiple scale of organization [14,15]. These models are useful as they can create realistic structures and test hypotheses about measured data. However, these constructions are not intended to provide any insights on the underlying mechanisms behind the growth of the system.

In contrast, generative models are quite successful at suggesting principles of organization leading to specific properties. For example, simple models exist to propose possible origins for scale independence [3] or of the small-world effect [4], but they fail to model the emergence of properties not included by design. Consequently, the identification of new universal properties requires the creation of new generative models. It is fair to say that a single unifying principle has yet to be proposed.

In this paper, we aim to close the gap between complex deterministic algorithms and simple stochastic growth models. The hierarchical nature of networks suggests that the observed links between nodes are merely projections of higher structural units [6,8,16] (e.g., people create groups within cities in given countries). These subsystems will be our focus. We use one general assumption to design an equally general model of hierarchical systems: all embedded levels of organization follow preferential attachment.

On the one hand, our model can be seen as a generalization of classical preferential attachment models [6,16–20]. We can thus apply methods developed in this context by generalizing them to hierarchical systems. On the other hand, our model fills the gap to previous studies wishing to introduce nontrivial structural properties, such as clustering and centrality. Past models manipulate the networks through local rules to add, remove, or rewire links: for instance, triadic closure [21,22] or copying mechanisms [23,24]. We find that complex properties emerge more naturally when changing the focus of the model from the actual network and its properties to the hierarchical system that produces it.

We validate this model on the well-documented data set of production entities in the film industry (i.e., producers produce films within companies in given countries). We then study the structure of the projection of this system onto a complex network of coproductions between film producers. Interestingly, the resulting networks feature a scale-independent hierarchical organization, community structure, fractality, and navigability.

The paper is structured as follows. In Sec. II, we provide a brief review of preferential attachment (PA), followed by an overview of structural preferential attachment (SPA) in Sec. III. In Sec. IV, we generalize this organization principle to a family of processes generating hierarchical systems of embedded structural levels. A particular process of this family is then algorithmically described in Sec. IV B and mathematically studied in Sec. IV C. In Sec. V, we explain how complex networks can be obtained from this process by projecting a hierarchical system onto a chosen structural level. Finally, our conclusions are presented in Sec. VI and a few technical details are covered in two short Appendixes.

II. PREFERENTIAL ATTACHMENT (PA)

The preferential attachment principle is a ubiquitous *rich-get-richer* mechanism modeling complex systems of all sorts [2,3,6,25–28]. It implies that the likelihood for a given entity to be involved in a new activity is roughly proportional to its total past activities. For instance, an individual with 10 acquaintances in a social network is roughly 10 times more likely to gain a new connection than one with a single acquaintance. This does not imply causation; the individual does not necessarily gain a new connection because of its existing ones, but merely that its past is a good indicator of its future activity. This simple mechanism leads to a scale-independent distribution of the activity in question, modeling any system where the distribution of a resource among a population roughly follows a power-law distribution. Consequently, the number N_s of individuals with a share s ($\in \mathbb{N}$) of the resource scales as $s^{-\gamma}$, where γ is called the scaling exponent.

In practice, we consider a discrete time process where, during a time step $\Delta t = 1$, a new element i of share $s_i = 1$ is introduced within the system with rate B (birth event) or the share s_j of an existing element j is increased to $s_j + 1$ with rate G (growth event). We can write a rate equation governing the expected number of individuals N_s with a given share s:

$$N_s(t+1) = N_s(t) + B\delta_{s,1}$$
$$+ \frac{G}{\sum s N_s(t)}[(s-1)N_{s-1}(t) - sN_s(t)], \quad (1)$$

where $\sum s N_s(t)$ is the sum of all shares (total resource) used to normalize the transition probabilities and which rapidly converges to $(B+G)t$. Consequently, we will hereafter use $\sum_s s N_s(t) = (B+G)t$ interchangeably whenever they appear. Since B is the birth rate, the evolution of the normalized distribution $\{\tilde{N}_s(t)\}$ can be obtained by replacing $N_s(t)$ by $tB\tilde{N}_s(t)$:

$$(t+1)B\tilde{N}_s(t+1) = tB\tilde{N}_s(t) + B\delta_{s,1}$$
$$+ \frac{GB}{B+G}[(s-1)\tilde{N}_{s-1}(t) - s\tilde{N}_s(t)]. \quad (2)$$

Solving at statistical equilibrium, i.e., $\tilde{N}_s(t+1) = \tilde{N}_s(t) = \tilde{N}_s^*$, yields

$$\left(1 + s\frac{G}{B+G}\right)\tilde{N}_s^* = \frac{G}{B+G}(s-1)\tilde{N}_{s-1}^* + \delta_{s,1} \quad (3)$$

or more directly for $s > 1$

$$\tilde{N}_s^* = \frac{\prod_{m=1}^{s-1} \frac{G}{B+G} m}{\prod_{m=1}^{s}\left(1 + \frac{G}{B+G}m\right)}. \quad (4)$$

Asymptotically for $s \to \infty$, this steady state can be shown to scale as a power law

$$\lim_{t,s\to\infty} \tilde{N}_s^*(t) \propto s^{-\gamma} \quad \text{with} \quad \gamma = 2 + \frac{B}{G}. \quad (5)$$

III. STRUCTURAL PREFERENTIAL ATTACHMENT (SPA)

With the ongoing focus on the modularity of complex systems, e.g., the community structure of networks, it is essential to be able to consider structural properties of real systems within preferential attachment processes. In a recent study, we have introduced colored balls to represent individuals in social systems where unique individuals (unique colors) are a resource for communities (boxes) and vice versa [6,16]. This can be mapped to a process where colored balls are placed in boxes. Balls of the same color are meant to represent different activities of the same individual, just as different boxes represent different structures growing by receiving new balls. We have extended preferential attachment to structured systems: just as an individual involved in more social groups is more likely to join a new group, a larger social group is more likely to gain new members than a small one. We have coined the name structural preferential attachment (SPA) to describe this first level of extension of the PA principle.

In SPA, the two important quantities are the membership of a given color, i.e., the number of structures in which that color is found, and the size of a given structure, i.e., the number of balls it contains. They can be followed by the rate equation approach of Eq. (1). In distinction to the previous section, we now have a first structural level, and our notation reflects this extension by an extra index on the associated quantities. In the case of memberships, the share of a ball is now the number m of apparitions in different structures, whereas in the case of sizes, the share of a structure is now the number n of balls it contains. Hence, in both cases, the total resource is given by the sum of all balls found in the system, regardless of their colors. We can thus write

$$N_{1,m}(t+1) = N_{1,m}(t) + {}^{(N)}\!B_1\,\delta_{m,1} + \frac{{}^{(N)}\!G_1}{({}^{(N)}\!B_1 + {}^{(N)}\!G_1)t}$$
$$\times [(m-1)N_{1,m-1}(t) - mN_{1,m}(t)] \quad (6)$$

for the number $N_{1,m}$ of different colors with memberships m at the first structural level. ${}^{(N)}\!B_1$ and ${}^{(N)}\!G_1$ now represent the rates of introducing a new color (birth) or reusing an old one (growth), respectively. Similarly, the number of structures $S_{1,n}$ of size n evolves as

$$S_{1,n}(t+1) = S_{1,n}(t) + {}^{(S)}\!B_1\,\delta_{n,1} + \frac{{}^{(S)}\!G_1}{({}^{(S)}\!B_1 + {}^{(S)}\!G_1)t}$$
$$\times [(n-1)S_{1,n-1}(t) - nS_{1,n}(t)]. \quad (7)$$

Since Eqs. (6) and (7) and (1) are similar, the normalized distributions $\{\tilde{N}_{1,s}(t)\}$ and $\{\tilde{S}_{1,s}(t)\}$, with $X_{1,s}(t) = t\,{}^{(X)}\!B\,\tilde{X}_{1,s}(t)$ (with $X = N$ or S), satisfy equations of the form (2) whose stationary solutions reproduce Eq. (4)

$$\tilde{X}_{1,s}^* = \frac{\prod_{m=1}^{s-1} \frac{{}^{(X)}\!G_1}{{}^{(X)}\!B_1 + {}^{(X)}\!G_1} m}{\prod_{m=1}^{s}\left(1 + \frac{{}^{(X)}\!G_1}{{}^{(X)}\!B_1 + {}^{(X)}\!G_1}m\right)}, \quad (8)$$
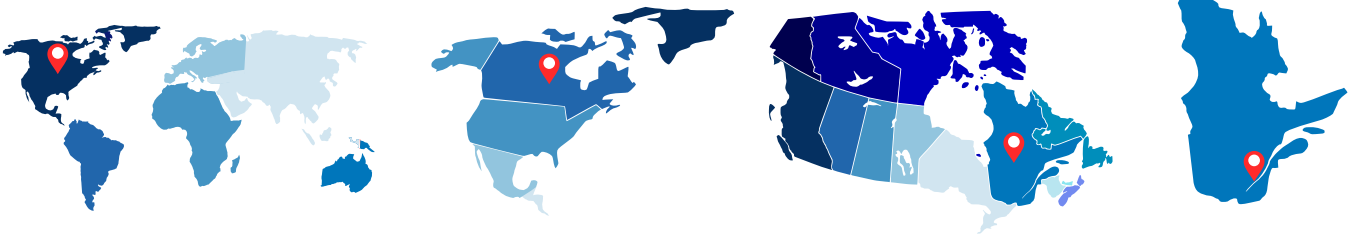
FIG. 1. (Color online) An example of hierarchical structure. Individuals involved in workshops taking place in given states or provinces forming countries within continents (four embedded structures).

from which we recover the scaling exponents directly:

$$\lim_{t,m\to\infty} \tilde{N}^*_{1,m}(t) \propto m^{-\gamma_{N,1}} \quad \text{with} \quad \gamma_{N,1} = 2 + \frac{^{(N)}B_1}{^{(N)}G_1}, \quad (9a)$$

$$\lim_{t,n\to\infty} \tilde{S}^*_{1,n}(t) \propto n^{-\gamma_{S,1}} \quad \text{with} \quad \gamma_{S,1} = 2 + \frac{^{(S)}B_1}{^{(S)}G_1}. \quad (9b)$$

In the context of social networks, this new process leads to a scale-independent community structure, where both the distribution of members per community and the distribution of communities per individual asymptotically follow a power-law organization. However, considering that this organization is found in distributions of friends [3], of members in social groups [6], and of city population [29], it is natural to ask the following: How would a preferential attachment occurring on *multiple* structural levels influence the created system? It is a popular idea that complexity frequently takes the form of hierarchy and that a hierarchical organization influences the property of the whole independently of the nature of its content [1]. With the recent successes of preferential attachment models, we hereafter propose a generalization for hierarchical systems.

IV. HIERARCHICAL PREFERENTIAL ATTACHMENT (HPA)

We now generalize the process of Sec. III by considering systems consisting of an arbitrary number d of embedded levels of organization. Hence, we can describe hierarchical preferential attachment (HPA) as a scheme of throwing colored balls in d embedded levels of structures, which can be pictured as Russian dolls but different.

A. Qualitative description

We will start with a tongue-in-cheek example of what a model of HPA could be. Its precise nature is of no concern, but serves its illustrative purpose well, for preferential attachment (classic, structural, hierarchical, or otherwise) is not meant to mimic the actual mechanisms or microscopic details at play in any system [27]. In the following, the proposed urn scheme should be seen as a potential abstract model for the statistical properties of the system.

Let us assume that we want to study the distribution of scientists who have attended a small workshop held yearly in different states around the world. The network can then be constructed by ignoring time and simply assigning scientists to the different editions: we assign individual scientists to

embedded structures (workshops held in geographical regions). Each structural level follows the preferential attachment principle based on the substructures they contain. Consider, for instance, the example of Fig. 1: we are assigning one scientist to one workshop and to do so we must progressively go down the hierarchy of $d = 4$ embedded structures. In this case, large-scale structures represent continents (level $k = 1$) containing countries (level $k = 2$) containing provinces or states (level $k = 3$) containing fine-grained structures representing workshops (level $k = 4$); this should be enough to describe the global system (planet Earth, level $k = 0$). Thus, the level k refers to different resolution of coarse graining, such that large-scale structures mean low resolution (small k) and fine-grained structures mean high resolution (large k).

We now choose a workshop. On Fig. 1, we associate the scientist to an existing continent $k = 1$, in this case North America, then within that continent we select a country $k = 2$, Canada, then a province or state $k = 3$, Québec, and finally a single workshop $k = 4$, a workshop in Québec City. At each level, the process follows the preferential attachment principle; e.g., the city was chosen proportionally to the number of workshops therein.

We can now determine the identity of the scientist. This is achieved by lowering the resolution progressively and probing all structural levels with the following question: Is the scientist a new participant? For instance, the scientist could be new to Canada (i.e., he has never attended a workshop in Canada), but not to North America, in which case his identity is borrowed from the United States or Mexico proportionally to his past activity in these two countries.

These embedded preferential attachment processes can be used to impose multiple constraints. Perhaps some countries host the workshop more often than others (preferential attachment at each structural level), and maybe some scientists seldom travel out of their own continent.

The HPA process can be mathematically described by using d different versions of Eq. (6) for the memberships of individuals (e.g., how many level k structures in which a given individual is found) and d more of Eq. (7) for the sizes of structures [e.g., how many level $(k + 1)$ structures in each level k structures]. The dynamics is then completely determined, assuming we set the birth $^{(N,S)}B_k$ and the growth $^{(N,S)}G_k$ rates properly at each level k.

B. Algorithmic description

We now describe a particular HPA model based on Simon's preferential attachment process [27] and explicitly show how it
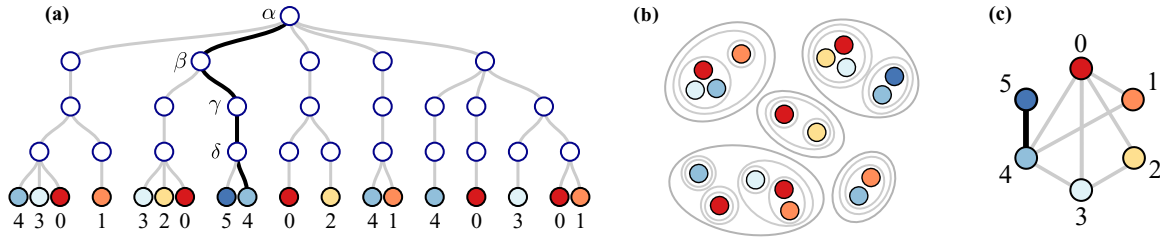
FIG. 2. (Color online) Schematization of hierarchical preferential attachment. HPA process frozen in time as a ball labeled 4 (the label representing a "color") is added to a $d = 3$ hierarchical structure. The process goes as follows. In this event, a structure at level 1 is chosen for growth (probability $1 - p_1$). Among the five structures of level 1 (total size 8), the structure β of size 2 is chosen for growth (probability 2/8). Then, into the selected structure β, a smaller structure labeled γ of size 1 is chosen for growth [probability $(1 - p_2) \times 1/2$] and finally a level 3 structure labeled δ of size 1 [probability $(1 - p_3) \times 1/1$]. Since $q_{d=3} = 1$ by construction, the "color" has to be new for δ (probability q_3). Then, the color is also new for γ because it is a size 1 structure and the logical constraint applies. The color is chosen to be new for β (probability q_1), but old for level 0 structure labeled α (probability $1 - q_0$). At this point, the accessible "colors" are those labeled 1 and 4. Balls 0, 2, 3, and 5 can not be chosen since the color should be new for structure β. Balls 1 and 4 have the same probability of being chosen as they both belong to three level 1 structures. The ball 4 is then chosen with probability 3/6 and placed in δ. (a) Hierarchical representation as an inverted tree. Navigating downwards corresponds to moving towards ever smaller structures until we reach the balls therein. (b) Representation as labeled balls in embedded levels of structures. (c) Possible network representation of the system. In this case, two nodes share an edge if they belong to a same level 3 structure. Adding ball number 4 to structure δ creates the link highlighted in bold.

can be followed algorithmically. The next subsection will then formalize the approach with an analytic description. Some visual representations of the model are given in Fig. 2 and a large hierarchical structure simulated with HPA is presented in Fig 3.

The model is represented either as a literal system of balls in embedded bins, or as the hierarchy it describes. For the rest of the paper, we will refer interchangeably to a level k structure as k-structure. Each event, or time step, is simply the act of throwing an additional ball in the system which depends on parameters p_k with $k \in [0, d + 1]$ and q_k with $k \in [0, d]$. It will soon become clear that some of these parameters are trivially assigned to avoid irregularities in the equations: $p_0 = 0$, $p_{d+1} = 1$, and $q_d = 1$.

The general process goes as follows. At every time step $\Delta t = 1$, an event takes place: a ball is thrown in d embedded structures. We first choose a set of structures. Starting at level $k = 1$, we have two options:

(1a) With probability p_k, we create a new structure. This forces the creation of one structure at all deeper levels $k' > k$ within that new structure. A larger structure cannot exist without containing at least one smaller structure.

(1b) With probability $1 - p_k$, an existing k-structure is chosen for growth. It is done preferentially to its size, i.e., the number of $(k + 1)$-structures that it contains. Repeat this step within the chosen structure (i.e., level $k + 1$), until level $k = d$ is reached or until a structure is created at level $k < d$.

Once a new structure has been created at level k, or once the level $k = d$ is reached by choosing existing structures, the color of the ball must be determined. By construction, the ball is new for the k-structure. We must, however, determine if it is new for the containing $(k - 1)$-structure. If not, we must sequentially examine lower resolution structures. This is determined by one of the two possible choices:

(2a) With probability q_k, the color is new for the k-structure. We then move to the level $k - 1$ and repeat the operation (2a or 2b). If level $k = 1$ is eventually reached, a new color is introduced in the system and the two-step process ends.

(2b) With probability $1 - q_k$, the color is chosen among all colors already occurring within this particular k-structure. This is done proportionally to the number of $(k + 1)$-structures, embedded in that selected k-structure, in which the colors appear. The two-step process concludes.
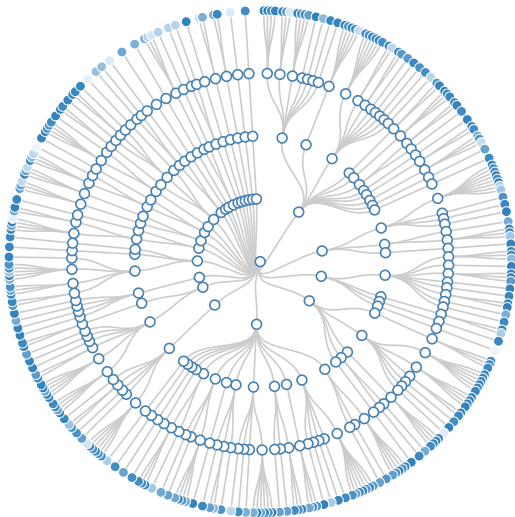

FIG. 3. (Color online) An example of HPA. HPA process for 250 steps on a structure of $d = 3$ levels. Each radius represents a level of organization. The nodes are found at the outermost circle and a unique color (shades of blue, more than 50) specifies their identity.

C. Mathematical description

The algorithmic rules just described can now be mapped onto an embedded system of preferential attachment equations. Table I gathers the different quantities involved.

TABLE I. Notation.

p_k	Probability to create a new k-structure
q_k	Probability to choose a new node for the selected k-structure
d	Number of structural levels of organization ($d = 1$ for SPA)
$^{(S)}B_k$	Rate of structural birth at level k
$^{(S)}G_k$	Rate of structural growth of a level k structure (it implies the creation of a new structure at level $k + 1$)
$^{(N)}B_k$	Rate of nodal birth at level k (equivalent to the rate of adding a new node to the system)
$^{(N)}G_k$	Rate of nodal membership growth at level k (rate at which a node acquires membership to an existing k-structure)
$S_{k,n}$	Number of k-structures of size n [i.e., containing n different $(k + 1)$-structures]
S_k	Number of structures at level k ($= \sum_n S_{k,n}$)
$N_{k,m}$	Number of nodes with m memberships at level k (i.e., appearing in m different k-structures)
P_k	Probability to choose a k-structure of size 1 under PA
$R_k(d)$	Probability that the construction process ends by choosing an existing node at level k, considering d levels of organization

1. Structural birth and growth

The structures of level k have a rate of birth $^{(S)}B_k$ and of growth $^{(S)}G_k$, for $k \leqslant d$, given by

$$^{(S)}B_k = \sum_{i=1}^{k} p_i \prod_{j=1}^{i-1} (1 - p_j) \qquad (10a)$$

and

$$^{(S)}G_k = p_{k+1} \prod_{i=1}^{k} (1 - p_i) \qquad (10b)$$

since birth events occur if structures are created at level k or at a lower resolution ($k' < k$). The growth of a k-structure requires to choose existing structures at levels $1 \leqslant i \leqslant k$ [probability $\prod_{i=1}^{k}(1 - p_i)$] and the creation of a structure at level $k + 1$

(probability p_{k+1}). In order to make every equation coherent, we adopt the product convention $\prod_{k=i}^{j} a_k = 1$ and the sum convention $\sum_{k=i}^{j} a_k = 0$ for $j < i$.

With these probabilities, the number $S_{k,n}(t)$ of k-structures with size n can be approximately followed using Eq. (7). However, while this is exact for the first structural level, it is an approximation for structures at a deeper level, $k > 1$. For example, the probability to choose a structure of size n at level 2 will depend on the size m of the level 1 structure in which it is nested. Mathematically, whereas level 1 evolves according to

$$S_{1,m}(t + 1) = S_{1,m}(t) + {}^{(S)}B_1 \delta_{m,1} + \frac{^{(S)}G_1}{[^{(S)}B_1 + {}^{(S)}G_1]t}$$
$$\times [(m - 1)S_{1,m-1}(t) - m S_{1,m}(t)], \qquad (11)$$

level 2 is governed by a somewhat more involved expression

$$S_{2,n,m}(t+1) = S_{2,n,m}(t) + \frac{m S_{1,m}(t)}{[^{(S)}B_1 + {}^{(S)}G_1]t} \left\{ {}^{(S)}G_2 \frac{(n-1)S_{2,n-1,m}(t) - n S_{2,n,m}(t)}{\sum_i i S_{2,i,m}(t)} - {}^{(S)}G_1 \frac{S_{2,n,m}(t)}{S_{1,m}(t)} \right\}$$
$$+ \frac{(m-1)S_{1,m-1}(t)}{[^{(S)}B_1 + {}^{(S)}G_1]t} \left\{ {}^{(S)}G_1 \frac{S_{2,n,m-1}(t)}{S_{1,m-1}(t)} + {}^{(S)}G_1 \delta_{n,1} \right\} + {}^{(S)}B_1 \delta_{n,1}\delta_{m,1}, \qquad (12)$$

where $S_{2,n,m}(t)$ is the number of level 2 structures of size n, nested in a level 1 structure of size m. This equation takes into account the choice of a 1-structure of size m or $(m - 1)$ and then the growth or the birth of a 2-structure.

To reduce Eq. (12) to a more manageable rate equation of the form (7), one must sum over all m to obtain an equation for $S_{2,n}(t) = \sum_m S_{2,n,m}(t)$. Under the approximation

$$\langle n \rangle_{S,2,m} \equiv \sum_n n \frac{S_{2,n,m}}{m S_{1,m}}$$
$$\simeq \sum_n n \frac{S_{2,n}}{\sum_j j S_{1,j}} = \sum_n n \frac{S_{2,n}}{\sum_j j S_{2,j}} \equiv \langle n \rangle_{S,2} \quad (13)$$

and using the relations $\sum_j j S_{k,j} = [\,{}^{(S)}B_k + {}^{(S)}G_k]t$ and $^{(S)}B_k + {}^{(S)}G_k = {}^{(S)}B_{k+1}$, the simplification follows immediately. This type of approximation can be applied successively

to all levels (e.g., $\langle n \rangle_{S,3,i,j} = \langle n \rangle_{S,3}$), yielding equations similar to Eq. (7). The stationary counterparts and scaling exponents [Eq. (9)] follow under the obvious replacements $(S_{1,n}, {}^{(S)}B_1, {}^{(S)}G_1) \rightarrow (S_{k,n}, {}^{(S)}B_k, {}^{(S)}G_k)$.

The resulting dynamical equations therefore describe a set of *uncorrelated* levels of structural organization with well-defined scaling exponents $\{\gamma_{S,k}\}$.

2. Nodal birth and membership growth

While the description of structure sizes is a straightforward problem, things get more involved for the number $N_{k,m}(t)$ of colors appearing in m structures of level k. An important logical constraint occurs for k-structures with size equal to one: if the color is new for the sole structure of level $k + 1$ therein (probability q_{k+1}), it must logically be new for the structure of level k (as seen in the example of Fig. 2). Thus, the probabilities $\{q_k\}$ must be corrected to account for this

logical constraint:

$$q'_k(t) = q_k + P_k(t)q_{k+1} = q_k + \frac{\tilde{S}_{k,1}(t)}{\sum_n n\tilde{S}_{k,n}(t)}q_{k+1}, \quad (14)$$

where $P_k(t)$ is the probability that the k-structure of interest had a size equal to 1. In other words, if the color is new at level k, it can either be because of the initial probability q_k, or because it was forced to be new by the aforementioned logical constraint. Equation (14) is only valid for $0 < k < d$ since there is no correction at $k = 0$ and $k = d$; $q_0 = q'_0$ and $q_d = q'_d = 1$. The probabilities $P_k(t)$ can be obtained from the rate equation for sizes of k-structures [Eq. (7)], as well as from their steady state values in the limit $t \to \infty$. Together this yields

$$\lim_{t \to \infty} q'_k(t) = q_k + \frac{\tilde{S}^*_{k,1}}{\sum_n n\tilde{S}^*_{k,n}}q_{k+1} \quad (15)$$

$$= q_k + \frac{{}^{(S)}B_k + {}^{(S)}G_k}{{}^{(S)}B_k + 2{}^{(S)}G_k}\frac{q_{k+1}}{\langle n^*\rangle_{S,k}}, \quad (16)$$

where the average size $\langle n^*\rangle_{S,k} = \sum_n n\tilde{S}^*_{k,n}$ corresponds intuitively to the ratio of the total rate to the birth rate:

$$\langle n^*\rangle_{S,k} = \frac{[{}^{(S)}B_k + {}^{(S)}G_k]}{{}^{(S)}B_k}. \quad (17)$$

This result can also be obtained analytically and its demonstration is relegated to Appendix A. Inserting this last expression in Eq. (16) finally leads to

$$\lim_{t \to \infty} q'_k(t) = q_k + \frac{q_{k+1}}{1 + 2{}^{(S)}G_k/{}^{(S)}B_k}. \quad (18)$$

It is then a matter of evaluating the birth ${}^{(N)}B_k$ and growth rates ${}^{(N)}G_k$ (see Table I). To obtain the growth rates, let us consider the probability $R_k(d)$ that the chosen color is an existing one selected according to level k. These probabilities are easily calculated for the three deepest levels. By definition, $R_d(d) = 0$, and

$$R_{d-1}(d) = (1 - q'_{d-1})\prod_{i=0}^{d-1}(1 - p_i) \quad (19)$$

and

$$R_{d-2}(d) = (1 - q'_{d-2})p_{d-1}\prod_{i=0}^{d-2}(1 - p_i)$$
$$+ (1 - q'_{d-2})q'_{d-1}\prod_{i=0}^{d-1}(1 - p_i). \quad (20)$$

These probabilities yield a recursive expression for $k \leqslant d - 2$:

$$R_k(d) = (1 - q'_k)p_{k+1}\prod_{i=0}^{k}(1 - p_i)$$
$$+ (1 - q'_k)q'_{k+1}\frac{R_{k+1}(d)}{1 - q'_{k+1}}, \quad (21)$$

starting from $R_{d-1}(d)$ given above. The terms ${}^{(N)}G_k$ can then be written as the sum of the probabilities of choosing an existing node according to level k or higher levels ($k' < k$):

$${}^{(N)}G_k = \sum_{i=0}^{k-1}R_i(d). \quad (22)$$

To obtain the birth rate, we calculate the probability of introducing a new individual at each time step. Since an individual has at least one membership at each level, the birth rate at each level is, namely, the rate of introducing a new color to the system. In consequence, ${}^{(N)}B_i = {}^{(N)}B_j$ for all i, j. Using the normalization ${}^{(N)}B_d(t) + {}^{(N)}G_d(t) = 1$ since we always add at least one ball to a d-structure, we infer

$${}^{(N)}B_k = 1 - {}^{(N)}G_d = \frac{q_0}{1 - q_0}R_0(d). \quad (23)$$

At this point, it is perhaps helpful to collect some of the explicit expressions of the birth and growth functions for a few hierarchical depths, say $d = 1$, 2, and 3. Table II illustrates the construction scheme of these functions. A few observations are worth noticing. First, for internal consistency and as already used previously, one observes that

$${}^{(S)}B_k + {}^{(S)}G_k = {}^{(S)}B_{k+1}, \quad 0 \leqslant k \leqslant d - 1 \quad (24)$$

as clearly seen from the definitions (10). Second, at level d, the birth and growth functions satisfy a normalization condition for both structures and nodes ($X = S$ or N)

$${}^{(X)}B_d + {}^{(X)}G_d = 1. \quad (25)$$

Finally, ${}^{(N)}B_k$ and ${}^{(N)}G_k$ depend explicitly on the probabilities $\{q'_k\}$ which themselves depend on the structural functions $\{{}^{(S)}B_k\}$ and $\{{}^{(S)}G_k\}$. In other words, the logical constraints, captured by the $\{q'_k\}$, induce correlations between the evolution of the hierarchical structure and the distribution of elements within this structure.

To validate the use of these birth and growth rates in Eqs. (6) and (7), we examine the pyramid of production entities in the film industry. Based on the Internet Movie Database (IMDb), we study a system with $d = 3$ structural levels where 363 571 producers (colored balls) are assigned to films (426 913 films, $k = 3$) associated with one principal production company (121 958 companies, $k = 2$), in a given country (198 countries, $k = 1$). The results of this case study are presented in Fig. 4.

While the mean-field description of the distributions $\{\tilde{N}_{k,m}(t)\}$ suffers from neglecting possible correlations from one resolution level i to level $i + 1$, the numerical simulations correctly reproduce the system and its finite size effects (distribution cutoff). The approximation of uncorrelated levels is also the source of the error observed in the mean-field description of the distributions $\{\tilde{S}_{k,n}(t)\}$ for $k > 1$ and becomes increasingly inadequate for larger k. The progression of error is essentially caused by the fact that a strict description of the third level, for instance, should not only be given in terms of $S_{3,n,m}(t)$ [see Eq. (12)], but of $S_{3,n,m,l}(t)$ describing the number of level 3 structures of size n nested in level 2 structures of size m themselves nested in level 1 structures of size l.

A few words are worth mentioning to describe our practical selection of the parameters $\{p_i\}_{i=1,d}$ and $\{q_j\}_{j=0,d-1}$. The heuristic method that we used amounts to a cascade from low to high resolution and back and mimics the generative

TABLE II. Birth and growth functions.

$d = 1$ (SPA)	$d = 2$	$d = 3$
$^{(S)}B_0 = 0$	$^{(S)}B_0 = 0$	$^{(S)}B_0 = 0$
$^{(S)}B_1 = p_1$	$^{(S)}B_1 = p_1$	$^{(S)}B_1 = p_1$
	$^{(S)}B_2 = p_1 + p_2(1 - p_1)$	$^{(S)}B_2 = p_1 + p_2(1 - p_1)$
		$^{(S)}B_3 = p_1 + p_2(1 - p_1) + p_3(1 - p_1)(1 - p_2)$
$^{(S)}G_0 = p_1$	$^{(S)}G_0 = p_1$	$^{(S)}G_0 = p_1$
$^{(S)}G_1 = (1 - p_1)$	$^{(S)}G_1 = p_2(1 - p_1)$	$^{(S)}G_1 = p_2(1 - p_1)$
	$^{(S)}G_2 = (1 - p_1)(1 - p_2)$	$^{(S)}G_2 = p_3(1 - p_1)(1 - p_2)$
		$^{(S)}G_3 = (1 - p_1)(1 - p_2)(1 - p_3)$
$R_0(1) = (1 - q_0)$	$R_0(2) = (1 - q_0)[p_1 + q_1'(1 - p_1)]$	$R_0(3) = (1 - q_0)\{p_1 + q_1'(1 - p_1)[p_2 + q_2'(1 - p_2)]\}$
	$R_1(2) = (1 - q_1')(1 - p_1)$	$R_1(3) = (1 - q_1')(1 - p_1)[p_2 + q_2'(1 - p_2)]$
		$R_2(3) = (1 - q_2')(1 - p_1)(1 - p_2)$
$^{(N)}B_1 = q_0$	$^{(N)}B_k = q_0[p_1 + q_1'(1 - p_1)]$	$^{(N)}B_k = q_0\{p_1 + q_1'(1 - p_1)[p_2 + q_2'(1 - p_2)]\}$
$^{(N)}G_1 = R_0(1)$	$^{(N)}G_k = \sum_{i=0}^{k-1} R_i(2)$ for $k = \{1,2\}$	$^{(N)}G_k = \sum_{i=0}^{k-1} R_i(3)$ for $k = \{1,2,3\}$

procedure: first moving towards finer structures to obtain sequentially p_1 to p_d and then choosing the memberships by subsequent coarse graining to select q_{d-1} to q_0. With $d = 3$, we have six parameters to set. This leaves us with a significant amount freedom bounded, however, by three empirical size distributions and three empirical membership distributions to limit our choices. The size distribution at level $k = 1$ depends on p_1 and p_2, so one can put a value on these parameters by following the scale exponent of the empirical distribution. The size distributions of the following levels k' then depend on $p_{k \leqslant k'}$, such that every $p_{k'}$ can be chosen to reproduce the scale exponent of the size distribution at that level given the previous values of $p_{k < k'}$. Memberships follow a somewhat similar but reverse logic because of the constraints mentioned at the beginning of this subsection. The membership distribution at $k = 3$ for instance will depend on all $\{p_i\}$ and on q_2, such that one should then start reproducing membership scale exponents at level $k = 3$ first.

The system of Fig. 4 is representative of a case where the number d of structural levels is known *a priori* and our experimentations show little variations in the final choice of parameters although an exhaustive optimization was not attempted. However, were the depth of the hierarchy unknown, any superfluous levels would lead to overfitting and nonuniqueness of the model parameters. To adequately infer the number of structural levels is still an open problem and its solution is currently being pursued.

V. PROJECTION ON NETWORKS

Despite the advent of large data sets, few hierarchical systems are categorized and referenced as such. Consequently, research tends to focus on a single level of activity. For instance, the IMDb is often studied as a network of coactors [3,4] or, as in Fig. 4, a network of coproductions where
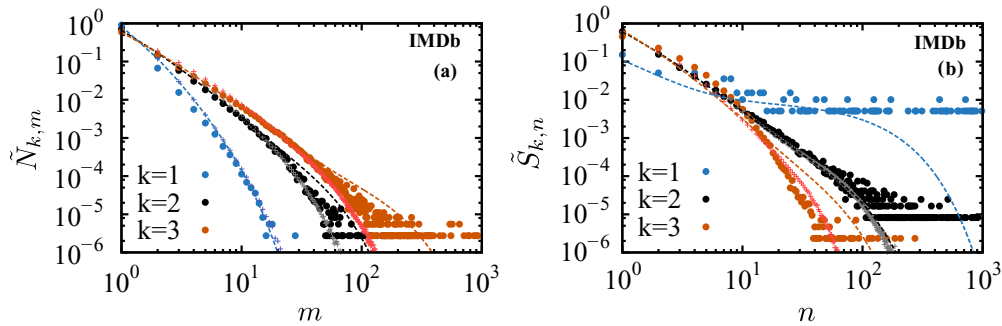

FIG. 4. (Color online) Hierarchical structure of film production. Events involving producers are distributed among $d = 3$ structural levels: films at $k = 3$ [sequence of upper dots in (a) and of lower dots in (b)], in production companies at $k = 2$ [sequence of middle dots in (a) and (b)] in countries at $k = 1$ [sequence of lower dots in (a) and of upper dots in (b)]. (a) Distribution of the number of films, companies, and countries a given producer is involved with. (b) Distribution of the number of producers, films, and companies involved within a given film, company, and country. The empirical data are shown with dots. Lines are obtained with Eqs. (6) and (7) using the corresponding birth and growth probabilities; crosses indicate direct Monte Carlo simulations. Both calculations are iterated for 10^6 time steps using $(p_1, p_2, p_3) = (0.0005, 0.185, 0.385)$, $(q_0, q_1, q_2) = (0.80, 0.60, 0.50)$. Simulated results of $S_{1,n}$ are not shown to avoid cluttering the figure (note that the plateau observed in the empirical data is due to finite size). The correspondence between the observed scale exponents and our mathematical results implies that the model is not overparametrized: $2d$ parameters for $2d$ scale exponents. The chosen parameters were hand selected to roughly reproduce the qualitative behavior of each distribution.
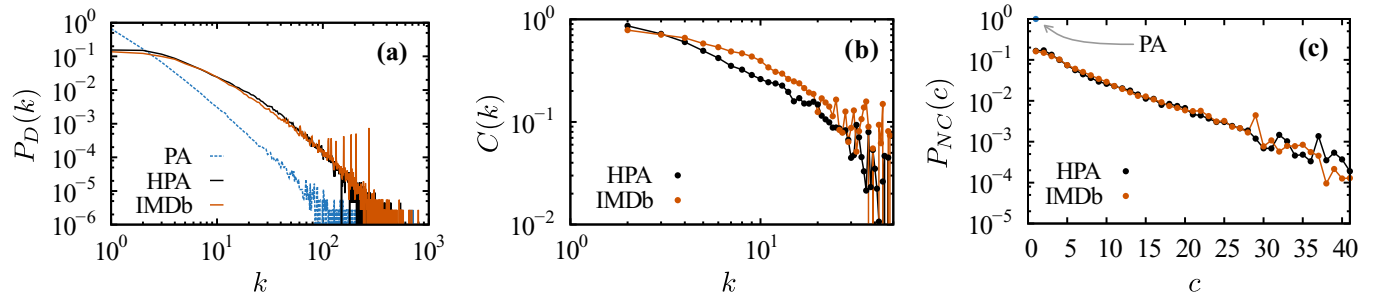
FIG. 5. (Color online) Scale independence and clustering of projected hierarchical systems. (a) Degree distribution $P_D(k)$ observed in networks created by projecting the systems of Fig. 4 on webs of coproductions (the actual data and one simulated system with the same parameters as those of Fig. 4). A network obtained through the classic preferential attachment model [3] (PA) is given for comparison. (b) Average clustering coefficient for nodes of degree k. PA leads to a vanishing clustering $C(k) = 0$ for all degree k in large networks. (c) Distribution of node centrality $P_{NC}(c)$ measured with their coreness c under k-core decomposition of the networks. PA leads to a unique shell of coreness $c = 1$ because of the treelike structure of the network.

producers are connected if they have produced a film together (if they are found within a common level d structure). Effectively, this implies that the system is reduced to a projection of all structural levels onto the chosen activity. While the involvement of actors and producers in films is well captured, their involvement in different companies and countries is somehow encoded, and more than often lost, in the resulting network.

A. Degree, clustering, and centrality

Following the projection procedure schematically illustrated on Fig. 2 (right), Fig. 5 presents some basic properties obtained by projecting the hierarchical system of film production onto a network of coproducers. We first investigate the degree distribution $P_D(k)$ (coproducing link per producer) and the clustering function $C(k)$ (probability that two links of a degree k producer are part of a triangle) of a network projection of a HPA system based on the parameters used in Fig. 4. The nontrivial clustering [4,7] and the power-law tail of the degree distribution [3], properties ubiquitous in real networks, are reproduced in our framework as emergent features of the HPA model. Essentially, by only fitting the hierarchical structure of the IMDb coproduction network, we obtain a good approximation of the complex properties of the network projection without having to directly account for them in the model. For an example of the calculation of the scaling exponents across multiple scales, we refer the reader to Appendix B.

Moreover, Fig. 5 also presents the result of a centrality analysis known as core decomposition. This analysis relies on the concept of c cores, i.e., the maximal subset where all nodes share c links amongst one another. A node is assigned coreness c if it belongs to the c core but not to the $(c + 1)$ core. This procedure effectively defines a periphery (low c) and core (high c) to the network and was recently shown to reflect structural organization beyond simple local correlations [30]. The HPA centrality distribution is seen to agree quite well with the data. This increases our confidence that the model effectively reproduces the structure of the real hierarchical system beyond the statistical properties previously considered in Fig. 4.

B. Fractality

Aside from scale-independent degree distribution and nontrivial clustering function, the fractality of complex networks is often a sign of hierarchical organization [9,10]. One can unravel the fractal nature of a network using a box counting method: groups of nodes within a distance (number of links) r of each other are assigned to the same box. The fractal dimension d_b of a network manifests itself as a scaling relation between the number N_b of boxes needed to cover all nodes and the size r of the boxes ($N_b \propto r^{-d_b}$). The self-similarity of network structure was previously assumed to stem from a repulsion or disassortativity between the most connected nodes [10]. However, Fig. 6 demonstrates that fractality can also emerge from a scale-independent hierarchical structure, without further assumptions. Interestingly, Fig. 6 (left) also illustrates how, even if fractality might imply hierarchy, the opposite is not necessarily true.

HPA can produce both fractal and nonfractal networks. It remains to be determined whether or not this box counting method is truly equivalent to an actual measure of *dimensionality*. However, it can, at the very least, be interpreted as an observation of how easily a network can be covered. Of course, since the definition of network fractality is somewhat ambiguous, so is the distinction between sets of HPA parameters leading to fractality or not. Nevertheless, a useful empirical rule can be established.

Most models of stochastic network growth produce networks with very low mean shortest paths, low clustering, and no long-range correlations. Consequently, the number of boxes needed to cover the whole network falls very rapidly. In HPA, we can control the manner in which boxes cover the network since the distance between higher structural levels is directly influenced by the memberships at this level. Hence, HPA can generate networks that are more robust to box covering [i.e., such that $N_b(r)$ falls slower with respect to r] if higher structural levels feature less nodes that act as bridges between structures and levels. For example, in Fig. 7, only nodes 0, 1, and 2 can be used by boxes to move from one level to the other (from workshops to countries, here illustrated as an inverted tree).

More precisely, let us consider the two different networks of Fig. 6 (left) built using the parameters given in the caption.
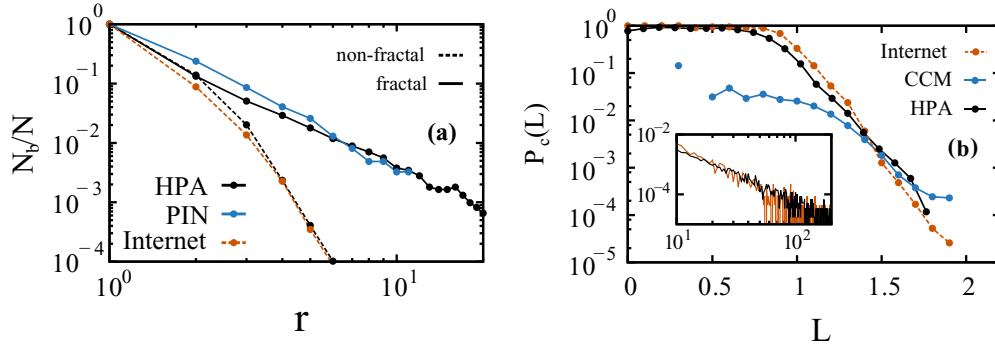
FIG. 6. (Color online) Fractality and navigability of projected hierarchical systems. (a) Box counting results on a well-known fractal network [protein interaction network (PIN) of *Homo sapiens*] and a nonfractal network (the Internet at the level of autonomous systems) [9]. HPA can approximately model how both of these networks span and cover their respective space, with $(p_1, p_2, p_3) = (0.01, 0.02, 0.30)$, $(q_0, q_1, q_2) = (0.95, 0.80, 0.30)$ (fractal) or $(p_1, p_2, p_3) = (0.005, 0.195, 0.395)$, $(q_0, q_1, q_2) = (0.60, 0.40, 0.30)$ (nonfractal). (b) Probability of connection $P_c(L)$ between nodes at a distance L after an inferred projection of the networks onto a hyperbolic space. (The distance is given as a fraction of the hyperbolic disk radius. See Boguñá *et al.* [12] for details on the method.) Both the Internet and its HPA model are the same as presented on the left and share a similar scaling exponent for their degree distribution [see inset: degree distribution $D(k)$ versus k]. The CCM corresponds to a rewired network preserving degree distribution and degree-degree correlations [31], but lacks the more complex structural correlations.

Roughly speaking, in the nonfractal network, 2-structures contain on average around three 3-structures whereas nodes belong to over four 3-structures. Therefore, a single node typically grants access to all of the 3-structures contained within its 2-structure, such that a box covering at least *part* of a 2-structure typically covers *most* of it. The network is thus easily covered as higher levels are not any harder to navigate.

In contrast, 2-structures of the fractal network contain on average ten 3-structures while an average node may still be found within around three 3-structures. An average 2-structure may thus have nodes at a distance greater than three steps. The network is consequently harder to cover and can be expected to be much more robust to box covering. As a general rule, we have found that to display *measurable* network self-similarity, the average size of a structure (at level k) has to be at least greater than the memberships of a node at the deeper level (at level $k + 1$).

C. Navigability

The box decomposition method tells us something about how networks cover the space in which they are embedded, and consequently at what speed a random walker might encounter new nodes in this network. However, it tells us nothing
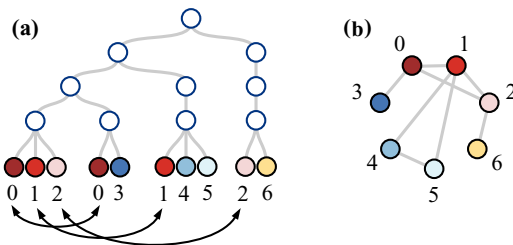

FIG. 7. (Color online) Example of bridges. (a) Inverted tree representation of a hierarchical structure and (b) the corresponding network projection which shows how nodes labeled 0, 1, and 2 act as bridges between structures.

about the geometrical space that supports the network, or how a walker could find one specific node. In that respect, the navigability of complex networks has recently been a subject of interest for two reasons. First, the development of a mapping of networks to a geometrical space allows to predict the probability of links as a function of geometrical distance between nodes, which in turn enables an efficient navigation through the network [11,12]. Second, network growth based on preferential attachment fails to capture this geometrical property [13]. In a recent paper [13], this metric was consequently considered as evidence of an opposition between two organizational forces: popularity (preferential attachment) and similarity (assortativity). Our last case study, shown in Fig. 6 (right), indicates that geometrical constraints, or network navigability, can emerge under a strict preferential attachment, which implies a growth driven by popularity only, but one occurring on multiple structural levels. The different hierarchical levels can *a posteriori* be interpreted as indicators of similarity, but are conceptually much more general.

We also compare in Fig. 6 (right) the results obtained on the actual network and on its HPA model with those obtained on a rewired network that preserves the degree distribution and degree-degree correlations (correlated configuration model, CCM) [31]. The fact that CCM does not preserve the navigability of the Internet structure indicates that it emerges mostly from clustering and long-range correlations. As the HPA network does reproduce the navigability of the Internet, these long-range correlations could very well be consequences of the hierarchical structure. It would be instructive to investigate whether the inferred structure corresponds to the actual hierarchy of the Internet (probably of geographical nature: continents, countries, regions).

VI. CONCLUSION

We have presented a proof of concept for a hierarchical preferential attachment (HPA) model in an attempt to reproduce the hierarchical nature of complex systems. We have

illustrated how complex networks could be better analyzed by first modeling their hierarchical structure, and then projecting this structure onto a network. Not only does this procedure yield the nontrivial clustering of networks and their degree (centrality) distributions at multiple levels, but it also gives access to the hidden geometrical metrics of these networks and the way they occupy space.

The fact that so many key features of the network structure are modeled using two minimal assumptions, hierarchy and preferential attachment, indicates that HPA provides more than theoretical insights; it leads support to the underlying assumptions. HPA could therefore be used to infer the possible hierarchical structure of networks when this information is not directly available.

Finally, while HPA is essentially a simple stochastic growth process, it nevertheless exemplifies eloquently how complex structural features of real networks, e.g., scale independence, clustering, self-similarity, fractality, and navigability, can emerge through the hierarchical embedding of scale independent levels. Perhaps this is the most important message: to study the structure of complex networks, one should avoid focusing on unique level of activity (e.g., links), but instead investigate the hidden hierarchical organizations from which the networks emerge.

APPENDIX A: AVERAGE STRUCTURAL SIZE

We wish to demonstrate the relation

$$\langle n^* \rangle_{S,k} = \sum_{n=1}^{\infty} n \tilde{S}_{k,n}^* = \frac{{}^{(S)}\mathrm{B}_k + {}^{(S)}\mathrm{G}_k}{{}^{(S)}\mathrm{B}_k} \qquad (A1)$$

for the average size of k-structures at equilibrium. We will support our intuition that the mean value $\langle n^* \rangle_{S,k}$ should simply be the ratio between the number of shares $({}^{(S)}\mathrm{B}_k + {}^{(S)}\mathrm{G}_k)t$ and the number of structures ${}^{(S)}\mathrm{B}_k t$. Inserting (8) for $\tilde{S}_{k,n}$, simplifying and rearranging, one finds

$$\sum_{n=1}^{\infty} n \tilde{S}_{k,n}^* = b \sum_{n=1}^{\infty} \prod_{m=1}^{n} \frac{m}{b+m}, \qquad (A2)$$

where $b = ({}^{(S)}\mathrm{B}_k + {}^{(S)}\mathrm{G}_k)/ {}^{(S)}\mathrm{G}_k$. The numerator and denominator are easily identified. The numerator $\prod_{m=1}^{n} m = n!$ is a factorial while the numerator $\prod_{m=1}^{s}(b+m) = (b+1)_s$ is a Pochhammer symbol, i.e., $(x)_n = x(x+1)\ldots(x+n-1)$. This reduces our expression to

$$\langle n^* \rangle_{S,k} = b \sum_{m=1}^{\infty} \frac{s!}{(b+1)_s}. \qquad (A3)$$

One recognizes the sum as a hypergeometric series (minus the leading term), namely,

$$b \sum_{m=1}^{\infty} \frac{s!}{(b+1)_s} = b \sum_{m=1}^{\infty} \frac{(1)_s(1)_s}{(b+1)_s} \frac{1}{s!}$$
$$= b\, [\, {}_2F_1(1,1;b+1;1) - 1]. \qquad (A4)$$

Since the argument of the ${}_2F_1$ is equal to 1, a useful transformation [32] asserts that

$${}_2F_1(\alpha,\beta;\gamma;1) = {}_2F_1(-\alpha, -\beta;\gamma - \alpha - \beta;1) \qquad (A5)$$

as long as Re $(\gamma) >$ Re$(\alpha + \beta)$. This property applies to our case where $\alpha = \beta = 1$ and $\gamma = b+1$ leading to a finite terminating series

$${}_2F_1(-1,-1;b-1;1) = 1 + \frac{(-1)(-1)}{(b-1)} \qquad (A6)$$

which, once inserted in (A4), leaves us with the final result

$$\langle n^* \rangle_{S,k} = b\left[\frac{1}{(b-1)} \right] = \frac{{}^{(S)}\mathrm{B}_k + {}^{(S)}\mathrm{G}_k}{{}^{(S)}\mathrm{B}_k}. \qquad (A7)$$

APPENDIX B: MULTIPLE SCALE INDEPENDENCE

By ignoring the interlevel correlations for the structural growth, we have obtained in Sec. IV C a set of coupled equations (6) and (7) for all levels k that enable us to follow approximately the time evolution of the size distributions $\{\tilde{S}_{k,n}\}$ and of the node membership distributions $\{\tilde{N}_{k,m}\}$. We were then able to derive their scale exponents $\{\gamma_{S,k}, \gamma_{N,k}\}$ in the limit $t \to \infty$ [Eq. (9)].

When investigating the projected properties of a hierarchical system, for instance the degree distribution of the resulting network, we can combine the membership and size distributions of the lowest level d (where links are created) to deduce the resulting scaling exponent. As done in [16], the idea is to define the following *probability generating functions* (pgf):

$$\mathcal{S}(x,t) = \sum_n \tilde{S}_{d,n}(t)x^n \quad \text{and} \quad \mathcal{N}(x,t) = \sum_m \tilde{N}_{d,m}(t)x^m.$$
$$(B1)$$

Since a community of size n implies $n-1$ links for each node, the generating function of the distribution of the number of links $\mathcal{L}(x,t)$ in a d-structure for a randomly chosen node can be generated by

$$\mathcal{L}(x,t) = \frac{\frac{d}{dx}\mathcal{S}(x,t)}{\frac{d}{dx}\mathcal{S}(x,t)|_{x=1}} = \frac{\sum_n n S_{d,n}(t)x^{n-1}}{\sum_n n S_{d,n}(t)}. \qquad (B2)$$

The degree distribution is then generated by $\mathcal{D}(x,t)$, a pgf combining the distribution of memberships m and of links obtained from each of these memberships:

$$\mathcal{D}(x,t) = \mathcal{N}(\mathcal{L}(x,t),t), \qquad (B3)$$

which will simply scale as the slowest decreasing function between $\mathcal{N}(x,t)$ and $\mathcal{L}(x,t)$. The scale exponent of the degree

distribution is thus given by

$$\min[\gamma_{N,d}, \gamma_{S,d} - 1]. \qquad (B4)$$

The same method could of course be used to determine the scaling of other projections (e.g., network of companies sharing or having shared at least one producer).

[1] H. A. Simon, The architecture of complexity, Proc. Am. Philos. Soc. **106**, 467 (1962).

[2] D. G. Champernowne, A model of income distribution, Econ. J. **63**, 318 (1953).

[3] A. Barabási and R. Albert, Emergence of scaling in random networks, Science **286**, 509 (1999).

[4] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).

[5] M. Girvan and M. E. J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).

[6] L. Hébert-Dufresne, A. Allard, V. Marceau, P.-A. Noël, and L. J. Dubé, Structural Preferential Attachement: Network Organization Beyond the Link, Phys. Rev. Lett. **107**, 158702 (2011).

[7] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Hierarchical organization of modularity in metabolic networks, Science **297**, 1551 (2002).

[8] A. Clauset, C. Moore, and M. E. J. Newman, Hierarchical structure and the prediction of missing links in networks, Nature (London) **453**, 98 (2008).

[9] C. Song, S. Havlin, and H. A. Makse, Self-similarity of complex networks, Nature (London) **433**, 392 (2005).

[10] C. Song, S. Havlin, and H. A. Makse, Origins of fractality in the growth of complex networks, Nat. Phys. **2**, 275 (2006).

[11] M. Boguná, D. Krioukov, and K. C. Claffy, Navigability of complex networks, Nat. Phys. **5**, 74 (2009).

[12] M. Boguñá, F. Papadopoulos, and D. Krioukov, Sustaining the internet with hyperbolic mapping, Nat. Commun. **1**, 62 (2010).

[13] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, Popularity versus similarity in growing networks, Nature (London) **489**, 537 (2012).

[14] E. Ravasz and A.-L. Barabási, Hierarchical organization in complex networks, Phys. Rev. E **67**, 026112 (2003).

[15] G. Palla, L. Lovász, and T. Vicsek, Multifractal network generator, Proc. Natl. Acad. Sci. USA **107**, 7640 (2010).

[16] L. Hébert-Dufresne, A. Allard, V. Marceau, P.-A. Noël, and L. J. Dubé, Structural preferential attachment: Stochastic process for the growth of scale-free, modular, and self-similar systems, Phys. Rev. E **85**, 026108 (2012).

[17] P. L. Krapivsky, S. Redner, and F. Leyvraz, Connectivity of Growing Random Networks, Phys. Rev. Lett. **85**, 4629 (2000).

[18] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Structure of Growing Networks with Preferential Linking, Phys. Rev. Lett. **85**, 4633 (2000).

[19] S. N. Dorogovtsev and J. F. F. Mendes, Scaling properties of scale-free evolving networks: Continuous approach, Phys. Rev. E **63**, 056125 (2001).

[20] P. L. Krapivsky and S. Redner, Organization of growing random networks, Phys. Rev. E **63**, 066123 (2001).

[21] E. Volz, Random networks with tunable degree distribution and clustering, Phys. Rev. E **70**, 056115 (2004).

[22] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Triadic closure as a basic generating mechanism of communities in complex networks, Phys. Rev. E **90**, 042806 (2014).

[23] P. L. Krapivsky and S. Redner, Network growth by copying, Phys. Rev. E **71**, 036118 (2005).

[24] A. Gabel, P. L. Krapivsky, and S. Redner, Highly dispersed networks by enhanced redirection, Phys. Rev. E **88**, 050802(R) (2013).

[25] G. U. Yule, A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S, Philos. Trans. R. Soc. London, Ser. B **213**, 21 (1925).

[26] R. Gibrat, Ph.D. thesis, Université de Lyon, 1931.

[27] H. A. Simon, *Models of Man* (Wiley, New York, 1961).

[28] D. de Solla Price, A general theory of bibliometric and other cumulative advantage processes, J. Am. Soc. Inf. Science **27**, 292 (1976).

[29] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, 1949).

[30] L. Hébert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé, Percolation on random networks with arbitrary k-core structure, Phys. Rev. E **88**, 062820 (2013).

[31] M. E. J. Newman, Assortative Mixing in Networks, Phys. Rev. Lett. **89**, 208701 (2002).

[32] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 7th ed. (Academic, Amsterdam, 2007).