Interfaces and the edge percolation map of random directed networks

M. Ángeles Serrano and Paolo De Los Rios

Institute of Theoretical Physics, LBS, SB, EPFL, 1015 Lausanne, Switzerland

The traditional node percolation map of directed networks is reanalyzed in terms of edges. In the percolated phase, edges can mainly organize into five distinct giant connected components, interfaces bridging the communication of nodes in the strongly connected component and those in the in and out components. Formal equations for the relative sizes in the number of edges of these giant structures are derived for arbitrary joint degree distributions in the presence of local and two-point correlations. The uncorrelated null model is fully solved analytically and compared against simulations, finding an excellent agreement. Interfaces, and their particular conformations giving place from "hairy ball" percolation landscapes to bottleneck straits, could bring new light to the discussion of how a structure is interwoven with functionality in flow networks.

　　　　　　　　　　　　　　　PACS number(s): 89.75.Hc, 64.60.Ak

I. INTRODUCTION

The theory of percolation applied to random networks [1] has proven to be one of the most notorious advances in complex network science [2–4]. Its importance goes beyond the production in the short term of theoretical results, which are general and relevant to systems in many different fields. The implications are far reaching. On one hand, a number of different problems have a direct interpretation in terms of percolation or can be mapped to it, such as the study of resilience or vulnerability in front of random failures [5] or susceptible-infected-recovered (SIR) epidemic spreading models [6]. On the other hand, the emergent percolation landscape of self-organized systems can strongly affect properties such as fluency or navigability. Hence, its conformation should ensure efficient communication at the global level for the whole to organize and develop functionality.

In the case of undirected networks, where elements are linked by channels operating in both directions, the basic percolation discussion was centered around the appearance of a macroscopic portion of connected nodes. The critical point for the appearance of this giant component and its relative size in the number of nodes and edges was determined [5,7–10], also in the presence of specific structural attributes [11–16,19]. In its turn, the standard picture in directed graphs [9,14,17,18] establishes that this giant connected component may become much more complex and internally organized in three main giant structures, the in component, the out component, and the strongly connected component, as well as other secondary aggregates such as tubes or tendrils. This conformation, sometimes represented as a bow-tie diagram [20], denotes a potential global flow—of matter, energy, information, etc.—organized around a core which usually processes input into output.

In this work, we will see that the percolation landscape, the aggregate of macroscopic connectivity structures in the percolated phase above the critical point, is further shaped when edges, the zero-level primary building blocks of networks along with nodes, are taken as starring elements. Five distinct components are found to be relevant in the edge percolation map of directed networks, the strongly connected component bridged by two newly identified interfaces to the peripheral in and out components. In Sec. II, we give their definitions and present analytical computations for their relative sizes based on the usual locally treelike assumption and the generating function formalism in the presence of local and two-point correlations. In Sec. III, the formal equations for the most general situation will be reformulated for uncorrelated networks as null models. The corresponding analytical results will be compared to simulations for networks with exponential in- and out-degree distributions and with scale-free in-degree and exponential out-degree distributions. A discussion of the implications coming out of this description will be provided in Sec. IV, where the concept of interface will be further examined along indications of the potential relevance of its particular conformation, that could produce from "hairy ball" percolation landscapes to bottleneck straits. We end by summarizing and giving some final remarks in Sec. V.

II. EDGE COMPONENTS IN DIRECTED NETWORKS

In the traditional node percolation map of directed networks the core structure is the giant strongly connected component (SCC), where all vertices within can reach each other by a directed path. When present, it serves as a connector of the giant in component (IN), composed by all vertices that can reach the SCC but cannot be reached from it, to the giant out component (OUT), made of all vertices that are reachable from the SCC but cannot reach it.

From the point of view of edges, the IN and the OUT unfold into two structures each, the edge in component (ICE) and the in interface (ITF), and the edge out component (OCE) and the out interface (OTF), respectively (see Fig. 1). The number of relevant structures is increased to five as a consequence of the fact that, whereas nodes are point objects belonging at most to one of the three node components, edges can be considered as extended objects in the sense that they could belong simultaneously to two different node components, having for instance one end in the IN or OUT and the other in the SCC. This fact points to the necessity of defining new classes for edges. We will not take into account aggregates such as tendrils or tubes, so edges will be classified into five different categories depending on the affiliation
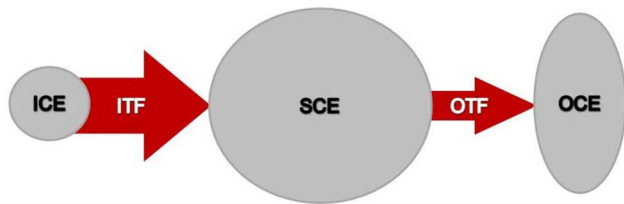
FIG. 1. (Color online) Schematic representation of the main giant components in the edge percolation map. As illustrated, the different components may be heterogeneous in their sizes.

of the nodes they are joining. Let us recall that, in the node percolation map, the out and in components of individual vertices are defined as the number of vertices (plus itself), s_i, that are reachable from a given vertex and the number of vertices (plus itself), s_o, that can reach that vertex, respectively. The SCC can be thus thought of as the set of vertices with infinite in and out components simultaneously, and the OUT and IN as the set of vertices with infinite in component and infinite out component, respectively (excluding the SCC). Taking this into consideration, we give the following definitions for the principal components of the edge percolation map of random directed networks:

(a) The edge in component, ICE, is the set of edges joining initial and terminal nodes with finite in component and infinite out component. These edges are connecting nodes within the IN.

(b) The in interface, ITF, is the set of edges joining initial nodes with finite in component and infinite out component and terminal nodes with infinite in and out components. These edges are bridging the ICE and the SCE (see below) by connecting nodes in the IN to nodes in the SCC.

(c) The edge strongly connected component, SCE, is the set of edges joining initial and terminal nodes with infinite in and out components. These edges are connecting nodes within the SCC.

(d) The out interface, OTF, is the set of edges joining initial nodes with infinite in and out components and terminal nodes with infinite in component and finite out component. These edges are bridging the SCE and the OCE by connecting nodes in the SCC to nodes in the OUT.

(e) The edge out component, OCE, is the set of edges joining initial and terminal nodes with infinite in component and finite out component. These edges are connecting nodes within the OUT.

The critical point for the simultaneous appearance of the giant node components trivially marks also the emergence of the five giant edge components. In general, the condition $\lambda_m > 1$ characterizes the percolated phase, where λ_m stands for the maximum eigenvalue of a characteristic matrix. In the case of purely directed random networks with two-point correlations it was found to be [18]

$$C^o_{\mathbf{kk'}} = k'_o P_o(\mathbf{k'}|\mathbf{k}),$$

$$C^i_{\mathbf{kk'}} = k'_i P_i(\mathbf{k'}|\mathbf{k}), \qquad (1)$$

where $\mathbf{k} \equiv (k_i, k_o)$, k_i being the incoming number of connections and k_o the outgoing ones. The transition probabilities

$P_i(\mathbf{k'}|\mathbf{k})$ and $P_o(\mathbf{k'}|\mathbf{k})$ measure the likelihood to reach a vertex of degree $\mathbf{k'}$ leaving from a vertex of degree $\mathbf{k}$ using an incoming and an outgoing edge, respectively.

In order to compute the sizes of the different giant components in the number of edges, the already traditional approach used in previous developments is also appropriate with necessary adjustments. The mathematical methodology is based on the generating function formalism while the physical methodology explores the network with branching processes which expand under the locally treelike assumption [9,17,18]. Maximally random purely directed networks with local and two-point correlations will be considered. This implies that the relevant information about the topology of the network is encoded in the joint degree distribution $P(\mathbf{k}, \mathbf{k'})$, where $\mathbf{k}$ is the degree of the initial node and $\mathbf{k'}$ the degree of the terminal node. Equivalently, we consider the degree distribution $P(\mathbf{k})$ along with the transition probabilities $P_i(\mathbf{k'}|\mathbf{k})$ and $P_o(\mathbf{k'}|\mathbf{k})$. These are related through the degree detailed balance condition [18,21]

$$k_o P(\mathbf{k}) P_o(\mathbf{k'}|\mathbf{k}) = k'_i P(\mathbf{k'}) P_i(\mathbf{k}|\mathbf{k'}), \qquad (2)$$

which is fulfilled whenever the network is closed and does not present dangling edge ends. Although it is satisfied for the whole graph, the three node components do not fulfill it separately. If one restricts to consider the nodes within the boundaries of each component along all their connections, dangling ends can be found. The interfaces are just the sets of edges that prevent the node components from fulfilling the detailed balance condition.

Apart from the distributions above, the calculations also rely on the edge joint distribution $G(s_i, s_o; s'_i, s'_o)$ which measures the number of vertices (plus itself), s_o, that are reachable from the initial vertex and the number of vertices (plus itself), s_i, that can reach the initial vertex, simultaneously to the number of vertices (plus itself), s'_o, that are reachable from the terminal vertex and the number of vertices (plus itself), s'_i, that can reach the terminal vertex. Notice that if computations are done for node components, the relevant distribution is $G(s_i, s_o)$ and refers to just one node. According to the definitions above, the relative sizes of the different giant edge components can be formally written as

$$g_{ice} = \sum_{s_i} \sum_{s'_i} G(s_i, s_o = \infty; s'_i, s'_o = \infty),$$

$$g_{oce} = \sum_{s_o} \sum_{s'_o} G(s_i = \infty, s_o; s'_i = \infty, s'_o),$$

$$g_{itf} = \sum_{s_i} G(s_i, s_o = \infty; s'_i = \infty, s'_o = \infty),$$

$$g_{otf} = \sum_{s'_o} G(s_i = \infty, s_o = \infty; s'_i = \infty, s'_o),$$

$$g_{sce} = G(s_i = \infty, s_o = \infty; s'_i = \infty, s'_o = \infty), \qquad (3)$$

where we have made use of the fact that if the terminal node has an infinite out component so it has the initial node and,

analogously, if the in component of the initial node is infinite so will be the in component of the terminal node. These functions can be computed from marginal distributions associated to $G(s_i, s_o; s_i', s_o')$. Their dependence on a given variable indicates that the corresponding in or out component of the corresponding vertex is finite regardless of the sizes of the rest of the involved single node components. In terms of these marginal probabilities, the relative sizes are

$$g_{ice} = \sum_{s_i, s_i'} G(s_i, ; s_i',) - \sum_{s_i, s_i', s_o'} G(s_i, ; s_i', s_o'),$$

$$g_{oce} = \sum_{s_o, s_o'} G(, s_o; , s_o') - \sum_{s_i, s_o, s_o'} G(s_i, s_o; , s_o'),$$

$$g_{itf} = \sum_{s_i} G(s_i, ; ,) - \sum_{s_i, s_i'} G(s_i, ; s_i',) - \sum_{s_i, s_o'} G(s_i, ; , s_o')$$
$$+ \sum_{s_i, s_i', s_o'} G(s_i, ; s_i', s_o'),$$

$$g_{otf} = \sum_{s_o'} G(, ; , s_o') - \sum_{s_o, s_o'} G(, s_o; , s_o') - \sum_{s_i, s_{out}'} G(s_i, ; , s_o')$$
$$+ \sum_{s_i, s_o, s_o'} G(s_i, s_o; , s_o'),$$

$$g_{sce} = 1 - \sum_{s_i} G(s_i, ; ,) - \sum_{s_o'} G(, ; , s_o') + \sum_{s_i, s_o'} G(s_i, ; , s_o'). \quad (4)$$

For instance, $G(s_i, ; s_i',)$ measures the probability of an edge connecting an initial node with finite in component of size s_i to a terminal node with finite in component of size s_i', regardless of the sizes of the out components of connected nodes, that could be finite or infinite (notice that for ease of notation we just left blank the spaces corresponding to the marginalized variables). In a heterogenous network, these functions depend on the degrees of the nodes at the ends of the edge under consideration. Edges connecting nodes in the same degree classes will be considered statistically equivalent, so that these marginal probabilities should be rewritten over joint degree classes, for instance $G(s_i, ; s_i, s_o') = \Sigma_{\mathbf{k},\mathbf{k}'} P(\mathbf{k}, \mathbf{k}') G(s_i, ; s_i, s_o' | \mathbf{k}, \mathbf{k}')$ and analogously for the rest. To calculate these conditional probabilities we have to introduce at this point the probability functions $g_o(s|\mathbf{k})$ and $g_i(s|\mathbf{k})$, which represent the distributions of the number of reachable vertices from a vertex, given that we have arrived to it from another initial vertex of degree $\mathbf{k}$ following one of its outgoing or incoming edges, respectively. These functions are exactly the same as those already introduced in previous works for the computation of the sizes of the IN, OUT, and SCC. The marginal conditional probabilities can then be expressed as functions of these single-node probabilities, that in its turn obey closed equations obtained from an iterative procedure which applies the techniques of random branching processes under the locally treelike assumption. This hypothesis is correct if the length of cycles present in the network is of the order of its diameter, so that the sizes of single node

components can be exposed by subsequent jumps from neighbors to neighbors of neighbors without returning to already visited ones. In this way, the problem can be formally solved in the general correlated case.

As a way of example, it will suffice here to provide the expression of one of the marginal conditional distributions as a function of $g_o(s|\mathbf{k})$ and $g_i(s|\mathbf{k})$ to illustrate the derivation. Assuming the locally treelike condition, one of the two relevant marginal conditional probabilities in the computation of the ICE can be written as

$$G(s_i, ; s_i', s_o' | \mathbf{k}, \mathbf{k}')$$
$$= \sum_{s_1^i \cdots s_{k_i}^i} g_i(s_1^i | \mathbf{k}) \cdots g_i(s_{k_i}^i | \mathbf{k}) \delta_{s_1^i + \cdots + s_{k_i}^i + 1, s_i}$$
$$\times \sum_{s_1'^i \cdots s_{k_i'-1}'^i} g_i(s_1'^i | \mathbf{k}') \cdots g_i(s_{k_i'-1}'^i | \mathbf{k}') \delta_{s_i + s_1'^i + \cdots + s_{k_i'-1}'^i + 1, s_i'}$$
$$\times \sum_{s_1'^o \cdots s_{k_o'}'^o} g_o(s_1'^o | \mathbf{k}') \cdots g_o(s_{k_o'}'^o | \mathbf{k}') \delta_{s_1'^o + \cdots + s_{k_o'}'^o + 1, s_o'}. \quad (5)$$

This expression involves three simultaneous computations: the number of vertices that can reach the initial node, the number of vertices that can reach the terminal node, and the number of nodes that the terminal node can reach itself. The procedure starts from an edge linking nodes of degrees $\mathbf{k}$ and $\mathbf{k}'$ and splits the sets s_i, s_i', and s_o' into the different contributions associated to the corresponding neighbors. For instance, the number of edges that bring to the degree-$\mathbf{k}$ initial node, s_i, can be computed as the sum of the different contributions that can reach each of its k_i incoming neighbors, $s_1^i + \cdots + s_{k_i}^i$. This corresponds to the first set of summations of the three that appear in Eq. (5). Independent equations for the functions g_i and g_o can be found by expanding iteratively this procedure,

$$g_i(s|\mathbf{k}) = \sum_{\mathbf{k}'} P_i(\mathbf{k}'|\mathbf{k}) g_i(s_1|\mathbf{k}') \cdots g_i(s_{k_i'}|\mathbf{k}') \delta_{S_{k_i'}, s},$$

$$g_o(s|\mathbf{k}) = \sum_{\mathbf{k}'} P_o(\mathbf{k}'|\mathbf{k}) g_o(s_1|\mathbf{k}') \cdots g_o(s_{k_o'}|\mathbf{k}') \delta_{S_{k_o'}, s}, \quad (6)$$

where $S_{k_i'} = s_1 + \cdots + s_{k_i'} + 1$ and $S_{k_o'} = s_1 + \cdots + s_{k_o'} + 1$. These equations become tractable using the generating function formalism. In mathematical terms, generating functions are obtained by applying the discrete Laplace space transformation $\hat{f}(z) \equiv \Sigma_s f(s) z^s$. Once transformed for the variables s, Eqs. (6) become closed for $\hat{g}_i$ and $\hat{g}_o$,

$$\hat{g}_i(z|\mathbf{k}) = z \sum_{\mathbf{k}'} P_i(\mathbf{k}'|\mathbf{k}) \hat{g}_i(z|\mathbf{k}')^{k_i'},$$

$$\hat{g}_o(z|\mathbf{k}) = z \sum_{\mathbf{k}'} P_o(\mathbf{k}'|\mathbf{k}) \hat{g}_o(z|\mathbf{k}')^{k_o'}. \quad (7)$$

All summations over finite sizes of the joint conditional size distributions correspond to their generating functions evaluated at $z=1$. Eventually, those depend on $\hat{g}_i(1|\mathbf{k})$ and $\hat{g}_i(1|\mathbf{k})$,

$$\hat{G}(1,;1,1) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_i(1|\mathbf{k})^{k_i}\hat{g}_i(1|\mathbf{k}')^{k_i'-1}\hat{g}_o(1|\mathbf{k}')^{k_o'},$$

$$\hat{G}(1,1;,1) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_i(1|\mathbf{k})^{k_i}\hat{g}_o(1|\mathbf{k})^{k_o-1}\hat{g}_o(1|\mathbf{k}')^{k_o'},$$

$$\hat{G}(,1;,1) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_o(1|\mathbf{k})^{k_o-1}\hat{g}_o(1|\mathbf{k}')^{k_o'},$$

$$\hat{G}(1,;1,) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_i(1|\mathbf{k})^{k_i}\hat{g}_i(1|\mathbf{k}')^{k_i'-1},$$

$$\hat{G}(1,;,1) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_i(1|\mathbf{k})^{k_i}\hat{g}_o(1|\mathbf{k}')^{k_o'},$$

$$\hat{G}(,;,1) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_o(1|\mathbf{k}')^{k_o'},$$

$$\hat{G}(1,;,) = \sum_{\mathbf{k},\mathbf{k}'} P(\mathbf{k},\mathbf{k}')\hat{g}_i(1|\mathbf{k})^{k_i}. \quad (8)$$

These expressions will allow us to compute easily the relative sizes of the different components,

$$g_{ice} = \hat{G}(1,;1,) - \hat{G}(1,;1,1),$$

$$g_{oce} = \hat{G}(,1;,1) - \hat{G}(1,1;,1),$$

$$g_{itf} = \hat{G}(1,;,) - \hat{G}(1,;1,) - \hat{G}(1,;,1) + \hat{G}(1,;1,1),$$

$$g_{otf} = \hat{G}(,;,1) - \hat{G}(,1;,1) - \hat{G}(1,;,1) + \hat{G}(1,1;,1),$$

$$g_{sce} = 1 - \hat{G}(1,;,) - \hat{G}(,;,1) + \hat{G}(1,;,1). \quad (9)$$

Notice that the sizes of the interfaces can also be written as

$$g_{itf} = \hat{G}(1,;,) - \hat{G}(1,;,1) - g_{ice},$$

$$g_{otf} = \hat{G}(,;,1) - \hat{G}(1,;,1) - g_{oce}. \quad (10)$$

The set of Eqs. (7)–(9) determines completely the relative sizes in number of edges of the main giant components of the edge percolation map of two-point correlated purely directed networks.

III. UNCORRELATED PURELY DIRECTED NETWORKS

The formal solution given in the previous section becomes simpler for the classical null model of uncorrelated networks. This will allow us to perform further analytical computations that will be checked against simulation results in order to contrast the accuracy of the theory.

The absence of two-point correlations make it possible to factorize the joint degree distribution, $P(\mathbf{k},\mathbf{k}') = [k_i' k_o P(\mathbf{k})P(\mathbf{k}')]/\langle k_i \rangle^2$, and the conditional degree distributions also simplify

$$P_o(\mathbf{k}'|\mathbf{k}) = \frac{k_i' P(\mathbf{k}')}{\langle k_i \rangle}, \quad P_i(\mathbf{k}'|\mathbf{k}) = \frac{k_o' P(\mathbf{k}')}{\langle k_i \rangle}. \quad (11)$$

In this situation, Eqs. (7) evaluated in $z=1$ reduce to

$$\hat{g}_i(1|\mathbf{k}) \equiv \hat{g}_i(1) = \sum_{\mathbf{k}} \frac{k_o P(\mathbf{k})}{\langle k_i \rangle}\hat{g}_i(1)^{k_i},$$

$$\hat{g}_o(1|\mathbf{k}) \equiv \hat{g}_o(1) = \sum_{\mathbf{k}} \frac{k_i P(\mathbf{k})}{\langle k_i \rangle}\hat{g}_o(1)^{k_o}, \quad (12)$$

so that the relative sizes in number of edges of the different components in the uncorrelated case just depend on the joint degree distribution $P(\mathbf{k})$ and the single-node in and out generating functions $\hat{g}_i(1)$ and $\hat{g}_o(1)$, and can be written as

$$g_{ice} = \sum_{\mathbf{k}} \frac{k_i P(\mathbf{k})}{\langle k_i \rangle}\hat{g}_i(1)^{k_i}[1 - \hat{g}_o(1)^{k_o}],$$

$$g_{oce} = \sum_{\mathbf{k}} \frac{k_o P(\mathbf{k})}{\langle k_i \rangle}\hat{g}_o(1)^{k_o}[1 - \hat{g}_i(1)^{k_i}],$$

$$g_{itf} = \hat{g}_i(1)[1 - \hat{g}_o(1)] - g_{ice},$$

$$g_{otf} = \hat{g}_o(1)[1 - \hat{g}_i(1)] - g_{oce},$$

$$g_{sce} = [1 - \hat{g}_i(1)][1 - \hat{g}_o(1)]. \quad (13)$$

If local correlations are also absent, the expressions above become even simpler,

$$g_{ice} = [1 - \hat{g}_o(1)]\sum_{k_i} \frac{k_i \hat{g}_i(1)^{k_i} P(k_i)}{\langle k_i \rangle},$$

$$g_{oce} = [1 - \hat{g}_i(1)]\sum_{k_o} \frac{k_o \hat{g}_o(1)^{k_o} P(k_o)}{\langle k_i \rangle},$$

$$g_{itf} = [1 - \hat{g}_o(1)]\hat{g}_i(1) - g_{ice},$$

$$g_{otf} = [1 - \hat{g}_i(1)]\hat{g}_o(1) - g_{oce},$$

$$g_{sce} = [1 - \hat{g}_i(1)][1 - \hat{g}_o(1)], \quad (14)$$

with

$$\hat{g}_i(1) = \sum_{k_i} P(k_i)\hat{g}_i(1)^{k_i},$$

$$\hat{g}_o(1) = \sum_{k_o} P(k_o)\hat{g}_o(1)^{k_o}. \quad (15)$$

In order to ascertain the accuracy of the theory, we contrast the analytical results with those obtained from simulating uncorrelated purely directed networks with given joint degree distribution of the form $P(\mathbf{k}) = P(k_i)P(k_o)$. Uncorrelated networks are generated according to a slightly modified version of the Molloy-Reed prescription [7,8]—which is based on the configuration model [22,23] and constructs

maximally random networks with a given degree sequence—to produce directed connections controlling that $\Sigma_i \, k_i = \Sigma_o \, k_o$ and also taking care of avoiding multiple and self-connections.

A. Exponential in- and out-degree distributions

For the first case study, we chose $P(k_i)$ and $P(k_o)$ of the form

$$P(x) = \frac{(1-P_0)^2}{\langle x \rangle}\left(1 - \frac{(1-P_0)}{\langle x \rangle}\right)^{x-1}, \quad x \geqslant 1, \quad (16)$$

with $x \equiv k_i$ and $P_0 \equiv P(0) \equiv P_0^i$, or $x \equiv k_o$ and $P_0 \equiv P(0) \equiv P_0^o$, so that a full analytical solution is available. The sizes of the giant components in the edge percolation map for this particular joint degree distribution just depend on the parameters P_0^i and P_0^o and the average degree $\langle k \rangle = \langle k_i \rangle = \langle k_o \rangle$. Substituting Eqs. (16) into Eqs. (15), the solutions are found to be

$$\hat{g}_i(1) = \frac{P_0^i}{q^i}, \quad \hat{g}_o(1) = \frac{P_0^o}{q^o}, \quad q = 1 - \frac{1-P_0}{\langle k_i \rangle}, \quad (17)$$

and the relative sizes

$$g_{ice} = \frac{\hat{g}_i(1)[1 - \hat{g}_o(1)]}{\langle k_i \rangle^2},$$

$$g_{oce} = \frac{\hat{g}_o(1)[1 - \hat{g}_i(1)]}{\langle k_i \rangle^2},$$

$$g_{itf} = \frac{\hat{g}_i(1)[1 - \hat{g}_o(1)]}{\langle k_i \rangle^2}[\langle k_i \rangle^2 - 1],$$

$$g_{otf} = \frac{\hat{g}_o(1)[1 - \hat{g}_i(1)]}{\langle k_i \rangle^2}[\langle k_i \rangle^2 - 1],$$

$$g_{sce} = [1 - \hat{g}_i(1)][1 - \hat{g}_o(1)]. \quad (18)$$

We compared these results with direct measures of the edge components on a synthetic set of purely directed random networks with $N = 10^5$. We fix the values $P_0^i = 0.4$, $P_0^o = 0.8$, and vary the average degree from $\langle k \rangle = 1$ to $\langle k \rangle = 10$. As Fig. 2 shows, the conformity of our formulas to the simulation results is excellent. Notice also that for this particular choice of the parameters P_0^i and P_0^o the out interface, OCE, is by far the biggest edge component in the percolated phase for all values of the average degree above approximately 1.5, followed with a noticeable difference by the edge strongly connected component, SCE, and the in interface, ICE. In this example, the interfaces are much stronger than the edge in and out components, practically absent for high degrees. This edge percolation map is seen to be quite stable for most of the average degree range.

B. Scale-free in-degree and exponential out-degree distributions

In some real networks, such as the World Wide Web (WWW) [24] for instance, the in-degree distribution exhibits
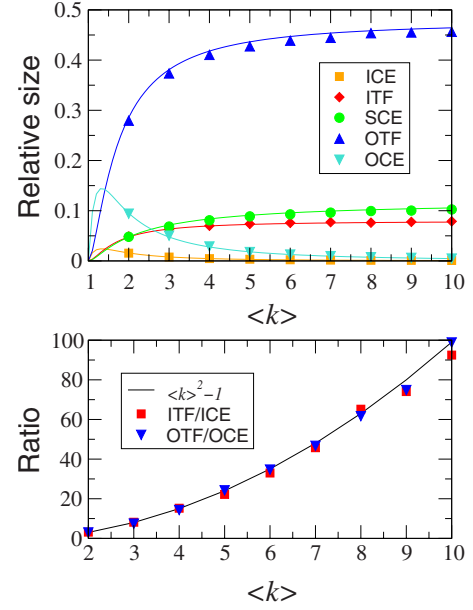

FIG. 2. (Color online) (Top) Relative sizes of the main giant components in the edge percolation map of networks with exponential in- and out-degree distributions as a function of the average degree $\langle k \rangle = \langle k_i \rangle = \langle k_o \rangle$. Simulation results (dots) correspond to one-realization measures on synthetic networks with $N = 10^5$ vertices, $P_0^i = 0.4$, and $P_0^o = 0.8$. Solid lines are the analytical solution equations (17) and (18). (Bottom) Ratio of the relative sizes of the giant interfaces to the edge in and out components as a function of the average degree for the same networks (dots). The solid line corresponds to the analytical ratio equation (20).

a heavy-tailed form well approximated by a power-law behavior $P(k_i) \sim k_i^{-\gamma}$, at the same time that the out-degree distribution $P(k_o)$ presents clear exponential cutoffs. In biology, transcriptional regulatory networks are characterized by the reflexive situation in which an incoming degree distribution that decays faster than a power law can be observed along with a scale-free outgoing degree distribution [25]. It is then particularly interesting to see what happens for power-law in- or out-degree distributions when combined to exponential out or in degree ones. In this example, the in-degree distribution is taken to follow a scale-free form of the type

$$P(k_i) = \frac{(1-P_0^i)}{\zeta(\gamma)k_i^\gamma}, \quad k_i \geqslant 1, \quad (19)$$

where $\zeta(\gamma)$ is the ζ Riemann function and $P(0) = P_0^i$. The out-degree distribution is given again by Eq. (16). The set of Eqs. (15) is solved numerically and plugged into Eqs. (14) to obtain the relative sizes of the edge components and the results are compared to direct measures of the edge percolation map on a set of synthetically generated networks with $N = 5 \times 10^5$ nodes. We take $\gamma = 2.2$, $P_0^o = 0.4$, and vary the average degree $\langle k \rangle = \langle k_i \rangle = \langle k_o \rangle$ until the maximum possible value is reached by adjusting P_0^i. This upper boundary in the average degree is due to the fact that, since $P_0^i = 1 - \langle k_i \rangle \zeta(\gamma)/\zeta(\gamma-1)$, values above the threshold impose a negative P_0^i and are not realizable. The theoretical value for this threshold is $\zeta(\gamma-1)/\zeta(\gamma)$.
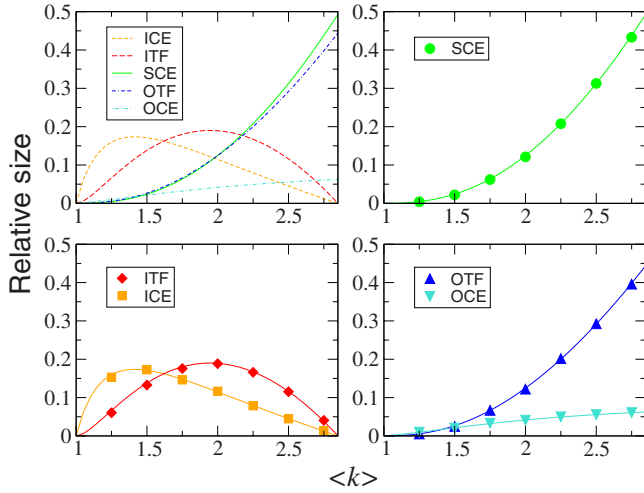
FIG. 3. (Color online) Relative sizes of the giant components in the edge percolation map of uncorrelated networks with scale-free in-degree distribution, $\gamma=2.2$, and exponential out-degree distribution, $P_0^{out}=0.4$, as a function of the average degree $\langle k \rangle=\langle k_i \rangle=\langle k_o \rangle$. Simulation results (dots) correspond to synthetic networks of size $N=5 \times 10^5$ vertices, three realizations for the first two points and one realization for the rest. Solid lines correspond to the numerical solutions of Eqs. (14) and (15).

Once again, our predictions compare extremely well with the measures on the simulated networks, see Fig. 3. Interestingly, and in contrast to what was obtained in the previous example, the edge percolation map changes dramatically depending on the average degree. For small values, but big enough to ensure that the system is in the percolated phase, the edge in-component, ICE, and the in interface, ITF, are predominant. After reaching a maximum they decay to eventually disappear at the average degree threshold. However, the rest of the edge components grow steadily with the average degree, so that for high values the edge strongly connected component, SCE, and the out interface, OTF, dominate.

IV. INTERFACES

Interfaces arise as distinctive elements of the edge percolation map. Their size could be much larger than that of the ICE and OCE, as shown in the bottom graph of Fig. 2. In the particular case of completely uncorrelated networks with exponential in- and out-degree distributions, the ratio of the relative sizes of the interfaces to those of the pure components can be calculated analytically to grow quadratically with the average degree

$$\frac{g_{itf}}{g_{ice}} = \frac{g_{otf}}{g_{oce}} = \langle k \rangle^2 - 1. \tag{20}$$

This result is very interesting because it suggests that, at least for exponential distributions, the traditional IN and OUT components of the node percolation map show a shallow architecture mainly formed by leaf edges emanating from or pointing to the SCC. As a consequence, and to give a mental image, the bow-tie structure of those networks becomes a "hairy ball."

All this points to a rich second-order fine structure that could play a central role in the investigation of how topology is related to functionality. In particular, the conformation of the interfaces and, more specifically, the distinction of leaf edges from connectors, is fundamental in order to assess the efficiency of the global flow or the risks of bottleneck effects. Further discussion about the internal structure of interfaces and possible implications will be provided in a forthcoming work.

Interfaces have a hybrid nature from the point of view of node components. In order to calculate internal average degrees, it is not clear whether they should be assigned to one node component or another. If one considers for instance the subset of nodes in the SCC with all their connections, internal or not, it is found that

$$\frac{\sum_{SCC} k_i}{E} = g_{sce} + g_{itf}, \quad \frac{\sum_{SCC} k_o}{E} = g_{sce} + g_{otf}, \tag{21}$$

where E is the total number of edges in the network. As a consequence, the detailed balance condition, Eq. (2), will not be accomplished in general, $\Sigma_{SCC}k_i \neq \Sigma_{SCC}k_o$, except when both interfaces are of equal sizes. The same happens for the subsets of nodes in the IN and the OUT, where from the point of view of detailed balance there is an excess out and in degree, respectively. The interfaces are precisely responsible for these imbalances.

We explore once more as a null model that can be fully calculated analytically the completely uncorrelated network, with no local or degree-degree correlations. The sizes of the main components in the node percolation map can be written (see Refs. [11,14,18])

$$g_{scc} = 1 - \hat{g}_o(1) - \hat{g}_i(1) + \hat{g}_o(1)\hat{g}_i(1),$$

$$g_{in} = 1 - \hat{g}_o(1) - g_{scc},$$

$$g_{out} = 1 - \hat{g}_i(1) - g_{scc}, \tag{22}$$

where $\hat{g}_i(1)$ and $\hat{g}_o(1)$ are the solutions of Eq. (15). Comparing Eq. (22) and Eq. (14), it is found that the relative sizes in number of nodes of the IN, OUT, and SCC are the same as the relative sizes in number of edges of the ICE+ITF, OCE+OTF, and SCE, respectively. In other words, the average degree of the whole network is preserved in the different components if the in and out interfaces are assigned to the in and out components, respectively. This is in particular valid for the previous examples of uncorrelated networks with exponential or scale-free in-degree distribution and exponential out-degree distribution.

V. CONCLUSIONS

We investigated analytically how edges organize in the percolated phase of purely directed random networks. The general case of local and degree-degree correlations is formally solved and the relative sizes of the five main giant edge components are characterized quantitatively. The results for uncorrelated networks are found to be in very good

agreement with direct measures on synthetic networks, which could present very different edge percolation maps depending on their topological properties. The node percolation map is in this way complemented by the edge percolation map, conforming to a percolation landscape that gives a more detailed topological description of the structure of globally connected systems. Moreover, interfaces play a key role bridging core a periphery. Their particular organization could produce from "hairy ball" percolation landscapes to bottleneck straits, two very different conformations regarding structural efficiency.

In this work, we have restricted to purely directed networks, a good approximation in many cases where flow or transport, when present in both directions, is asymmetric. Nevertheless, the same ideas can be extended to semidirected networks, the most general and realistic ones. For those, analytical calculations could be a bit more intricate due to the nontrivial correlations associated to reciprocity.

[1] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, e-print arXiv:0705.0010v2.

[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[3] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[5] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[6] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).

[7] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[8] M. Molloy and B. Reed, Combinatorics, Probab. Comput. **7**, 295 (1998).

[9] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).

[10] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).

[11] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[12] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).

[13] A. Vázquez and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).

[14] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **64**, 066110 (2001).

[15] P. L. Krapivsky and B. Derrida, Physica A **340**, 714 (2004).

[16] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, Phys. Rev. E **66**, 015104(R) (2002).

[17] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **64**, 025101(R) (2001).

[18] M. Boguñá and M. A. Serrano, Phys. Rev. E **72**, 016106 (2005).

[19] M. A. Serrano and M. Boguñá, Phys. Rev. Lett. **97**, 088701 (2006).

[20] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, S. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).

[21] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).

[22] E. A. Bender and E. R. Canfield, J. Comb. Theory, Ser. A **24**, 296 (1978).

[23] B. Bollobás, Eur. J. Comb. **1**, 311 (1980).

[24] M. A. Serrano, M. Boguñá, A. G. Maguitman, S. Fortunato, and A. Vespignani, ACM Trans. Web **1**, 10 (2007).

[25] A. Vázquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z. N. Oltvai, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **101**, 17940 (2004).