UNIVERSITY OF
BATH

University of Bath

Heterogeneous micro-structure of percolation in sparse networks

Reimer Kühn[1] and Tim Rogers[2]

[1] *Department of Mathematics, King's College London,*
Strand, London WC2R 2LS, UK
[2] *Centre for Networks and Collective Behaviour,*
Department of Mathematical Sciences,
University of Bath, Bath, BA2 7AY, UK.

We examine the heterogeneous responses of individual nodes in sparse networks to the random removal of a fraction of edges. Using the message-passing formulation of percolation, we discover considerable variation across the network in the probability of a particular node to remain part of the giant component, and in the expected size of small clusters containing that node. In the vicinity of the percolation threshold, weakly non-linear analysis reveals that node-to-node heterogeneity is captured by the recently introduced notion of non-backtracking centrality. We supplement these results for fixed finite networks by a population dynamics approach to analyse random graph models in the infinite system size limit, also providing closed-form approximations for the large mean degree limit of Erdős-Rényi random graphs. Interpreted in terms of the application of percolation to real-world processes, our results shed light on the heterogeneous exposure of different nodes to cascading failures, epidemic spread, and information flow.

I. INTRODUCTION

Since the very beginning of the modern fascination with networked systems, researchers have been interested in questions of propagation. Across many applications, the analysis of bond percolation provides a simple framework with which to analyse the capability of a network to transmit information, disease, influence, or failure [1–4]. Early work in this area mainly concentrated on understanding the global properties of percolation in the ensemble average of randomly generated model networks. Surprisingly, detailed results for single instances of fixed networks have only been available relatively recently [5, 6], and very little is known exactly about the responses of individual nodes [7].

We consider bond percolation for fixed networks defined as follows: starting from an arbitrary large (connected) network, we evaluate each edge independently, keeping it with probability ρ and deleting it with probability $1 - \rho$. The largest connected component remaining after this random edge removal process is referred to as the *percolating cluster* or *giant component*; write S for its size measured as a fraction of the total number N of nodes in the network. For large sparse networks it was shown in [5, 8] that this quantity can be computed to close approximation using a message-passing protocol.

Knowledge of S gives global information about the robustness of a network to attack or infection; in particular, there is a critical value of ρ below which no percolating cluster survives in the thermodynamic limit, and $S = 0$. One of the main results of [5, 6] was to identify the percolation threshold ρ_c as the reciprocal of the largest eigenvalue of the *non-backtracking* (or *Hashimoto*) matrix that encodes the relationship between variables in the message-passing equations. The eigenstructure of
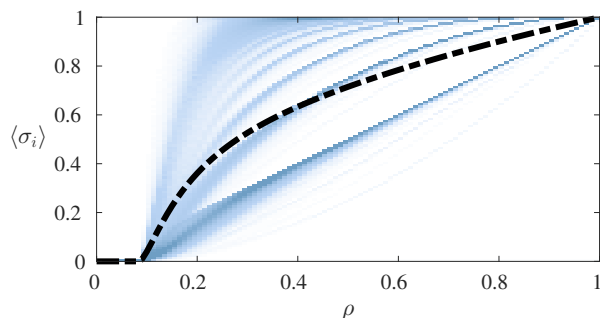

FIG. 1. Micro-structure of percolation in a sample network with 62,586 nodes taken from the *gnutella* file sharing platform [11]. Each vertical slice of the density plot shows the distribution $\varphi(s)$ of probability to appear in the percolating cluster for given edge occupation probability ρ. The thick dashed line shows the expected size of the percolating cluster S, which is equal to the mean of s under φ.

this matrix has received considerable attention recently, having been proposed as an efficient tool for both network clustering [9] and centrality analysis [10].

In this article we will be concerned with more detailed questions about the typical outcomes for individual nodes in the network, when averaged over many instances of the percolation process. For a given random instantiation of percolation, write $\sigma_i = 1$ if node i appears in the largest connected component, and $\sigma_i = 0$ if not. Taking the ensemble average of this variable yields the probability $\langle \sigma_i \rangle$ for node i to appear in the percolating cluster. Heterogeneity in the responses of individual nodes to percolation is captured by the empirical distribution of $\langle \sigma_i \rangle$, defined as

$$\varphi(s) = \frac{1}{N} \sum_i \delta(s - \langle \sigma_i \rangle). \tag{1}$$

Notice that the total fractional size of the percolating cluster is given by the mean of φ, that is, $S = \int s\,\varphi(s)\,\mathrm{d}s$. As mentioned in the introduction, there is a phase transition in S as the edge occupation probability ρ is rasied above a critical threshold ρ_c, determined by the top eigenvalue of the non-backtracking matrix. Recently in [12], the slope of S as it departs from zero at ρ_c was computed as a function of the corresponding top eigenvector. We will be concerned with the responses of indivudual nodes, rather than the average behaviour.

When node i does not appear in the percolating cluster, we write n_i for the size of the component it belongs to, and $\langle n_i \rangle$ for the average over many instances. The node average $\frac{1}{N}\sum_i \langle n_i \rangle$ was again analysed for finite networks in [5], and previously results for the distribution of finite cluster sizes in large random graphs was presented in [2]. The empirical distribution small clusters is defined analogously to φ in Eq. (1);

$$\psi(n) = \frac{1}{N}\sum_i \delta(n - \langle n_i \rangle)\,. \tag{2}$$

It turns out that many networks exhibit extreme differences between nodes in both $\langle \sigma_i \rangle$ and $\langle n_i \rangle$, which are not well-represented by the average value of percolation probability or small cluster size. Figure 1 shows an illustrative example for the probability to appear in the percolating cluster of nodes in a real-world dataset. This behaviour was previously observed in the particular context of epidemic spreading on networks, and exploited in [13] to formulate and analyse a heterogeneous dynamic mean-field approximation of epidemic spreading. Within that approximation the infection probability of nodes is postulated to depend *only* on the degrees of individual nodes, an assumption which allows of a self-consistent solution. This degree-based approximation has become a mainstay in the analysis of epidemic spreading on networks, including in particular also in the search for optimal vaccination strategies; see [14] for a recent review. The heterogeneity was again observed within a cavity formulation of the problem [7] which for the analysis of SIR dynamics allows an exact mapping on bond-percolation. In that work it was found that node degrees play a dominant role in the behaviour of $\langle \sigma_i \rangle$ near $\rho = 1$, but also that the picture becomes much more complex near ρ_c.

In this paper we explore these issues in detail. After recapping the message passing formulation in the next section we move on in section III to consider node variability in fixed finite networks in the neighbourhood of the percolation transition. In this regime we apply a weakly non-linear analysis to compute the probability of a node to appear in the percolating cluster, and the expected size of non-percolating clusters containing that node. In fact, we will show that the measure of non-backtracking centrality proposed in [10] determines the leading-order behaviour of both $\langle \sigma_i \rangle$ and $\langle n_i \rangle$ near percolation. This gives and interesting physical interpretation to a quantity that was originally devised for purely practical reasons. In section IV we move on to use a pop-

ulation dynamics approach compute the distributions of node percolation probability and expected small component size in the ensemble average for large random graphs with specified degree distributions. This analysis is more generally applicable to the whole range of ρ. Numerical simulations reveal the fine structure of the distributions φ and ψ. We are also able to compute a closed-form approximation for the large mean degree limit of Erdős-Rényi random graphs.

II. MESSAGE PASSING

As detailed previously in [5], analysis of the probability generating function of component sizes yields a set of self-consistency equations which can be solved efficiently by iteration. For a network with M edges we define the $2M$-vector $\boldsymbol{H}$ to be the smallest solution in $[0, 1]$ of the system

$$H_{i \leftarrow j} = (1 - \rho) + \rho \prod_{\ell \in \mathcal{N}_j \setminus i} H_{j \leftarrow \ell}\,. \tag{3}$$

Here we write $\mathcal{N}_j$ for the neighbourhood of node j, and the entries of the vector $\boldsymbol{H}$ are indexed by ordered pairs of nodes attached by an edge. The expected size of the percolating cluster is then given by

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \prod_{j \in \mathcal{N}_i} H_{i \leftarrow j}\right)\,. \tag{4}$$

Unpacking the sum above yields additional information about the likely outcomes of the percolation process for individual nodes, specifically

$$\langle \sigma_i \rangle = 1 - \prod_{j \in \mathcal{N}_i} H_{i \leftarrow j}\,. \tag{5}$$

Similarly, for the expected sizes of small clusters one obtains

$$\langle n_i \rangle = 1 + \sum_{j \in \mathcal{N}_i} \frac{H'_{i \leftarrow j}}{H_{i \leftarrow j}}\,, \tag{6}$$

where $\boldsymbol{H'}$ solves

$$H'_{i \leftarrow j} = \rho\left(1 + \sum_{\ell \in \mathcal{N}_j \setminus i} \frac{H'_{j \leftarrow \ell}}{H_{j \leftarrow \ell}}\right)\prod_{\ell \in \mathcal{N}_j \setminus i} H_{j \leftarrow \ell}\,. \tag{7}$$

For fixed finite graphs, both (3) and (7) can be solved quickly by simple iteration, making analysis of percolation computationally tractable even in very large networks.

Notice that equation (3) admits a trivial fixed point in which $H_{i \leftarrow j} \equiv 1$, yielding $S = 0$. This solution may or may not be stable, and the transition boundary exactly corresponds to the percolation threshold ρ_c. In the following section we examine in detail the behaviour of node-specific probability $\langle \sigma_i \rangle$ to belong to the giant cluster and the expected size $\langle n_i \rangle$ of finite clusters in the vicinity of the percolation phase transition.

III. WEAKLY NON-LINEAR ANALYSIS

A. Percolation probability

The value of the percolation threshold for a fixed finite network is revealed by linear stability analysis of the message passing equations. For a given ordered pair $i \leftarrow j$, the pair $k \leftarrow \ell$ will appear on the right hand side of (3) if and only $j = k$ and $\ell \neq i$. The matrix B encoding this relationship is exactly the non-backtracking matrix mentioned in the introduction. The instability of the $H_{i \leftarrow j} \equiv 1$ solution is thus seen to occur at ρ_c satisfying $\rho_c \lambda_{\max}(B) = 1$.

Here we will consider $\rho = \rho_c + \varepsilon$ for small positive ε, and postulate a Taylor expansion for the $2M$-vector of messages

$$\mathbf{1} - \boldsymbol{H} = \varepsilon \boldsymbol{a} + \varepsilon^2 \boldsymbol{b} + \varepsilon^3 \boldsymbol{c} + \mathcal{O}(\varepsilon^4), \qquad (8)$$

where $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are constant vectors, and $\mathbf{1}$ is the vector of ones, i.e. $\mathbf{1} = (1, 1, \dots, 1)^T$. Inserting into both sides of (3) and matching powers of ε to third order we obtain the equations

$$\begin{aligned}
\boldsymbol{a} &= \rho_c B \boldsymbol{a}, \\
\boldsymbol{b} &= \rho_c B \boldsymbol{b} - \frac{\rho_c}{2}\left((B\boldsymbol{a})^2 - B\boldsymbol{a}^2\right) + B\boldsymbol{a}, \\
\boldsymbol{c} &= \rho_c B \boldsymbol{c} - \rho_c \left(B\boldsymbol{a}B\boldsymbol{b} - B(\boldsymbol{ab})\right) \\
&\quad + \frac{\rho_c}{6}\left((B\boldsymbol{a})^3 - 3B\boldsymbol{a}B\boldsymbol{a}^2 + 2B\boldsymbol{a}^3\right) \\
&\quad + B\boldsymbol{b} - \frac{1}{2}\left((B\boldsymbol{a})^2 - B\boldsymbol{a}^2\right).
\end{aligned} \qquad (9)$$

Here we use the notational shorthand of applying multiplication and exponentiation element-wise so that, for example, $\boldsymbol{ab}$ denotes the vector with entries $a_{i \leftarrow j} b_{i \leftarrow j}$ and $\boldsymbol{a}^2$ that with entries $a_{i \leftarrow j}^2$.

The first equation in (9) simply states that $\boldsymbol{a}$ is an eigenvector of B associated to the Frobenius eigenvalue $\lambda_{\max}(B) = 1/\rho_c$. To obtain the proper normalisation for $\boldsymbol{a}$ it is necessary to check the solvability of the second order equation. Let us write $\boldsymbol{a} = \alpha \boldsymbol{v}$, where $\boldsymbol{v}$ is the right Frobenius eigenvector of B with positive entries, summing to one. We also introduce the corresponding left eigenvector $\boldsymbol{u}$, again with positive entries. That is,

$$\boldsymbol{v}/\rho_c = B\boldsymbol{v}, \quad \boldsymbol{u}^T/\rho_c = \boldsymbol{u}^T B, \quad \|\boldsymbol{v}\|_1 = 1. \qquad (10)$$

Following standard conventions by normalising left and right eigenvectors to form a bi-orthonormal system and, having fixed $\|\boldsymbol{v}\| = 1$, we need to choose $\|\boldsymbol{u}\|_1$ such as to achieve $\boldsymbol{u}^T \boldsymbol{v} = 1$. Then, replacing $\boldsymbol{a}$ by $\alpha \boldsymbol{v}$, the second order equation from (9) reads as:

$$\boldsymbol{b} = \rho_c B \boldsymbol{b} - \alpha^2 \frac{\boldsymbol{v}^2}{2\rho_c} + \alpha^2 \frac{\rho_c}{2} B\boldsymbol{v}^2 + \alpha \frac{\boldsymbol{v}}{\rho_c}. \qquad (11)$$

Multiplying through on the left by $\boldsymbol{u}^T 2\rho_c/\alpha$ cancels some terms, yielding $\alpha = 2/\boldsymbol{u}^T \boldsymbol{v}^2(1 - \rho_c)$. Returning to (5),
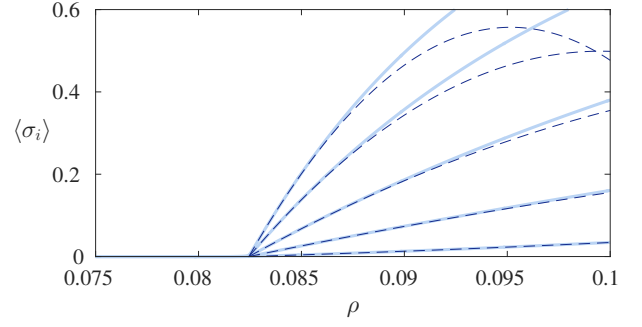

FIG. 2. Solid lines show the probability $\langle \sigma_i \rangle$ to appear in the percolating cluster, compared to the second order theory of Eq. (16) in dashed lines. A selection of results are shown for vertices from a graph with power-law degree distribution, $p(k) \propto k^{-3}$ with $k_{\min} = 2$.

we find the first order approximation

$$\langle \sigma_i \rangle \approx \alpha(\rho - \rho_c) \sum_{j \in \mathcal{N}_i} v_{i \leftarrow j}. \qquad (12)$$

In fact the right hand side here is exactly the so-called "non-backtracking centrality" that was proposed in [10]. If the original system was simply connected to begin with, our choice of normalisation for $\boldsymbol{v}$ implies that $0 < v_{\min} \leq v_{i \leftarrow j} \leq v_{\max} < 1$, entailing that close to the percolation threshold we have lower and upper bounds for $\langle \sigma_i \rangle$ which are proportional to degree and of the form $\alpha(\rho - \rho_c)v_{\min}k_i \lesssim \langle \sigma_i \rangle \lesssim \alpha(\rho - \rho_c)v_{\max}k_i$. If the system was not simply connected to begin with, we have $v_{\min} = 0$, and the lower bound becomes trivial.

By the same methodology we are also able to obtain the curvature of $\langle \sigma_i \rangle$ near criticality. The $\boldsymbol{b}$ equation from (9) can be rewritten as

$$(I - \rho_c B)\boldsymbol{b} = -\frac{\boldsymbol{a}^2}{2\rho_c} + \frac{\rho_c B\boldsymbol{a}^2}{2} + \frac{\boldsymbol{a}}{\rho_c}, \qquad (13)$$

and thus

$$\boldsymbol{b} = (I - \rho_c B)^+ \left(\frac{\rho_c B\boldsymbol{a}^2}{2} - \frac{\boldsymbol{a}^2}{2\rho_c} + \frac{\boldsymbol{a}}{\rho_c}\right) + \beta \boldsymbol{v}, \qquad (14)$$

where $^+$ denotes the Moore-Penrose pseudoinverse and $\beta \in \mathbb{R}$ is a constant yet to be established. Introducing $\boldsymbol{w} = \boldsymbol{b} - \beta \boldsymbol{v}$, we multiply third order equation in (9) through on the left by $\boldsymbol{u}^T \rho_c$ to obtain

$$\begin{aligned}
0 &= -\boldsymbol{u}^T \rho_c \left(\boldsymbol{a} B \boldsymbol{w} - \boldsymbol{a}\boldsymbol{w}\right) + \frac{1}{6\rho_c}\boldsymbol{u}^T \boldsymbol{a}^3 - \frac{\rho_c}{2}\boldsymbol{u}^T(\boldsymbol{a}B\boldsymbol{a}^2) \\
&\quad + \frac{\rho_c}{3}\boldsymbol{u}^T \boldsymbol{a}^3 + \boldsymbol{u}^T \boldsymbol{w} - \frac{1}{2\rho_c}\boldsymbol{u}^T \boldsymbol{a}^2 + \frac{1}{2}\boldsymbol{u}^T \boldsymbol{a}^2 \\
&\quad + \beta\left(\rho_c \boldsymbol{u}^T \boldsymbol{a}\boldsymbol{v} - \boldsymbol{u}^T \boldsymbol{a}\boldsymbol{v} + \boldsymbol{u}^T \boldsymbol{v}\right).
\end{aligned} \qquad (15)$$

The constant β is easily obtained by rearranging.

Having solved for $\boldsymbol{a}$ and $\boldsymbol{b}$, a complete second order expansion for the probability of appearing in the percolating cluster is then given by

$$\langle \sigma_i \rangle \approx (\rho - \rho_c) \sum_{j \in \mathcal{N}_i} a_{i \leftarrow j}$$
$$+ (\rho - \rho_c)^2 \sum_{j \in \mathcal{N}_i} \left(b_{i \leftarrow j} - \frac{1}{2} a_{i \leftarrow j} \sum_{\ell \in \mathcal{N}_i \backslash j} a_{i \leftarrow \ell} \right) \tag{16}$$

Figure 2 shows some numerical examples.

B. Small clusters

Turning attention now to the expected size of finite clusters containing a particular node, we first note that below percolation we have $\boldsymbol{H} = \boldsymbol{1}$ and hence (7) admits the exact solution

$$\boldsymbol{H}' = \rho(I - \rho B)^{-1} \boldsymbol{1} . \tag{17}$$

The matrix inverse here implies an order one pole at ρ_c, around which we develop a first order perturbation theory. If $\rho = \rho_c - \varepsilon$ and $\boldsymbol{H}' = \varepsilon^{-1} \boldsymbol{x} + \boldsymbol{y} + \mathcal{O}(\varepsilon)$ then

$$\varepsilon^{-1} \boldsymbol{x} + \boldsymbol{y} + \mathcal{O}(\varepsilon) = (\rho_c - \varepsilon)(\boldsymbol{1} + \varepsilon^{-1} B \boldsymbol{x} + B \boldsymbol{y} + \mathcal{O}(\varepsilon))$$

Separating orders yields

$$\begin{aligned} \boldsymbol{x} &= \rho_c B \boldsymbol{x} \\ \boldsymbol{y} &= \rho_c B \boldsymbol{y} + \rho_c \boldsymbol{1} - B \boldsymbol{x} . \end{aligned} \tag{18}$$

Evidently $\boldsymbol{x} = \xi^{(-)} \boldsymbol{v}$ for some constant $\xi^{(-)}$. To determine the constant $\xi^{(-)}$, we consult the second order equation, as usual multiplying on the left by $\boldsymbol{u}^T$, to find

$$\xi^{(-)} = \rho_c^2 \|\boldsymbol{u}\|_1 . \tag{19}$$

To examine behaviour on the other side of the critical point, we set $\rho = \rho_c + \varepsilon$ and do the expansion again, this time using the result for α from the previous calculation.

$$\varepsilon^{-1} \boldsymbol{x} + \boldsymbol{y} + \mathcal{O}(\varepsilon)$$
$$= (\rho_c + \varepsilon) \Big(\boldsymbol{1} + B(\varepsilon^{-1} \boldsymbol{x} + \boldsymbol{y} + \mathcal{O}(\varepsilon))(\boldsymbol{1} + \varepsilon \boldsymbol{a} + \mathcal{O}(\varepsilon^2)) \Big)$$
$$\times \Big(\boldsymbol{1} - B(\varepsilon \boldsymbol{a} + \mathcal{O}(\varepsilon^2)) \Big)$$
$$= \varepsilon^{-1} \rho_c B \boldsymbol{x} + \rho_c \Big(\boldsymbol{1} + B(\boldsymbol{x} \boldsymbol{a}) + B \boldsymbol{y} - B \boldsymbol{x} B \boldsymbol{a} \Big)$$
$$+ B \boldsymbol{x} + \mathcal{O}(\varepsilon) . \tag{20}$$

At leading order we again find a multiple of the Frobenius eigenvector, $\boldsymbol{x} = \xi^{(+)} \boldsymbol{v}$. The same trick of multiplying the second order term by the left eigenvector determines the constant:

$$0 = \rho_c \|\boldsymbol{u}\|_1 + \alpha \xi^{(+)} \boldsymbol{u}^T \boldsymbol{v}^2 - \frac{\alpha \xi^{(+)}}{\rho_c} \boldsymbol{u}^T \boldsymbol{v}^2 + \frac{\xi^{(+)}}{\rho_c} \tag{21}$$
$$\Rightarrow \quad \xi^{(+)} = \xi^{(-)} = \xi := \rho_c^2 \|\boldsymbol{u}\|_1 .$$
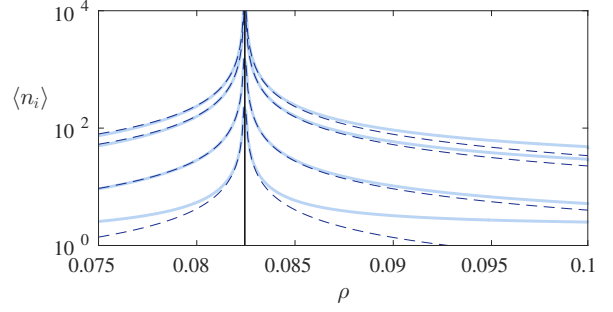

FIG. 3. Solid lines show expected size $\langle n_i \rangle$ of finite clusters containing given nodes, compared to the asymptotic theory of Eq. (22) in dashed lines. A selection of results are shown for vertices from a graph with power-law degree distribution, $p(k) \propto k^{-3}$ with $k_{\min} = 2$.

We conclude that near percolation, typical size of finite clusters involving node i is symmetric around ρ_c, with the asymptotic form

$$\langle n_i \rangle = \frac{\rho_c^2 \|\boldsymbol{u}\|_1}{|\rho - \rho_c|} \sum_{j \in \mathcal{N}_i} v_{i \leftarrow j} + \mathcal{O}(1) . \tag{22}$$

Note that this expression is again proportional to the non-backtracking centrality of node i. Figure 3 shows some numerical examples. Note also that Eq. (22) also allows to obtain an upper bound on the $\langle n_i \rangle$ in the vicinity of the percolation transition. As before, we can exploit the bounds $0 < v_{\min} \le v_{i \leftarrow j} \le v_{\max} < 1$ for systems which are originally simply connected; they entail that in the vicinity of the percolation transition we have diverging lower and upper bounds for $\langle n_i \rangle$ of the form

$$\frac{\rho_c^2 \|\boldsymbol{u}\|_1 v_{\min}}{|\rho - \rho_c|} k_i \lesssim \langle n_i \rangle \lesssim \frac{\rho_c^2 \|\boldsymbol{u}\|_1 v_{\max}}{|\rho - \rho_c|} k_i . \tag{23}$$

Upper and lower bounds for the $\langle n_i \rangle$ therefore proportional to the degree k_i of the site in question, with the constant of proportionality diverging as the transition is approached. As for the $\langle \sigma_i \rangle$ the lower bound becomes trivial for systems which are not simply connected to begin with.

IV. POPULATION DYNAMICS FOR $N \to \infty$

A. Percolation probability

We now proceed to analyse the bond percolation problem in the thermodynamic limit $N \to \infty$ for random networks in the configuration model class. In this limit Eqs. (3) can be interpreted as a stochastic recursion for the values of $H_{i \leftarrow j}$ of randomly chosen neighbouring nodes i and j. A probability density function $\widehat{\varphi}$ for the values of $H_{i \leftarrow j}$ will then have to satisfy a self-consistency
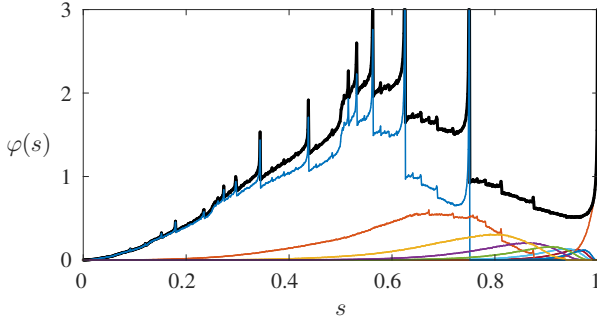
FIG. 4. Distribution $\varphi(s)$ of probabilities to be part of the giant cluster for the percolation problem on a graph with power-law degree distribution, $p(k) \propto k^{-3}$ with $k_{\min} = 2$, at $\rho = 0.5$ (black). Results of population dynamics shown together with its unfolding according to degree for $k = 2, 3, \ldots, 9$, and $\{k \geq 10\}$ (blue, red, green,... from left to right).

equation of the form

$$\widehat{\varphi}(h) = \sum_{k \geq 1} \frac{kp(k)}{c} \int \prod_{\ell=1}^{k-1} d\widehat{\varphi}(h_\ell)\, \delta\left(h - 1 + \rho - \rho \prod_{\ell=1}^{k-1} h_\ell\right) \tag{24}$$

in which $p(k)$ is the degree distribution, and $d\widehat{\varphi}(h_\ell)$ is shorthand for $dh_\ell \widehat{\varphi}(h_\ell)$. It is obtained by averaging the r.h.s of (3) over all realisations for which $H_{i \leftarrow j} \in [h, h + dh]$. The factor of $kp(k)/c$ above is the branching distribution of the configuration model, giving the distribution of the number of additional edges incident upon node j from the randomly chosen pair $i \leftarrow j$. Equation (24) simply states that if the $H_{j \leftarrow \ell}$ appearing on the right hand side of (3) are random variables with distribution $\widehat{\varphi}$, then across neighbourhoods $\mathcal{N}_j$ the resulting $H_{i \leftarrow j}$ should have the same distribution.

The distribution φ of the probabilities to remain in the percolating cluster is expressed in terms of the solution $\widehat{\varphi}$ of Eq. (24) via

$$\varphi(s) = \sum_{k \geq 0} p(k) \int \prod_{\ell=1}^{k} d\widehat{\varphi}(h_\ell)\, \delta\left(s - 1 + \prod_{\ell=1}^{k} h_\ell\right). \tag{25}$$

Equation (24) does not offer much hope for exact solution, however, highly accurate numerical solutions can be obtained by iterating a large sample population $\{h_\omega\}_{\omega=1}^{\Omega}$, a technique known in this context as population dynamics [15]. In Figure 4 we show the results of the population dynamics algorithm for random graphs with a power-law degree distributions of the form $p(k) \propto k^{-3}$, for $k \geq k_{\min} = 2$, and a edge occupation probability of $\rho = 1/2$. Also shown is a partial unfolding of $\varphi(s)$ according to degree k. The distributions show a significant amount of structure, including some sharply peaked 'band-edges' of a type that also appear in the Gnutella data shown in Fig. 1 for sufficiently large ρ.

More insight into the structure of the distribution φ can be gained by disentangling the contributions coming

from nodes of different degrees. We decompose (25) into $\varphi(s) = \sum_{k \geq 0} p(k) \varphi_k(s)$, where

$$\varphi_k(s) = \int \prod_{\ell=1}^{k} d\widehat{\varphi}(h_\ell)\, \delta\left(s - 1 + \prod_{\ell=1}^{k} h_\ell\right). \tag{26}$$

This unfolding according to degree is also shown in Figure 4. It reveals that – to the resolution shown – the peaked band-edges mentioned above are associated with degree 2 and degree 3 vertices, with the exception of the peak at $s = 1$ which originates from $\varphi_{k \geq 10}(s)$, i.e. from large-degree vertices. Moreover, for each degree k, there is an upper cutoff $s_k < 1$ of the probability, beyond which $\varphi_k(s) = 0$. A bound for this cutoff is easily read off from the single instance equations. Indeed Eq. (3) implies that $H_{i \leftarrow j} \geq 1 - \rho$, which via Eq. 5 in turn entails that $\langle \sigma_i \rangle \leq 1 - (1 - \rho)^{k_i}$. Hence we have the bounds $s_k \leq 1 - (1 - \rho)^k$ for the k-dependent cutoffs. Numerical evidence suggests that these bounds are rather sharp. Their exact form is very likely amenable to a more detailed probabilistic analysis.

B. Small clusters

The population dynamics approach is also available for the distribution of expected finite cluster sizes in the thermodynamic limit. From Eqs. (3) and (7) we derive a recursion equation for the joint distribution $\widehat{\psi}(h, h')$ of messages H and H',

$$\widehat{\psi}(h, h') =$$
$$\sum_{k \geq 1} \frac{kp(k)}{c} \int \prod_{\ell=1}^{k-1} d\widehat{\psi}(h_\ell, h'_\ell)\, \delta\left(h - 1 + \rho - \rho \prod_{\ell=1}^{k-1} h_\ell\right)$$
$$\delta\left(h' - \rho\left(1 + \sum_{\ell=1}^{k-1} \frac{h'_\ell}{h_\ell}\right) \prod_{\ell=1}^{k-1} h_\ell\right),$$

in which we use an analogous shorthand for integration measures, $d\widehat{\psi}(h_\ell, h'_\ell) \equiv dh_\ell dh'_\ell\, \widehat{\psi}(h_\ell, h'_\ell)$. This equation generalises (24), as $\widehat{\varphi}$ is recovered as the marginal $\widehat{\varphi}(h) = \int dh'\, \widehat{\psi}(h, h')$. The distribution of finite cluster sizes is then given by $\psi(n) = \sum_{k \geq 0} p(k) \psi_k(n)$, where

$$\psi_k(n) = \int \prod_{\ell=1}^{k} d\widehat{\psi}(h_\ell, h'_\ell)\, \delta\left(n - \left(1 + \sum_{\ell=1}^{k} \frac{h'_\ell}{h_\ell}\right)\right) \tag{27}$$

Figure. 5 shows the results of numerical solution of the population dynamics equations for $\psi(n)$ in the case of a Poisson random graph of mean degree $c = 4$ at bond occupation probability $\rho = 0.3$, together with a deconvolution according to degree k.

The figure exhibits two δ-peaks at $n = 1$ and $n = 1 + \rho = 1.3$, the first corresponding to isolated sites with $k_i = 0$, which remain isolated after deleting bonds, the
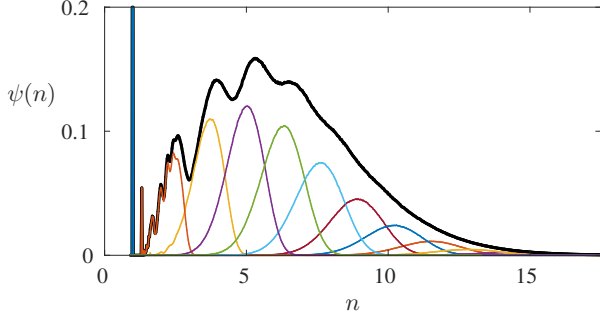
FIG. 5. Distribution $\psi(n)$ of average cluster sizes for percolation on an Erdős-Rényi network of mean degree $c = 4$ at $\rho = 0.3$. Results of population dynamics shown together with its unfolding according to degree for $k = 0, 1, 2, \ldots, 9$, and $\{k \geq 10\}$ (blue, red, green,... from left to right).
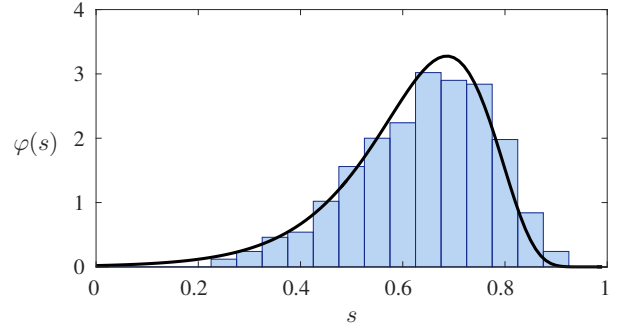

FIG. 6. Comparison of the large c asymptotic given in Eq. (32), with a histogram taken from a finite single instance of a Poisson graph with $N = 1000$ nodes and mean degree $c = 8$. The edge occupancy is $\rho = 0.2$.

second giving the average cluster size of pairs of vertices, which formed isolated dimers in the original graph (before deleting bonds), and consequently appears in $\psi_1(n)$. There are in principle many more such δ-peaks related to the contribution of originally isolated clusters to the $\langle n_i \rangle$-statistics. For the present system their weight is, however, too small to be detected in the continuum at the resolution chosen. The continuous part of $\psi(n)$ reveals several discernible peaks, which according to the deconvolution can be associated with vertices of degrees $k = 1, 2, 3$, and 4, whereas the peaks of $\psi_k(n)$ for larger k are too narrowly separated compared to their widths to give rise to discernible features in the overall distribution $\psi(n)$ at larger n. Each of the degree dependent $\psi_k(n)$ also exhibits an upper cutoff n_k beyond which $\psi_k(n) = 0$.

C. Large mean degree

Further analytical progress can be made in the limit of large mean degree and small edge occupancy; here we develop a single defect approximation [16] for Poisson random graphs. In this approximation scheme we assume that $\widehat{\varphi}(h)$ is for large mean degree c well approximated by a a Dirac delta distribution, $\widehat{\varphi}(h) = \delta(h - h_\star)$. Inserting this ansatz into (24) and integrating over h we obtain $h_\star = 1 - \rho + \rho e^{-c(1 - h_\star)}$, from which it follows that

$$
h_\star = \begin{cases} 1 - \rho - \frac{1}{c} W\left(-c\rho e^{-c\rho}\right) & \rho \geq \rho_c \\ 1 & \rho < \rho_c, \end{cases} \quad (28)
$$

where W is the Lambert W function, and the percolation transition occurs at $\rho_c = 1/c$. To obtain non-trivial behaviour in the limit $c \to \infty$, it is therefore necessary to introduce the scaling $\rho = \varrho/c$ where $\varrho > 1$. Then $h_\star = 1 - \eta/c$, where $\eta = \varrho + W(-\varrho e^{-\varrho})$.

Imagining a single "defect" node i with degree k_i attached to the otherwise homogeneous network, we find the probability of this node to appear in the percolating

cluster to be

$$
\langle \sigma_i \rangle = 1 - h_\star^{k_i} \approx 1 - e^{-\eta k_i/c}. \quad (29)
$$

Taking the Gaussian limit of the Poisson distribution at large mean degree c, we note that the distribution of the ratio k_i/c of a randomly selected node i is well-approximated by a normal random variable x with mean one and variance $1/c$. Under these assumptions we change probability variables from x to s, computing

$$
\varphi(s) \approx \left| \frac{d}{ds} f^{-1}(s) \right| \sqrt{\frac{c}{2\pi}} \exp\left\{ -\frac{c}{2}\left(1 - f^{-1}(s)\right)^2 \right\}, \quad (30)
$$

where $f(x) = 1 - e^{-\eta x}$. The inverse and its derivative are

$$
f^{-1}(s) = -\frac{\log(1 - s)}{\eta}, \quad \frac{d}{ds} f^{-1}(s) = \frac{1}{\eta(1 - s)}. \quad (31)
$$

Putting all this together we arrive at

$$
\varphi(s) \approx \frac{\exp\left\{ -\frac{c}{2}\left(1 + \frac{1}{\eta}\log(1 - s)\right)^2 - \log(1 - s) \right\}}{\sqrt{2\pi/c}\,\eta}. \quad (32)
$$

Although it is exact only in the limit of large c, this approximation holds remarkably well for smaller values; see Fig. 6 for an example with $c = 8$. As $c \to \infty$, the distribution φ approaches a Gaussian with mean $1 - e^{-\eta}$ and variance $\eta^2 e^{-2\eta}/c$.

Applying the same approach to ψ, the distribution of finite clusters, we obtain a simple Gaussian law

$$
\psi(n) \approx \frac{\sqrt{c}}{\sqrt{2\pi}\gamma} \exp\left\{ -\frac{c}{2\gamma^2}(n - 1 - \gamma)^2 \right\}, \quad (33)
$$

where $\gamma = \varrho/(e^{\varrho + W(-\varrho e^{-\varrho})} - \varrho)$.

V. SUMMARY AND DISCUSSION

To summarise, we have taken a new look at the message passing approach [5, 8] to bond percolation on complex networks, revealing a considerable degree of heterogeneity. We have seen that this approach allows one to

determine the distribution of probabilities of individual nodes to belong to the percolating cluster, as well as the distribution of the average sizes of non-percolating clusters to which individual nodes may belong. We found both distributions to be typically broad, so that the *average* percolation probabilities and *average* sizes of finite clusters that are typically reported in analyses of percolation on random networks must be regarded as poor representations of the actual heterogeneity that is present in this problem. We have also obtained deconvolution of both distributions according to degree, in analogy to the deconvolution of sparse random matrix spectra advocated in [17], and found that the component distributions φ_k and ψ_k are themselves non-degenerate (except at $k = 0$) and indeed typically broad as well. A fairly detailed analysis of node-specific percolation probabilities $\langle \sigma_i \rangle$ and average cluster-sizes $\langle n_i \rangle$ was provided in the vicinity of the percolation transition, and formulated in terms of spectral properties of the Hashimoto matrix B, notably the right and left eigenvectors corresponding to its largest eigenvalue.

In the present paper we have used the message passing approach to analyse a specific finite large real world instance, percolation on the network of the Gnutella file sharing platform [11], and presented methods to analyse the problem for random graphs in the configuration model class in the thermodynamic limit of infinite system size, $N \to \infty$. We have presented examples for the distribution of percolation probabilities on a scale free graph with power-law degree distribution, and for the distribution of finite cluster sizes in the case of an Erdős-Rényi network. These two examples can only scratch the surface of the variability of phenomena that might be observed. A few general trends may be noted though. If the original random network ensemble contains finite clusters to begin with, as is indeed the case for Erdős-Rényi networks of finite mean degree, then the resulting distribution $\psi(n)$ of the average finite cluster sizes will contain a family of δ-peaks originating from the distribution of the resulting finite cluster sizes generated by bond removal, whereas the broad continuum is generated by clusters obtained as a result of disconnecting a finite connected set of vertices from the original percolating cluster. Only if the non-percolating fraction in the original graph is sufficiently large will the delta-peaks carry sufficient weight to be detectable in the population dynamics. This is in particular the case for low mean degree Erdős-Rényi graphs. Cluster size-distributions typically get broader as the percolation transition is approached from above or below; the same is true after deconvolution meaning that peaks contributed by individual ψ_k will cease to be discernible in the sum. The same broad trends are observed in $\varphi(s)$ and it's deconvolutions.

We have been able to obtain closed form analytic approximations for the distribution of percolation probabilities and mean cluster sizes in the large mean connectivity limit of Erdő-Rényi graphs, which produces excellent results already for fairly moderate values of c. The same methodology could be adapted to other degree distributions, and should be efficient whenever these distributions become narrow in the large mean degree limit.

Returning to the heterogeneity of percolation probabilities across a network, and the practical relevance of this phenomenon in the context of cascading failures, epidemic spreading, or probabilistic information spreading, we note that the considerable detail which our methods allow to unearth might be useful for instance in the design of optimal vaccination strategies that exploit information beyond degree. Indeed the fact that percolation probabilities conditioned on degrees are themselves broadly distributed is a clear indicator of the fact that the degree based approximation on which the majority of attempts to design optimal vaccination strategies has been based misses important information which could be exploited to improve upon such strategies. We believe that this point is worth investigating.

[1] P. Grassberger, Mathematical Biosciences **63**, 157 (1983).
[2] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, Physical review letters **85**, 5468 (2000).
[3] M. E. Newman, Physical review E **66**, 016128 (2002).
[4] D. J. Watts, Proceedings of the National Academy of Sciences **99**, 5766 (2002).
[5] B. Karrer, M. E. J. Newman, and L. Zdeborová, Phys. Rev. Lett. **113**, 208702 (2014).
[6] K. E. Hamilton and L. P. Pryadko, Phys. Rev. Lett. **113**, 208701 (2014).
[7] T. Rogers, Europhysics Letters **109**, 28005 (2015).
[8] B. Karrer and M. Newman, Physical Review E **82**, 016101 (2010).
[9] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Proceedings of the National Academy of Sciences **110**, 20935 (2013).
[10] T. Martin, X. Zhang, and M. Newman, Physical Review E **90**, 052808 (2014).
[11] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," `http://snap.stanford.edu/data` (2014).
[12] G. Timár, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **95**, 042322 (2017).
[13] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. Jour. B **26**, 521 (2002).
[14] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).
[15] M. Mézard and G. Parisi, The European Physical Journal B-Condensed Matter and Complex Systems **20**, 217 (2001).
[16] G. Biroli and R. Monasson, Journal of Physics A: Mathematical and General **32**, L255 (1999).
[17] R. Kühn, Journal of Physics A : Mathematical and General **41**, 295002 (2008).