An alternative expression of message passing on networks

Peter Mann*
Data Insights AI, CodeBase Edinburgh, Argyle House,
3 Lady Lawson street, Edinburgh, EH3 9DR, United Kingdom

Simon Dobson
School of Computer Science, University of St Andrews, St Andrews, Fife KY16 9SX, United Kingdom
(Dated: April 24, 2025)

Message passing techniques on networks encompasses a family of related methods that can be employed to ascertain many important properties of a network. It is widely considered to be the state of the art formulation for networked systems and advances in this method have a wide impact across multiple literatures. One property that message passing can yield is the size of the largest connected component in the network following bond percolation. In this paper, we introduce an alternative method of finding this value that differs from the standard approach. Like the canonical approach, our method is exact on trees and an approximation on arbitrary graphs. We show that our method lends itself to the description of a variety of generalisations of bond percolation such as sequential percolation and non-binary percolation and can yield information about the local environment of a node in percolation equilibrium that the traditional approach cannot.

I. INTRODUCTION

Bond percolation is a stochastic process in which the edges of a graph, $G(N, E)$, are *occupied* with probability ϕ, where N is the number of nodes and E is the number of edges in the graph. A cornerstone of statistical physics, bond percolation is a mathematical model that can be used to represent a range of phenomena from epidemics spreading through populations, fluid flow through porous media as well as the structural properties of crystalline materials.

Of primary interest to the bond percolation problem are the size of the percolating clusters (sets of nodes connected by occupied edges). In particular, the largest of these connected clusters - the giant connected component (GCC) displays phase properties as ϕ is varied. When ϕ is below some critical value ϕ_c, the GCC contains $\mathcal{O}(1)$ nodes and the network is fractured into lots of disjoint components. As ϕ is increased past ϕ_c, the GCC occupies a finite part of the graph. The fraction of the network occupied by the GCC, $S(\phi)$, is the order parameter of the percolation model, with $S(0) = 0$ and $S(1) = 1$.

Message passing is an analytical technique originating from statistical physics that has seen widespread use throughout computer science since the 1980s across a range of applications [1–6]. Also known as belief propagation, message passing generalises transfer matrix method, the forward-backward algorithm, Kalman filtering and various peeling algorithms. It allows us to answer the question of whether a particular node, i, belongs to the GCC or not; and to find the size of the GCC itself, $S(\phi)$. The answer is exact when the graph is a tree and is otherwise an (often very good) approximate answer. For now we will assume that our graph is in fact a tree. Due to its ubiquity across several domains, advances in the fundamental techniques of the message passing algorithm have a broad range of applications to those scientific fields that consume it, such as computer science,

artificial intelligence, deep learning, network science and statistical physics. In particular various neural network techniques such as Graph Neural Networks (GNNs) and Graph Convolution Networks (GCNs) use message passing to exchange messages between neighbours in order to update the feature representations of their nodes.

Let $\mathcal{N}(i)$ be the set of neighbours of node i. The probability, $\mu_i(\phi)$, that node i does not belong to the GCC is a function of the probabilities that its neighbours $j \in \mathcal{N}(i)$, fail to connect i to the GCC when i is removed from the graph. The function can be written as the product that each edge j fails to lead to the GCC. In turn, this is the sum of the probabilities of the different ways that j might fail to connect i during the percolation process. We can consider the possible ways that an edge between i and j might not lead to the GCC. Firstly, the edge could be unoccupied with probability $1 - \phi$ and so not connect to i in the first place; or, it could be occupied, but j itself not be connected to the GCC, $\phi\mu_{i\leftarrow j}$. The probability μ_i is then given as the product that each neighbour fails

$$\mu_i = \prod_{j \in \mathcal{N}(i)} (1 - \phi + \phi\mu_{i\leftarrow j}). \tag{1}$$

The removal of i is important as it ensures that the neighbours are uncorrelated to one another (assuming G is a tree). This means that Eq 1 *can* in fact be written as the product of edges considered in isolation, over index j. If one of the neighbours did still belong to the GCC despite node is removal, node i must not have been its source of connection to the GCC. The graph without node i is called the *cavity graph* of G at i, denoted $G(i)$. If $G(N, E)$ was not a tree, then two neighbours of i, $j_1, j_2 \in \mathcal{N}(i)$, might still be connected in the cavity graph of i by some longer route, and this would mean that the probabilities $\mu_{i\leftarrow j_1}$ and $\mu_{i\leftarrow j_2}$ were correlated and the subsequent analysis, including 1, would be inexact.

To use Eq 1 we must calculate each probability $\mu_{i\leftarrow j}$. The probability that j is not in the giant component in

the cavity graph at i is equal to the probability that none of j's neighbors, other than i, belong to the GCC

$$\mu_{i\leftarrow j} = \prod_{k\in\mathcal{N}(j)\setminus i} (1 - \phi + \phi\mu_{j\leftarrow k}), \tag{2}$$

where $k \in \mathcal{N}(j)\setminus i$ is the set of neighbours of j apart from node i. There is an equation of the form of Eq 2 for each directed edge pair in the network, totalling a system of $2E$ expressions. This system of equations is known as the message passing equations and can be solved by numerical iteration to its fixed point. Once each $\mu_{i\leftarrow j}$ has been solved for all directed edge pairs in the graph, Eq 1 can be evaluated for each node. In turn, these values can be used to find the size of the GCC

$$S(\phi) = 1 - \frac{1}{N}\sum_i \mu_i. \tag{3}$$

For ϕ less than some critical value ϕ_c, the system has a trivial solution of $S = 0$ and no GCC exists. For $\phi \geq \phi_c$ the solution space bifurcates and an additional fixed point of $\mu_i < 1$ appears, indicating that a GCC exists in the graph. Eq 3 relies on the mutual exclusivity of the binary state equilibrium of the percolation process and for the purpose of this paper we refer to Eqs 1, 2 and 3 as the *canonical approach* to message passing.

In this paper, we introduce an alternative expression of $S(\phi)$ by considering a different family of $2E$ messages ν_i over the edge set. The collective knowledge of $4E$ variables can then be used to introduce a third reformulation for $S(\phi)$ in terms of both μ and ν message sets. In this new re-formulation we gain insight into the local neighbourhood of vertices embedded into the GCC that is not present with the canonical approach. This is because we can examine piecewise the contribution of each edge configuration to the percolation equilibrium. We do not claim that our method is *easier* to apply than the canonical approach: in fact, it is more complicated and computationally slower. However, our re-formulation is particularly suitable when applied to processes that extend the ordinary percolation model. We explore this by considering *sequential percolation* and *non-binary percolation* processes using our approach.

II. THEORETICAL

A. Reformulation of message passing

The essence of our reformulation is as follows. In the canonical approach to message passing, we considered all possible combinations of edge occupations in which a neighbour failed to connect node i to the GCC. In our reformulation, we enumerate all possible combinations that node i *is* successfully connected to the GCC by its set of neighbours. This is a non-trivial enumeration and we must consider the combinatorics of the neighbour set as a whole as the edges are no longer independent of one

another. The probability of connection ν can be treated as an alternative message set over G. Knowledge of both μ and ν message sets gives us complete knowledge over the percolating system, with detail that is not found in the canonical approach.

To begin, let $\pi_i(s, \phi)$ be the probability that the sum of the number of nodes reachable along each edge from i is $s-1$. For brevity, we will drop the explicit dependence on ϕ from our notation, $\pi_i(s, \phi) \equiv \pi_i(s)$. We can write this as

$$\pi_i(s) = \sum_{\{s_j : j\in\mathcal{N}(i)\}} \left[\prod_{j\in\mathcal{N}(i)} \pi_{i\leftarrow j}(s_j)\right] \delta\left(s-1, \sum_{j\in\mathcal{N}(i)} s_j\right), \tag{4}$$

where $\pi_{i\leftarrow j}(s_j)$ is the probability that the edge (i, j) leads to s_j nodes, $\mathcal{N}(i)$ is the set of neighbours of i, $\delta(a, b)$ is the Kronecker delta and $\{s_j : j \in \mathcal{N}(i)\}$ is the set of all values of s_j for each neighbour j to node i. To calculate $\pi_i(s)$ we require a closed form expression for $\pi_{i\leftarrow j}(s_j)$. Let us distinguish three kinds of edges: i) edges that are occupied and lead to the GCC, ii) edges that are occupied and lead to a finite cluster in $G(i)$, and iii) edges that are not occupied, see Fig 1.
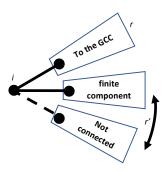

FIG. 1. A graphical representation of the enumeration. Solid edges are occupied whilst dashed edges are unoccupied.

Partitioning the neighbours $\mathcal{N}(i)$ into those that lead to the GCC, X, and those that do not, $\bar{X}$, we can write

$$\prod_{j\in\mathcal{N}(i)} \pi_{i\leftarrow j}(s_j) = \prod_{r\in X} \pi_{i\leftarrow r}(s_r) \prod_{r'\in\bar{X}} \pi_{i\leftarrow r'}(s_{r'}). \tag{5}$$

The set $\bar{X}$ is the complement set of X, containing all members of $\mathcal{N}_i$ that are not in X. Variables r and r' index the members of X and $\bar{X}$, respectively. Substituting this into Eq 4 we find

$$\pi_i(s) = \sum_X \sum_{\{s_r : r\in X\}} \sum_{\{s_{r'} : r'\in\bar{X}\}} \left[\prod_{r\in X} \pi_{i\leftarrow r}(s_r) \prod_{r'\in\bar{X}} \pi_{i\leftarrow r'}(s'_r)\right]$$
$$\times \delta\left(s-1, \sum_{r\in X} s_r + \sum_{r'\in\bar{X}} s'_r\right), \tag{6}$$

where the summation over X is over all non-empty subsets of $\mathcal{N}(i)$. The set $\{s_r : r \in X\}$ is the set of all possible

values of s_r, and similarly for $\{s_{r'} : r' \in \bar{X}\}$. The generating function $\nu_i(z)$ for this expression is

$$\nu_i(z) = \sum_{s=1}^{\infty} \pi_i(s) z^s, \tag{7}$$

which is the probability that node i belongs to the GCC in the percolation equilibrium. Inserting the following logic

$$z^s = z z^{s-1} = z z^{\sum_r s_r + \sum_{r'} s_{r'}}, \tag{8}$$

we have

$$\nu_i(z) = z \sum_X \left(\prod_{r \in X} \left[\sum_{\{s_r : r \in X\}} \pi_{i \leftarrow r}(s_r) z^{\sum_r s_r} \right] \right.$$
$$\left. \times \prod_{r' \in \bar{X}} \left[\sum_{\{s_{r'} : r' \in \bar{X}\}} \pi_{i \leftarrow r'}(s_{r'}) z^{\sum_{r'} s_{r'}} \right] \right). \tag{9}$$

The inner summation and the product can be exchanged due to the independence of each edge in the cavity graph technique [7]. This quantity can be enumerated by accounting for all of the configurations of occupied edges that connect the focal node to the GCC, see Fig 2.

FIG. 2. A graphical representation of the first few terms of the polynomials that arise during the direct enumeration of all configurations that connect the focal node to the GCC. Occupied edges are solid lines whilst unoccupied edges are dashed; in each case the focal node i is the bottom node. Configurations that are related by symmetry are enclosed in brackets subscripted by their degeneracy.

Defining $\nu_{i \leftarrow r}(z)$ as the probability that neighbour r is connected to the GCC

$$\phi^{|X|} \nu_{i \leftarrow r} = \sum_{\{s_r : r \in X\}} \pi_{i \leftarrow r}(s_r) z^{\sum_r s_r}, \tag{10}$$

we finalise the expression for $\nu_i(z)$ as

$$\nu_i(z) = z \sum_{l=1}^{|\mathcal{N}(i)|} \sum_{X \in Y_l} P_\phi(X), \tag{11}$$

where $P_\phi(X)$ is the following probability

$$P_\phi(X) = \left(\phi^{|X|} \prod_{r \in X} \nu_{i \leftarrow r} \prod_{r' \in \bar{X}} [1 - \nu_{i \leftarrow r'} \phi] \right), \tag{12}$$

and where Y_l is the set of all subsets of length l of the set of neighbours $\mathcal{N}(i)$ without replacement; set X belongs to Y_l, $|X|$ is the cardinality of X and similarly for $|\mathcal{N}(i)|$. Note the leading factor of $\phi^{|X|}$ in Eq 10 is because there

must be exactly $|X|$ occupied edges for a given configuration of set X over $\mathcal{N}(i)$. The probability $P_\phi(X)$ is simply the product of probabilities that a subset X of the neighbours in $G(i)$ connects to the GCC and a subset $\bar{X}$ does not. The summation over $X \in Y_l$ simply accounts for the different configurations that the subset of size l could be arranged in.

The probabilities $\nu_{i \leftarrow r}$ are the probabilities that neighbour r connects i to the GCC, which can be solved by writing a $2E$-dimensional message passing system of equations of the form

$$\nu_{i \leftarrow r}(z) = z \sum_{l=1}^{|\mathcal{N}(r) \setminus i|} \sum_{T \in W_l \setminus i} \left(\phi^{|T|} \prod_{t \in T} \nu_{r \leftarrow t} \prod_{t' \in \bar{T}} [1 - \nu_{r \leftarrow t'} \phi] \right), \tag{13}$$

where $W_l \setminus i$ is the set of subsets of length l of the set of neighbours $\mathcal{N}(r)$ of node r apart from node i and $|T|$ is the cardinality of set T. As previously, $\bar{T}$ is the complement set of T, containing all nodes in $\mathcal{N}(r) \setminus i$ that do not belong to T. From these quantities, the expectation value for the size of the GCC is given by

$$S(\phi) = \frac{1}{N} \sum_i \nu_i(1), \tag{14}$$

where N is the number of nodes in the graph.

If we like, these new variables can be combined with the original set from Eq 2 to re-write Eq 13 compactly in terms of $4E$ variables as

$$\nu_{i \leftarrow r}(z) = z \sum_{l=1}^{|\mathcal{N}(r) \setminus i|} \sum_{T \in W_l \setminus i} \left(\phi^{|T|} \prod_{t \in T} \nu_{r \leftarrow t} \prod_{t' \in \bar{T}} (1 - \phi + \mu_{r \leftarrow t'} \phi) \right), \tag{15}$$

although this substitution is not required to solve for $\nu_{i \leftarrow r}(z)$. With reference to Fig 1, $1 - \phi$ accounts for the unoccupied edges, whilst $\mu_{r \leftarrow t'} \phi$ accounts for the edges that lead to finite clusters. Comparison of Eqs 14 and 3 yields the following relation between the two descriptions

$$1 = \frac{1}{N} \sum_i \nu_i(\phi) + \frac{1}{N} \sum_i \mu_i(\phi). \tag{16}$$

This final relation subtly exhibits the mutual exclusivity of the two analytical descriptions of the percolation equilibrium and shows that we have complete knowledge of the system by accounting for all possible configurations of the equilibrium. We show a comparison of the two approaches alongside simulation in Fig 3.

Below the critical point, the probability that a node belongs to the GCC is zero, since a GCC does not exist in the network. The system described by Eq 13 has the trivial solution $0 = 0$. At the critical value of the control parameter, ϕ_c, the solution space bifurcates and non-zero solutions can be found. By linearising Eq 13 around $\nu_{i \leftarrow r} \approx \epsilon_{i \leftarrow r}$ for some small parameter $\epsilon_{i \leftarrow r}$, the system at the critical point is given by

$$\epsilon_{i \leftarrow r} \approx \sum_{l=1}^{|\mathcal{N}(r) \setminus i|} \sum_{T \in W_l \setminus i} \left(\phi_c^{|T|} \prod_t \epsilon_{r \leftarrow t} \right) \Bigg|_{z=1}. \tag{17}$$
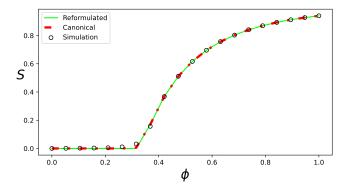
FIG. 3. A comparison of the calculated values of S using both the canonical and reformulated message passing models alongside Monte Carlo simulations of bond percolations. The two methods predict the same numerical result to machine precision. The network is an Erős-Renyi graph with 5000 nodes and an average degree of 3 whilst simulations are the average of 100 repeats of percolation.

Solving this expression numerically for ϕ yields the critical point ϕ_c and it is dominated by terms linear in $\epsilon_{i \leftarrow r}$. This means that configurations of occupied edges that have single occupancies have the highest measure associated around a GCC node at the critical point and therefore that the GCC is expected to be tree-like at this point.

The relation of this expression to the Hashimoto non-backtracking matrix and the implications for the community detection and centrality literature [8, 9], remains unknown.

B. The partition function $\mathcal{Z}_{i,\phi}$

The polynomial $\sum_{l,X} P_\phi(X)$ in Eq 11 is an important contribution to our understanding of the local environment of a node within the GCC of a percolation equilibrium. It can be regarded as a partition function $\nu_i(1) = \mathcal{Z}_{i,\phi}$ for the configurations of neighbours that also belong to the GCC with node i and we will denote it as follows

$$\mathcal{Z}_{i,\phi} = \sum_{l=1}^{|\mathcal{N}(i)|} \sum_{X \in Y_l} P_\phi(X). \qquad (18)$$

Here we will show that $\mathcal{Z}_{i,\phi}$ contains more information than can be obtained via the canonical approach and we will later use it to investigate the fine structure of the GCC surrounding a given node in the network.

One of the primary benefits of Eq 11 over Eq 1 is that it contains each possible configuration of edges that a node might experience in the percolation equilibrium. Immediately, this allows us to identify which edge configurations are associated with the highest measure and to better investigate occupation patterns across the network, tailored to each node and region in possibly heterogeneous

environments. This yields information on the fine structure of a network beyond that which can be achieved using the canonical approach.

Whilst the sets Y_l and X from Sec II A can easily be enumerated computationally for arbitrary $k = |\mathcal{N}(i)|$, for small degrees they can be written by hand also. For example, when $k = 1$ we have $P(\{j_1\}) = \phi \nu_{i \leftarrow j_1}$ for neighbour $j_1 \in \mathcal{N}(i)$. For $k = 2$ and $X \in [\{j_1\}, \{j_2\}, \{j_1, j_2\}]$ we have

$$\mathcal{Z}_{i,\phi} = \phi \nu_{i \leftarrow j_1} (1 - \phi \nu_{i \leftarrow j_2})$$
$$+ \phi \nu_{i \leftarrow j_2} (1 - \phi \nu_{i \leftarrow j_1}) + \phi^2 \nu_{i \leftarrow j_1} \nu_{i \leftarrow j_2}.$$

For $k = 3$ with neighbours $j_1, j_2, j_3 \in \mathcal{N}(i)$ we have

$$X \in [\{j_1\}, \{j_2\}, \{j_3\}, \{j_1, j_2\}, \{j_1, j_3\},$$
$$\{j_2, j_3\}, \{j_1, j_2, j_3\}],$$

from which we obtain

$$\mathcal{Z}_{i,\phi} = \phi^3 \nu_{i \leftarrow j_1} \nu_{i \leftarrow j_2} \nu_{i \leftarrow j_3}$$
$$+ \phi^2 \big[\nu_{i \leftarrow j_1} \nu_{i \leftarrow j_2} (1 - \nu_{i \leftarrow j_3} \phi)$$
$$+ \nu_{i \leftarrow j_1} (1 - \nu_{i \leftarrow j_2} \phi) \nu_{i \leftarrow j_3}$$
$$+ (1 - \nu_{i \leftarrow j_1} \phi) \nu_{i \leftarrow j_2} \nu_{i \leftarrow j_3} \big]$$
$$+ \phi \big[\nu_{i \leftarrow j_1} (1 - \nu_{i \leftarrow j_2} \phi)(1 - \nu_{i \leftarrow j_3} \phi)$$
$$+ (1 - \nu_{i \leftarrow j_1} \phi) \nu_{i \leftarrow j_2} (1 - \nu_{i \leftarrow j_3} \phi)$$
$$+ (1 - \nu_{i \leftarrow j_1} \phi)(1 - \nu_{i \leftarrow j_2} \phi) \nu_{i \leftarrow j_3} \big].$$

It can be shown that these expressions are numerically equivalent to those obtained from the polynomial generated by the canonical approach from Eq 1 by the following expression

$$1 = \mathcal{Z}_{i,\phi} + \prod_j \big(1 - \phi + \mu_{i \leftarrow j} \phi \big). \qquad (19)$$

However, whilst the polynomial in the canonical approach sums up the ways node i fails to connect to the GCC, $\mathcal{Z}_{i,\phi}$ instead enumerates the configurations where node i does connect.

There are a number of different ways to extract useful information about the network from this polynomial and we will detail these next. Firstly, we can obtain the probability $\mathcal{Q}_{X'}(X)$ that a particular configuration X' of neighbours appears around a given node i in the GCC

$$\mathcal{Q}_{i,X'}(X) = \frac{1}{\mathcal{Z}_{i,\phi}} P_\phi(X'). \qquad (20)$$

This is a useful expression to examine the measure associated to each microstate that the focal node could be in.

Next, we can find the probability that a particular node $j \in \mathcal{N}(i)$ belongs to the GCC along with node i by summing over all configurations that contain j

$$\mathcal{R}_{i,j}(X) = \frac{1}{\mathcal{Z}_{i,\phi}} \sum_{X'} P_\phi(X'), \qquad (21)$$

with

$$\{X' \in Y_l : j \in X'; \ l \in 1, \ldots, |\mathcal{N}(i)|\}. \tag{22}$$

This can also be generalised to an arbitrary subset of is neighbours

$$\mathcal{R}_{i,j_1,\ldots,j_n}(X) = \frac{1}{\mathcal{Z}_{i,\phi}} \sum_{X'} P_\phi(X'), \tag{23}$$

such that

$$\{X' \in Y_l : (j_1, \ldots, j_n) \in X'; \ l \in n, \ldots, |\mathcal{N}(i)|\}. \tag{24}$$

Finally, we can also obtain the expectation value $\mathcal{T}_{i,k'}(X)$ that the focal node has degree k' within the GCC by constraining the summation over X' to those sets such that $|X'| = k'$

$$\mathcal{T}_{i,k'}(X) = \frac{1}{\mathcal{Z}_{i,\phi}} \sum_{X'} P_\phi(X'), \tag{25}$$

and

$$\{X' \in Y_{k'} : |X'| = k'\}. \tag{26}$$

This is useful for investigating the structure of the giant component and can be thought of as the ϕ-dependent degree distribution for a given node in the GCC. As with any distribution we have $\sum_{k'} \mathcal{T}_{i,k'} = 1$.

We can use Eq 25 to compute the average degree for a given ϕ that a node will have in the GCC. For example in Fig 6 (top) we show how the components of $\mathcal{T}_{i,k'}$ change as a function of ϕ giving insight into the substructure of the GCC from the perspective of the local environment of the node within it. The average number of occupied edges of a node, i, in the GCC is given by the following expectation value

$$\langle k_i' \rangle = \sum_{k'} k' \mathcal{T}_{i,k'}. \tag{27}$$

We anticipate that the assortativity of the GCC can equally be investigated using this quantity in a similar, yet simpler fashion than [10].

With all of these functions, $\mathcal{Q}_{i,X'}$, $\mathcal{R}_{i,j_1,\ldots,j_n}$ and $\mathcal{T}_{i,k'}$ the important point to note is that we cannot obtain this information from the canonical approach. Whilst we can numerically obtain the value of $\mathcal{Z}_{i,\phi}$ by rearranging Eq 19, it is only because we have access to the full polynomial from the right hand side of Eq 18 that we can write the numerator of these quotients.

III. APPLICATIONS

In this section we will consider some examples of the reformulation that exhibit its advantage over the canonical approach.
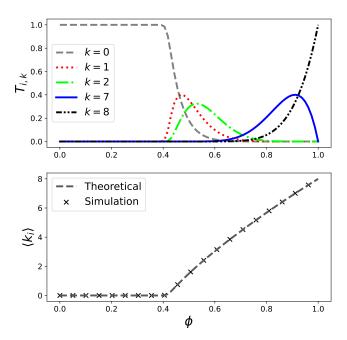

FIG. 4. (Top) Components of the degree distribution of occupied edges $T_{i,k}$ from Eq 25 for a degree 8 node in the GCC as a function of ϕ. (Bottom) The average degree given by Eq 27 for the same node alongside simulation. The network is an Erdős-Renyi graph with 10000 nodes and an overall average degree of 2.5.

A. Sequential percolation

Let us define *independent sequential percolation* (ISP) as the dynamical process of n sequential and independent bond percolation processes over a network. At each pass of percolation, edges are occupied uniformly with probability $\phi_\ell \in \vec{\phi} = (\phi_1, \ldots, \phi_n)$. The fraction of the network occupied by the largest component of nodes that are connected by maximally occupied edges can be calculated by passing the following messages over the edges of the graph

$$\sigma_{i \leftarrow r}(z) = z \sum_{l=1}^{|\mathcal{N}(r) \backslash i|} \sum_{T \in W_l \backslash i} \left(\psi^{|T|} \prod_{t \in T} \sigma_{r \leftarrow t} \prod_{t' \in \bar{T}} [1 - \sigma_{r \leftarrow t'} \phi_\ell] \right), \tag{28}$$

with W_l taking its definition as before and where

$$\psi = \prod_{\ell=1}^{n} \phi_\ell. \tag{29}$$

The size of the GCC is given by $S(\vec{\phi}) = \sum_i \sigma_i(1)/N$, with

$$\sigma_i(z) = z \sum_{l=1}^{|\mathcal{N}(i)|} \sum_{X \in Y_l} \left(\psi^{|X|} \prod_{r \in X} \sigma_{i \leftarrow r} \prod_{r' \in \bar{X}} [1 - \sigma_{i \leftarrow r'} \phi_\ell] \right), \tag{30}$$

The fraction of nodes that were occupied by only a subset of the percolation processes can also be calculated by allowing only those ϕ_ℓ pertaining to the subset in

the product defining ψ. This is an immediate generalisation of Newman and Ferrario's coinfection model [11] from two coinfecting diseases to n. By modulating ψ we can also obtain quantities that are not available via their method, such as the fraction of the population that was coinfected by $n' < n$ diseases. We show an example of the ISP process in Fig 6 (top) for $n = 2$; the circle markers track the simulation results of the primary percolation, whilst the triangle scatter points are the secondary process.

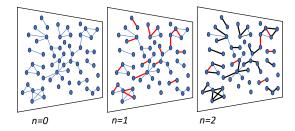FIG. 5. A depiction of the sequential percolation process for $\phi_1 < \phi_2$.

Treating the sequential processes as the composition of *independent* dynamical processes is a simplification of the reality of many phenomenological process. Often, temporally separated processes are *coupled* together in some manner. For instance, seasonal diseases can confer partial immunity to an individual, making them less likely to contract a subsequent strain [12]; conversely, immunosuppressant effects could act to increase the likelihood of catching something else [13, 14]. To model this behaviour we must occupy edges with probabilities based on their occupation state of all of the previous percolations, which is analytically a harder task. Due to this, the ℓth percolation process in the sequence observes $2^{\ell-1}$ possible edge types and therefore requires this many parameters. For the case of $n = 2$, we require 3 bond occupation parameters: ϕ_1 for the first process; and ϕ_2 and ϕ_3 for the second process that occupies 1-occupied and 1-unoccupied edges, respectively. We can compute the GCC of maximally occupied edges following *coupled sequential percolation* (CSP) by passing the following message sets over the edges

$$\varphi_{i \leftarrow r} = \sum_{l=1}^{|\mathcal{N}(r)\backslash i|} \sum_{T \in W_l \backslash i} \left(\prod_{t \in T} f_{r \leftarrow t}(\vec{\phi}) \prod_{t' \in \bar{T}} g_{r \leftarrow t'}(\vec{\phi}) \right) / \nu_{i \leftarrow r},$$
(31a)

$$\omega_{i \leftarrow r} = \prod_{t \in \mathcal{N}(r)\backslash i} g_{r \leftarrow t}(\vec{\phi}) / \mu_{i \leftarrow r}$$
(31b)

with

$$f_{r \leftarrow t}(\vec{\phi}) = \nu_{r \leftarrow t} \phi_1 h_\varphi(\phi_2), \tag{32}$$
$$h_y(\phi) = 1 - \phi + y_{r \leftarrow t}\phi, \tag{33}$$

and

$$g_{r \leftarrow t}(\vec{\phi}) = \mu_{r \leftarrow t}(1 - \phi_1) + (1 - \mu_{r \leftarrow t})(1 - \phi)h_\varphi(\phi_2)$$
$$+ \mu_{r \leftarrow t}\phi_1 h_\omega(\phi_3). \tag{34}$$

Expression 31a accounts for nodes in the GCC of the first process and reduces to Eq 15 when $\phi_2 = \phi_3 = 0$; whilst Eq 31b accounts for nodes in the residual graph and reduces to Eq 2 under the same limit. The size of the GCC formed by maximally occupied edges is given by

$$S = \frac{1}{N} \sum_i \nu_i - \frac{1}{N} \sum_i \varphi_i, \tag{35}$$

where ν_i is defined in Eq 11 and φ_i is

$$\varphi_i = \sum_{l=1}^{|\mathcal{N}(i)|} \sum_{X \in Y_l} \left(\prod_{r \in X} f_{i \leftarrow r}(\vec{\phi}) \prod_{r' \in \bar{X}} g_{i \leftarrow r'}(\vec{\phi}) \right). \tag{36}$$

This model is the 2-strain message passing equivalent of Eq 37 in [15] re-written in the alternative formulation presented here. It can instantly be seen that the reformulated description presented here is far simpler.

B. Non-binary percolation

Let us now consider the application of our expression to a third kind of percolation model. The equilibrium of ordinary bond percolation is a binary state fixed point containing occupied and unoccupied edge states. Let us next consider a Potts-like process [16, 17] over the edges of network in which edges can belong to one of m states $q_1, \ldots, q_m$ with occupation probabilities $\vec{\phi} = (\phi_1, \ldots, \phi_m)$ with the constraint that $\sum_\ell \phi_\ell = 1$ for $\ell \in [1, m]$. When $m = 2$ the model reduces to ordinary bond percolation and $\vec{\phi} = (\phi, 1 - \phi)$. For each edge-state we are concerned with finding the largest connected component that can be created by following edges of a particular state. Nodes may belong to the GCCs of multiple edge-states. For edge-state q_ℓ let $\nu_{\ell,i}$ be the probability that node i with degree $|\mathcal{N}_i|$ belongs to the GCC of q_ℓ-state occupied edges

$$\nu_{\ell,i}(\vec{\phi}) = \sum_{\substack{X_1, \ldots, X_m = \mathcal{N}(i) \\ X_\ell \neq \varnothing}} \left(\prod_{k=1}^m \prod_{r \in X_k} \nu_{k, i \leftarrow r} \phi_k \right), \tag{37}$$

with the condition $X_\ell \neq \varnothing$ indicating the node belongs to the q_ℓ-state GCC. The summation is over all covering disjoint subsets (partitions) of the set of neighbours $\mathcal{N}(i)$ such that X_ℓ is not empty

$$\left\{ (X_1, \ldots, X_m) \mid \bigcup_k X_k = \mathcal{N}(i), \ X_\ell \neq \varnothing \right\} \tag{38}$$

A given covering set $X_1, \ldots, X_m$ represents an edge configuration of the neighbours among the states. Index k

runs over the edge states $1, \ldots, m$ and $r \in X_k$ is an index over the neighbours that are in state q_k for the edge configuration. The message updates $\nu_{k,i \leftarrow r}$ are given by

$$\nu_{k,i \leftarrow r}(\vec{\phi}) = \sum_{\substack{X_1, \ldots, X_m = \mathcal{N}_r \backslash i \\ X_k \neq \varnothing}} \left(\prod_{j=1}^{m} \prod_{t \in X_j} \nu_{j, r \leftarrow t} \phi_j \right), \quad (39)$$

which is the probability that at least one of neighbour r's neighbours (other than i) connects to the GCC of the k edge-state. The size of the giant connected component for the ℓth state is given by

$$S_\ell(\vec{\phi}) = \frac{1}{N} \sum_i \nu_{\ell,i}(\vec{\phi}). \quad (40)$$

We show an example of this in Fig 6 (bottom) for $n = 3$ with the following bond occupation probabilities for the three states: $\phi, \alpha(1-\phi)$ and $1-\phi-\alpha(1-\phi)$ setting $\alpha = 0.45$.
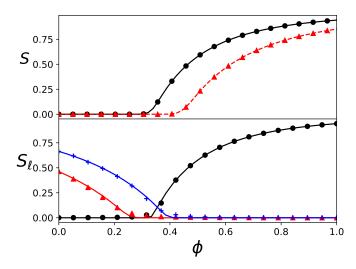

FIG. 6. Results for the ISP model (top) and the Potts-like model (bottom) on Erdős-Renyi graphs with $10,000$ nodes. Scatter points are the average of 100 repetitions of Monte Carlo simulation whilst lines are the theoretical predictions of our models when passing Eqs 30 and 37, respectively. The black circles are $\ell = 1$, the red triangles are $\ell = 2$ and the blue crosses are $\ell = 3$.

IV. CONCLUSION

To conclude, we have introduced an alternative exact analytical formulation of the message passing algorithm on graphs. Our theoretical models provide additional insight into the environment of nodes embedded into the GCC beyond those that the traditional approach affords. We have shown that this approach lends itself to multiple extensions of message passing on graphs, in particular to systems such as sequential percolation and Potts-like models with multiple edge states. In both of those cases, the model presented here is simpler and more tractable than those presented previously and is arguably the natural description of nodes that are embedded in the GCC.

In future work we intend to explore higher-order networks using this framework, such as motif [18], neighbourhood [19], hypergraph and simplicial [20] models. In particular, the ability to pick apart different occupation probabilities, both *along* and *through* each higher-order structure, could yield important information on community structure, centrality metrics or optimal motif projection [18, 21]. Additionally, we would like to investigate how this representation could be used to describe complex contagions such as opinion dynamics and percolation models where multiple edge occupations are required to include a node in the GCC [22]. We also intend to generalise the CSP model to include different kinds of coupling patterns between the percolation processes in the spirit of [15, 23, 24]. Finally, this method can be used for gaining insight into the training of neural network models for a multitude of deep learning applications which we also intend to explore.

* peter.mann@data-insights-ai.co.uk

[1] B. Karrer, M. E. J. Newman, and L. Zdeborová, "Percolation on sparse networks," *Phys. Rev. Lett.*, vol. 113, p. 208702, Nov 2014.

[2] B. Karrer and M. E. J. Newman, "Message passing approach for general epidemic models," *Phys. Rev. E*, vol. 82, p. 016101, Jul 2010.

[3] M. E. Newman, "Message passing methods on complex networks," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 479,

no. 2270, 2023.

[4] M. Mézard, G. Parisi, and R. Zecchina, "Analytic and algorithmic solution of random satisfiability problems," *Science*, vol. 297, p. 812–815, Aug 2002.

[5] S. N. Dorogovtsev and M. J. F. F., *The nature of Complex Networks*. Oxford University Press, 2022.

[6] M. E. Newman, *Networks*. Oxford University Press, 2019.

[7] M. E. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, Jul 2001.

[8] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, "Spectral redemption in clustering sparse networks," *Proceedings of the National Academy of Sciences*, vol. 110, no. 52, p. 20935–20940, 2013.

[9] T. Martin, X. Zhang, and M. E. J. Newman, "Localization and centrality in networks," *Phys. Rev. E*, vol. 90, p. 052808, Nov 2014.

[10] T. Hasegawa and S. Mizutaka, "Structure of percolating clusters in random clustered networks," *Physical Review E*, vol. 101, Jun 2020.

[11] M. E. J. Newman and C. R. Ferrario, "Interacting epidemics and coinfection on contact networks," *PLOS ONE*, 2013.

[12] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, "Two-pathogen model with competition on clustered networks," *Phys. Rev. E*, vol. 103, p. 062308, Jun 2021.

[13] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, "Cooperative coinfection dynamics on clustered networks," *Phys. Rev. E*, vol. 103, p. 042307, Apr 2021.

[14] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, "n-strain epidemic model using bond percolation," *Phys. Rev. E*, vol. 106, p. 014304, Jul 2022.

[15] P. Mann, V. A. Smith, J. B. Mitchell, and S. Dobson, "N-strain epidemic model using bond percolation," *Physical Review E*, vol. 106, Jul 2022.

[16] R. B. Potts, "Some generalized order-disorder transfor-mations," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, p. 106–109, Jan 1952.

[17] F. Y. Wu, "The Potts model," *Reviews of Modern Physics*, vol. 54, p. 235–268, Jan 1982.

[18] P. Mann and S. Dobson, "Belief propagation on networks with cliques and chordless cycles," *Phys. Rev. E*, vol. 107, p. 054303, May 2023.

[19] G. T. Cantwell and M. E. Newman, "Message passing on networks with loops," *Proceedings of the National Academy of Sciences*, vol. 116, no. 47, p. 23398–23403, 2019.

[20] F. Baccini, F. Geraci, and G. Bianconi, "Weighted simplicial complexes and their representation power of higher-order network data and topology," *Phys. Rev. E*, vol. 106, p. 034319, Sep 2022.

[21] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, "Degree correlations in graphs with clique clustering," *Phys. Rev. E*, vol. 105, p. 044314, Apr 2022.

[22] L. A. Keating, J. P. Gleeson, and D. J. P. O'Sullivan, "Multitype branching process method for modeling complex contagion on clustered networks," *Phys. Rev. E*, vol. 105, p. 034306, Mar 2022.

[23] S. Funk and V. A. Jansen, "Interacting epidemics on overlay networks," *Physical Review E*, vol. 81, Mar 2010.

[24] P. Mann, V. A. Smith, J. B. Mitchell, and S. Dobson, "Symbiotic and antagonistic disease dynamics on networks using bond percolation," *Physical Review E*, vol. 104, Aug 2021.