

Coverage versus supply cost in facility location: Physics of frustrated spin systemsChi Ho Yeung,^{1,2,3} K. Y. Michael Wong,² and Bo Li²¹*The Nonlinearity and Complexity Research Group, Aston University, Birmingham B4 7ET, United Kingdom*²*Department of Physics, The Hong Kong University of Science and Technology, Hong Kong*³*Department of Science and Environmental Studies, The Hong Kong Institute of Education, Tai Po, Hong Kong*

(Received 18 December 2013; published 10 June 2014)

A comprehensive coverage is crucial for communication, supply, and transportation networks, yet it is limited by the requirement of extensive infrastructure and heavy energy consumption. Here, we draw an analogy between spins in antiferromagnet and outlets in supply networks, and apply techniques from the studies of disordered systems to elucidate the effects of balancing the coverage and supply costs on the network behavior. A readily applicable, coverage optimization algorithm is derived. Simulation results show that magnetized and antiferromagnetic domains emerge and coexist to balance the need for coverage and energy saving. The scaling of parameters with system size agrees with the continuum approximation in two dimensions and the tree approximation in random graphs. Due to frustration caused by the competition between coverage and supply cost, a transition between easy and hard computation regimes is observed. We further suggest a local expansion approach to greatly simplify the message updates which shed light on simplifications in other problems.

DOI: [10.1103/PhysRevE.89.062805](https://doi.org/10.1103/PhysRevE.89.062805)

PACS number(s): 89.75.Hc, 02.50.-r, 05.20.-y, 89.20.-a

The effectiveness of communication and supply networks relies crucially on the extent of network coverage. For instance, networks of fire sensors, surveillance video cameras, and local weather monitors require a comprehensive coverage of a geographical region to guard safety and security or collect extensive information [1]. A comprehensive coverage is essential in transportation [2] and water distribution networks [3]. Likewise, the locations of supermarket branches, teller machines, or restaurant outlets determine market share, and hence business success [4,5]. Governments set up public facilities such as schools and clinics to cover their administrative regions so as to uphold the well being of the community [6].

On the other hand, it is costly to extend and maintain coverage. Sensor networks usually consume a lot of transmission power for broad area coverage, which is a problem for maintaining the life span of the sensor networks given the limited battery size of the sensors [7,8]. The logistics supply for networks of facilities requires abundant resources as well as extensive infrastructure. In military history, overstretching the supply lines increased the vulnerability of military operations, as commonly believed in the Russian campaigns of Napoleon and Hitler [9]. These create practical concerns on coverage expansion, and lead to a tradeoff between supply costs and service coverage.

To balance the need for coverage and the cost of maintaining supplies, efforts have been devoted in various disciplines. For instance, deployment of sensors in sensor networks has been intensely studied by engineers [10,11]; retail strategies and optimal branch allocation are studied in operations research [4,5]; airline destinations are frequently revised by airline management. However, many of the above problems remain heuristically tackled, which result in suboptimal solutions. In addition, centralized techniques such as linear programming are usually employed [12]. These work for problems with a global optimizer, but not for distributive applications such as the distribution of P2P file storage on the Internet [13].

This competition shares similarities with frustrated spin systems in statistical physics [14]. To see the correspondence, we note that local variables in a spin model are represented by

“spins” with binary states (but can be generalized to multiple states) and correspond to whether a location is selected for facility location or not. The propensity to increase coverage can be modeled by antiferromagnetic interactions. At the same time, we introduce a transportation cost from a warehouse to each selected node. This will result in a regional bias of selected nodes in the neighborhood of the warehouse, in competition with antiferromagnetism. Our formulation is also relevant to sensor networks where distributed sensors communicate with a base station. In the context of spin models, the transportation cost creates a long-range correlation between spins, a new feature absent in conventional antiferromagnet models.

In this paper, we elucidate the different patterns resultant from the interplay between these factors. We will find the coexistence of magnetized and antiferromagnetic or glassy domains. This multiplicity of possible states is a consequence of frustration, and leads to the appearance of many suboptimal solutions, and interferes severely with the search for the optimal solution. By applying techniques in the studies of disordered systems [15,16], we derive a readily applicable distributed algorithm capable of optimizing real instances. In addition, the comparison between square lattices and random regular graphs leads to important implications on their coverage efficiency and cost scalability.

Other than the above implications, we have extended the local expansion approach in [17] to intensely simplify the suggested message-passing procedures. This algorithm sheds light on similar simplifications in other physical and optimization problems. A simple example of convergence to metastable states also reveals the limitation of fundamental message-passing algorithms.

I. THE MODEL

Specifically, we consider a network of N nodes; each node $i = 1, \dots, N$ is connected to a set of k_i neighbors denoted by \mathcal{N}_i . We denote $s_i = 1, -1$ when node i is *active* or *idle*, respectively. An active node can be regarded as

an outlet, a service station, or a sensor which serves or monitors the surrounding area. In addition, each of the active nodes establishes an individual communication path to a *central terminal*, labeled T , where commodities are stored or information is processed. We denote the total flow on the edge from i to j as $I_{ij}(= -I_{ji})$, which is the sum of communication paths passing the edge (ij) . We then minimize the Hamiltonian H in the space of s_i and I_{ij} , given by

$$\mathcal{H} = J \sum_{(ij)} s_i s_j + U \sum_{i=1}^N \frac{1 - s_i}{2} + \sum_{(ij)} \phi(I_{ij}). \quad (1)$$

In the first term, a positive J encourages neighbors to be in opposite states. In supply networks, J may correspond to the *outlet redundancy*, namely, the redundant operational cost due to the duplication of service in neighboring nodes. It tends to spread the outlets and hence increase the coverage. The second term represents the loss in revenue due to nodes being uncovered. The *idle penalty* U may correspond to unsatisfied demand, such that a large U encourages outlet installation. The third term is the *supply cost*. Although our analytic and algorithmic results are generic to the form of $\phi(x)$, we assume $\phi(x) = x^2$ in this study. Considering situations that a single path is established between individual active nodes and the terminal, we restrict the current I_{ij} on all edges (ij) to be integers [18], i.e.,

$$I_{ij} \in \mathbb{Z}, \quad \forall (ij). \quad (2)$$

The variables are subject to the following constraint to conserve the flow of resources:

$$\sum_{j \in \mathcal{N}_i} I_{ji} = \frac{1 + s_i}{2}. \quad (3)$$

In this paper, we will mainly study the interplay between parameter J and U and the optimal coverage state. We remark that it is straightforward to generalize the model to accommodate heterogeneous onsite demand through multiplying the right-hand side of Eq. (3) by the demand of node i .

The above setup is analogous to spin models. The first and second terms of H correspond to an antiferromagnetic interaction and an external magnetic field, respectively. The last term is ubiquitous to communication networks [17]. If one fixes the active or inactive states of the nodes and optimizes the supply energy in the space of I_{ij} , one would obtain an effective long-range antiferromagnetic interaction between the spins since activating two nodes which share part of their paths to the terminal will increase the nonlinear total supply cost. The interplay of these three interactions would lead to new physical phenomena.

To further illustrate the analogy in this paper, we will consider the Hamiltonian on the square lattice and random regular graphs. Spin models with competing interactions on the square and other regular lattices have been studied in the axial next nearest neighbor Ising (ANNNI) model [19], and the next nearest neighbor Ising model (NNNI) [20]. When the ratio of the nearest to next nearest couplings (both antiferromagnetic) changes, there are phase changes in the periodicity of the modulated configuration, and may even lead to the absence of long range order at some special coupling ratio. At nonzero

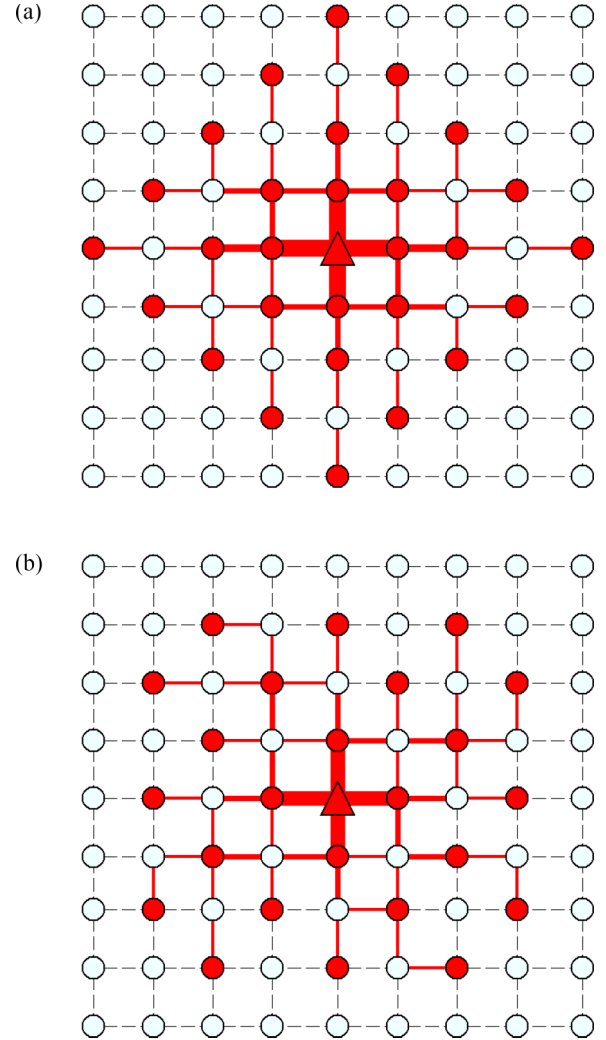


FIG. 1. (Color online) The optimized node configuration on a square lattice with 81 nodes, (a) $J = 0.3$ and $U = 20$, (b) $J = 5$ and $U = 0$. The triangle, red filled nodes, and empty nodes correspond to the terminal, active, and idle nodes, respectively. Dashed edges are idle, and the thickness of the solid edges is proportional to $|I|$, the flow on individual edges.

temperature, further modulation patterns appear, and there are metastable states [21]. The periodicity of the modulations has infinite steps when the coupling ratio changes, exhibiting the so-called devil's staircase behavior [22]. All these point to the existence of many metastable states produced by frustration. Progress on the analysis of many ANNNI and NNNI models was made by approximating the local structures to be treelike [23]. They reproduce many thermodynamic features of the finite dimensional NNNI models on random graphs including the special ratios [24,25].

The new observation in our model is the coexistence of active, idle, and antiferromagnetic domains in a single network. We show in Fig. 1(a) an example of the ground state of H on a square lattice. One can identify that (i) the emergence of a core of active nodes is due to the presence of an external field, (ii) the alternating active-idle outer layer is due to the antiferromagnetic interaction, and (iii) the four idle corner regions are due to the presence of supply cost.

We note that there are some fundamental differences between the present model and the conventional facility location problem, such that algorithms of the latter can not be readily applied to solve the problem. Specifically, the presence of the antiferromagnetic term to encourage the spread of facilities has significantly increased the extent of frustration in the system. Besides, the quadratic transportation cost which discourages active nodes from sharing their path is a major feature of the present formulation. Among the various versions of conventional facility location problems, the version comparable to our model would have the objective of minimizing the population deprived of facilities [26–28]. However, these problems paid little attention to whether the deprived population is clustered topographically, or to the cost incurred by congested transportation paths. These differences render the application of existing algorithms on the present model difficult and make the algorithm derived in Sec. III essential. Nevertheless, we have compared the performance of our algorithm with the IBM CPLEX solver [29] and the results are discussed in Sec. III C.

II. CAVITY EQUATIONS

Based on the treelike picture, we will obtain the optimal state of the present model by employing the cavity approach developed in the field of spin glasses and disordered systems [15,16]. The cavity approach assumes that only large loops exist in a network such that neighbors of a node become independent. In regular lattices, the cavity approach is effectively only approximate. Nevertheless, we found that the derived algorithm yields accurate results in networks with short loops, including regular lattices. We will see in Sec. IV A that the algorithmic solution is identical to the known optimal solution in simple cases. On random regular graphs, the absence of short loops renders the application of cavity approach more justified. In Sec. IV B, we will find more evidence of frustration, and a transition from replica symmetric to replica symmetry-breaking phase.

To obtain various macroscopic properties, we first denote $E_{i \rightarrow l}(s_i, I_{il})$ as the optimized energy of the tree terminated at the edge from node i to l , where the state of node i and the current to node l are s_i and I_{il} , respectively. This energy includes the supply cost $\phi(I_{ij})$ but excludes the coupling term $J s_i s_j$. One can write a recursive relation relating $E_{i \rightarrow l}(s_i, I_{il})$ to the energy $E_{j \rightarrow i}(s_j, I_{ji})$ from neighbor j to node i , i.e.,

$$E_{i \rightarrow l}(s_i, I_{il}) = \phi(I_{il}) + \min_{\{(s_j, I_{ji}) | f_i = I_{il}\}} \left[\sum_{j \in \mathcal{N}_i \setminus l} J s_i s_j + U \frac{1 - s_i}{2} + \sum_{j \in \mathcal{N}_i \setminus l} E_{j \rightarrow i}(s_j, I_{ji}) \right], \quad (4)$$

where $f_i = \sum_{j \in \mathcal{N}_i \setminus l} I_{ji} - \frac{1+s_i}{2}$. We have assumed that the minimization on the right-hand side of Eq. (4) does not depend on the states s_k of the next nearest neighbors $k \in \mathcal{N}_j \setminus i$. As we shall see, it turns out that the assumption yields good results. For the terminal node T , we assume $s_T = -1$ since it does not serve as an active node. The optimized energy function

$E_{T \rightarrow l}(s_T, I_{Tl})$ from the terminal to its neighbor l is given by

$$E_{T \rightarrow l}(s_T, I_{Tl}) = \phi(I_{Tl}) + C(1 + s_T) \quad (5)$$

for some arbitrary large constant $C > 0$ such that $s_T = -1$ always has a lower energy compared to $s_T = 1$.

Since $E_{i \rightarrow l}(s_i, I_{il})$ is an extensive quantity, we define its intensive counterpart by $E_{i \rightarrow j}^V(s_i, I_{il}) = E_{i \rightarrow l}(s_i, I_{il}) - \min_{\{s'_i, I'_{il}\}} [E_{i \rightarrow l}(s'_i, I'_{il})]$, and further denote the right-hand side of Eq. (4) as

$$\mathcal{M}[s, I; \{E_{j \rightarrow i}\}_{j \in \mathcal{S}}] \equiv \phi(I) + \min_{\{(s_j, I_{ji}) | f_i = I\}} \left[\sum_{j \in \mathcal{S}} J s_i s_j + U \frac{1 - s}{2} + \sum_{j \in \mathcal{S}} E_{j \rightarrow i}(s_j, I_{ji}) \right], \quad (6)$$

where \mathcal{S} denotes a set of nodes. The recursive equation of $E_{i \rightarrow l}^V(s_i, I_{il})$ in Eq. (4) can be written as

$$E_{i \rightarrow l}^V(s_i, I_{il}) = \mathcal{M}[s_i, I_{il}; \{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i \setminus l}] - \min_{s, I} \{ \mathcal{M}[s, I; \{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i \setminus l}] \}. \quad (7)$$

One can obtain a stable functional distribution $P[E^V(s, I)]$ of $E^V(s, I)$ by the following self-consistent equation:

$$P[E^V(s, I)] = \sum_{k=1}^{\infty} \frac{k\rho(k)}{\langle k \rangle} \int \prod_{\substack{j=1 \\ s_j, I_j}}^k dE_j^V(s_j, I_j) P[E_j^V(s_j, I_j)] \times \prod_{s, I} \delta(E^V(s, I) - \mathcal{M}[s, I; \{E_j^V\}_{j=1, \dots, k}]) + \min_{s', I'} \{ \mathcal{M}[s', I'; \{E_j^V\}_{j=1, \dots, k}] \}, \quad (8)$$

where $k\rho(k)/\langle k \rangle$ is the so-called *excess degree distribution*, which describes the probability of finding an edge connected to a node of degree k . The above equation can be solved, for instance, by population dynamics. By averaging over $P[E^V(s, I)]$, the optimized energy can be computed by $\langle E \rangle = (\langle E_{\text{node}} \rangle - \frac{\langle k \rangle}{2} \langle E_{\text{link}} \rangle)$ following the conventional cavity approach [30], where $\langle E_{\text{node}} \rangle$ and $\langle E_{\text{link}} \rangle$ are given by

$$\langle E_{\text{node}} \rangle = \langle \min_s \{ \mathcal{M}[s, 0; \{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i}] \} \rangle_{\{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i}}, \quad (9)$$

$$\langle E_{\text{link}} \rangle = \langle \min_{s_1, s_2, I} \{ E_{j_1}^V(s_1, I) + E_{j_2}^V(s_2, -I) - \phi(I) + J s_1 s_2 \} \rangle_{E_{j_1}^V, E_{j_2}^V}. \quad (10)$$

III. ALGORITHMS AND LOCAL EXPANSION OF MESSAGES

The recursion of $E_{j \rightarrow i}^V$ in Eq. (7) gives rise to a message-passing algorithm. To optimize the total cost, one randomly chooses a node i and gathers the functional messages $E_{j \rightarrow i}^V$ from all neighbors j of i . We then randomly choose one of the neighbors to be the ancestor l , and update the message $E_{i \rightarrow l}^V$ according to Eq. (7). The procedure continues until the messages converge on all edges in both directions, i.e., E^V does not change in consecutive updates for a sufficiently large number of steps. Using equations resembling Eq. (10), one can

compute the optimal flow I_{il}^* on the edge from i to l by

$$I_{il}^* = \operatorname{argmin}_{s_i, s_l, I} \{ E_{i \rightarrow l}^V(s_i, I) + E_{l \rightarrow i}^V(s_l, -I) - \phi(I) + J s_i s_l \}. \quad (11)$$

The optimal state s_i^* of a node i is given by

$$s_i^* = \operatorname{argmin}_s \{ \mathcal{M}[s, 0; \{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i}] \}. \quad (12)$$

Alternatively, one can also compute s_i^* using Eq. (3) and $\{I_{ji}^*\}$ obtained from Eq. (11), such that

$$s_i^* = -1 + 2 \sum_{j \in \mathcal{N}_i} I_{ji}^*. \quad (13)$$

Once the optimal flow on all the edges is obtained, the communication path from individual active nodes to the terminal can be identified by tracing from the active node a path to the terminal through edges of nonzero flow un-occupied by other communications.

Although the algorithm is easily formulated by Eqs. (7), (12), and (11), it is difficult to compute since (i) the algorithm involves functional messages with indefinite domain, and (ii) the recursion Eq. (7) involves constrained optimization on integer domains.

A. Local expansion of messages

To cope with these difficulties, we employ an approach introduced in [17] to simplify the message-passing algorithm. Since the derivation is rather involved, we refer readers who would like to skip the derivation to the summary of the algorithm below Eq. (20).

To simplify the present message-passing algorithm, we first note that in [17], an optimization algorithm is derived for a resource allocation problem involving functional messages. Unlike the present case, the messages in [17] have a continuous domain, which renders their exact computation infeasible except for some specific form of cost function. Nevertheless, instead of a complete functional message, only the values on a small domain are necessary to verify the optimality of the converged state. This allows Wong and Saad [17] to parametrize the functional messages by only the first two coefficients in its Taylor's series expansion around a *working point*, which is given by the ancestor. The working points are continuously updated by the ancestors until they coincide with the optimal state. Here, we employ a similar idea of local expansion to effectively parametrize the functional messages on integer domains. The same approach has been adopted in routing optimization in [31]. To locally expand the messages, we define

$$h_{i \rightarrow l}^{(\sigma)} = E_{i \rightarrow l}^V(1, I_{il}^* + \sigma) - E_{i \rightarrow l}^V(-1, I_{il}^*), \quad (14)$$

$$g_{i \rightarrow l}^{(\sigma)} = E_{i \rightarrow l}^V(-1, I_{il}^* + \sigma) - E_{i \rightarrow l}^V(-1, I_{il}^*), \quad (15)$$

where I_{il}^* is the working point of flow given by node l for node i , and $-M \leq \sigma \leq M$ is an integer. The parameter M determines the range of flows around the working point to be parametrized, the smaller the the value of M , the fewer variables used to parametrize the functional messages $E_{i \rightarrow l}^V(s, I)$. When $M \rightarrow \infty$, the complete profile of $E_{i \rightarrow l}^V(s, I)$ is recovered. We

observed that the algorithmic performance does not improve beyond $M = 2$, such that $M = 2$ is already sufficient for the systems studied. We have chosen $E_{i \rightarrow l}^V(-1, I_{il}^*)$ to be a common reference point in both Eqs. (14) and (15), while other choices would yield identical physical results. In this case, $h_{i \rightarrow l}^{(0)} = E_{i \rightarrow l}^V(1, I_{il}^*) - E_{i \rightarrow l}^V(-1, I_{il}^*)$ and $g_{i \rightarrow l}^{(0)} = 0$. The $h_{i \rightarrow l}$ and $g_{i \rightarrow l}$ variables can be regarded as the cavity fields at zero temperature.

We then continue to derive a recursion relation involving only h and g . We first denote the vectors $(h_{i \rightarrow l}^{(-M)}, h_{i \rightarrow l}^{(-M+1)}, \dots, h_{i \rightarrow l}^{(M)})$ and $(g_{i \rightarrow l}^{(-M)}, g_{i \rightarrow l}^{(-M+1)}, \dots, g_{i \rightarrow l}^{(M)})$ by $\vec{h}_{i \rightarrow l}$ and $\vec{g}_{i \rightarrow l}$, respectively. From Eqs. (7) and (14), one can express $h_{i \rightarrow l}^{(\sigma)}$ as

$$\begin{aligned} h_{i \rightarrow l}^{(\sigma)} &= \mathcal{M}[1, I_{il}^* + \sigma; \{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i \setminus l}] \\ &\quad - \mathcal{M}[-1, I_{il}^*; \{E_{j \rightarrow i}^V\}_{j \in \mathcal{N}_i \setminus l}] \\ &\approx \phi(I_{il}^* + \sigma) - \phi(I_{il}^*) + \mathcal{Q}[1, \sigma, I_{il}^*; \hat{h}_{i \setminus l}, \hat{g}_{i \setminus l}, \hat{I}_{i \setminus l}^*] \\ &\quad - \mathcal{Q}[-1, 0, I_{il}^*; \hat{h}_{i \setminus l}, \hat{g}_{i \setminus l}, \hat{I}_{i \setminus l}^*], \end{aligned} \quad (16)$$

where $\hat{h}_{i \setminus l} = \{\vec{h}_{j \rightarrow i} | j \in \mathcal{N}_i \setminus l\}$ and $\hat{g}_{i \setminus l} = \{\vec{g}_{j \rightarrow i} | j \in \mathcal{N}_i \setminus l\}$. The set $\hat{I}_{i \setminus l}^* = \{I_{ji}^* | j \in \mathcal{N}_i \setminus l\}$ contains the working points where the messages $\hat{h}_{i \setminus l}$ and $\hat{g}_{i \setminus l}$ were computed. The functional \mathcal{Q} is defined as

$$\begin{aligned} \mathcal{Q}[s, \sigma, I_{il}^*; \hat{h}_{i \setminus l}, \hat{g}_{i \setminus l}, \hat{I}_{i \setminus l}^*] &= \min_{\{\{s_j, \sigma_j\} | f_i = I_{il}^* + \sigma\}} \left[\sum_{j \in \mathcal{N}_i \setminus l} J s s_j + U \frac{1-s}{2} \right. \\ &\quad \left. + \sum_{j \in \mathcal{N}_i \setminus l} (h_{j \rightarrow i}^{(\sigma_j)} \delta_{s_j, 1} + g_{j \rightarrow i}^{(\sigma_j)} \delta_{s_j, -1}) \right], \end{aligned} \quad (17)$$

where $f_i = \sum_{j \in \mathcal{N}_i \setminus l} (I_{ji}^* + \sigma_j) - \frac{1+s}{2}$. The approximation sign in Eq. (16) comes from the restricted range of $-M \leq \sigma_j \leq M$ when M is finite, such that the minimization in \mathcal{Q} in Eq. (17) is only an approximation of \mathcal{M} in Eq. (6). In other words, Eq. (16) becomes exact only when $M \rightarrow \infty$. Similarly, from Eqs. (7) and (15), one can express $g_{i \rightarrow l}^{(\sigma)}$ as

$$\begin{aligned} g_{i \rightarrow l}^{(\sigma)} &\approx \phi(I_{il}^* + \sigma) - \phi(I_{il}^*) + \mathcal{Q}[-1, \sigma, I_{il}^*; \hat{h}_{i \setminus l}, \hat{g}_{i \setminus l}, \hat{I}_{i \setminus l}^*] \\ &\quad - \mathcal{Q}[-1, 0, I_{il}^*; \hat{h}_{i \setminus l}, \hat{g}_{i \setminus l}, \hat{I}_{i \setminus l}^*]. \end{aligned} \quad (18)$$

In addition to the update of messages $\vec{h}_{i \rightarrow l}$ and $\vec{g}_{i \rightarrow l}$ from node i to ancestor l , node i has to update the working points for its descendants j at which new messages $\vec{h}_{j \rightarrow i}$ and $\vec{g}_{j \rightarrow i}$ will be computed in the next round. These working points are updated by

$$I_{ji}^* + \sigma_j^* \rightarrow I_{ji}^*, \quad (19)$$

where

$$\begin{aligned} \sigma_j^* &= \operatorname{argmin}_{\{\{s_j, \sigma_j\} | f_i = I_{il}^*\}} \left[\sum_{j \in \mathcal{N}_i \setminus l} J s_i s_j + U \frac{1-s_i}{2} \right. \\ &\quad \left. + \sum_{j \in \mathcal{N}_i \setminus l} (h_{j \rightarrow i}^{(\sigma_j)} \delta_{s_j, 1} + g_{j \rightarrow i}^{(\sigma_j)} \delta_{s_j, -1}) \right]. \end{aligned} \quad (20)$$

In summary, the algorithm proceeds as follows:

(1) Randomly draw a node i and choose at random one of its neighbors to be the ancestor l .

(2) Obtain the updated working point I_{il}^* from node l and gather $\vec{h}_{j \rightarrow i}$, $\vec{g}_{j \rightarrow i}$ and the corresponding working point I_{ji}^* at which $\vec{h}_{j \rightarrow i}$ and $\vec{g}_{j \rightarrow i}$ were computed from all the neighbors of i other than l .

(3) Compute $\vec{h}_{i \rightarrow l}$ and $\vec{g}_{i \rightarrow l}$ by Eqs. (16) and (18), respectively. Update the working points I_{ji}^* by Eqs. (19) and (20) for all neighbors j except l .

(4) Repeat steps 1 to 3 until all messages and working points converge, i.e., do not change for a sufficiently long time.

(5) The optimal flows on the edges are given by the converged working points. The optimal state of node i is given by Eq. (13). The communication path between individual active nodes and the terminal is identified by tracing from each node i through edges with nonzero flow unoccupied by the communication paths of other active nodes.

B. Random biases

As suggested by previous models of communication networks [18,31,32], the presence of degenerate optimized states may hinder algorithm convergence. One can adopt the same approach to break degeneracy by assigning to each edge (ij) a small random bias $\epsilon_{(ij)}$ and optimize a cost $\phi(I_{ij}) + |I_{ij}|\epsilon_{(ij)}$, which helps the algorithm converge to one of the optimized states.

In the parameter regimes where the algorithm fails to converge even with the set of small $\epsilon_{(ij)}$ implemented, we found that the following procedures may improve convergence:

(1) Start with a set of large $\epsilon_{(ij)}$ such that the algorithm converges. Obtain a converged state.

(2) Decrease the bias by $\epsilon_{(ij)} \rightarrow f\epsilon_{(ij)}$ for some fraction $f < 1$.

(3) Use the previous converged state as the initial condition to obtain a new convergent state with the new set of biases.

(4) Repeat steps 1–3 to decrease the value of $\epsilon_{(ij)}$ until the algorithm fails to converge or $\epsilon_{(ij)}$ is smaller than a threshold. The converged solution obtained by the set of smallest $\epsilon_{(ij)}$ is considered as the optimized configuration.

By repeating steps 1 to 3, we found that the energy of the converged state is decreasing with decreasing $\epsilon_{(ij)}$. In regimes where the algorithm converges without the above procedures, in most cases we obtain an identical algorithmic solution with the above process. We thus expect that an optimal or close-to-optimal state is obtained in the regime where the algorithm ceases to converge with small ϵ_{ij} . Physically, these procedures are equivalent to the case when one applies a strong external field to facilitate the convergence to an anticipated state, and slowly decrease the field such that the system is attracted to one of the local minima nearby.

C. Comparison with CPLEX

We have compared the performance of our algorithm with the IBM CPLEX solver [29], which is capable of solving integer and quadratic programming problems. We found that in parameter regimes with intermediate idle penalties and

nonzero antiferromagnetic couplings, the computational complexity of the solver scales up rapidly, or even exponentially increases with system size. It leads to a long running time even for relatively small size of $O(10^2)$. On the other hand, our algorithm does not converge on all instances in the same regime, but it does show a quadratic complexity with system size by averaging over the converged realizations. For those nonconverged instances which can not be solved by the approach described in the previous subsection, techniques such as decimation can be applied to find a solution [30]. In general, our algorithm compares favorably with integer quadratic programming techniques in terms of computational complexity.

IV. RESULTS

A. Square lattice

We first examine the optimized location of active nodes on a square lattice where the terminal is located at the center. We note that the algorithm derived in Sec. III is based on the cavity approach which assumes the absence of short loops, and is only a crude approximation on square lattices. Nevertheless, the algorithm converges in a majority of the parameter regime. For regimes where the algorithm ceases to converge, we found that the procedures suggested in Sec. III B facilitate the algorithmic convergence.

To verify the algorithmic results on square lattices, we examine a simple system of size $N = 25$ with $J = 0$. As shown in Figs. 2(a)–2(f), a core of active nodes grows around the center when U increases. In this case, one can analytically compute the number of active nodes N_a as a function of U . For instance, when U increases gradually from $U = 0$, we expect the system to change at a threshold value from the complete idle state in Fig. 2(a) to the state where the four neighbors of the terminal become active, as shown in Fig. 2(b). To compute the threshold value, we consider the change in energy of the configuration from Figs. 2(a) to 2(b) given by $[(N - 5)U + 4] - (N - 1)U$, a negative energy change at $U = 1$ signals the stability of Fig. 2(b) over 2(a) such that N_a increases from 0 to 4. We remark that there is no intermediate optimized configuration with $1 \leq N_a \leq 3$ since the system has a fourfold symmetry; if an active node is energetically favorable in one direction, then other active nodes are also favorable in the other three directions.

Similarly, one can obtain other threshold values of U by considering the energetic stability of consecutive configurations in Fig. 2. Degenerate states, i.e., states with the same optimized energy but different configuration, do exist for Figs. 2(c) and 2(e). Nevertheless, their existence has no influence on the threshold values.

In Fig. 3, we show the excellent agreement between the results from the message-passing algorithm and the above analytical N_a . The agreement suggests that, although the cavity approach is an approximation on square lattices, the derived algorithm leads to the optimal state at $J = 0$. We thus expect that an optimal state, or at worst a close-to-optimal state, is obtained by the algorithm when it converges. Other than the square lattice of size $N = 25$, the analytical N_a also agrees well with algorithmic results in larger systems, suggesting

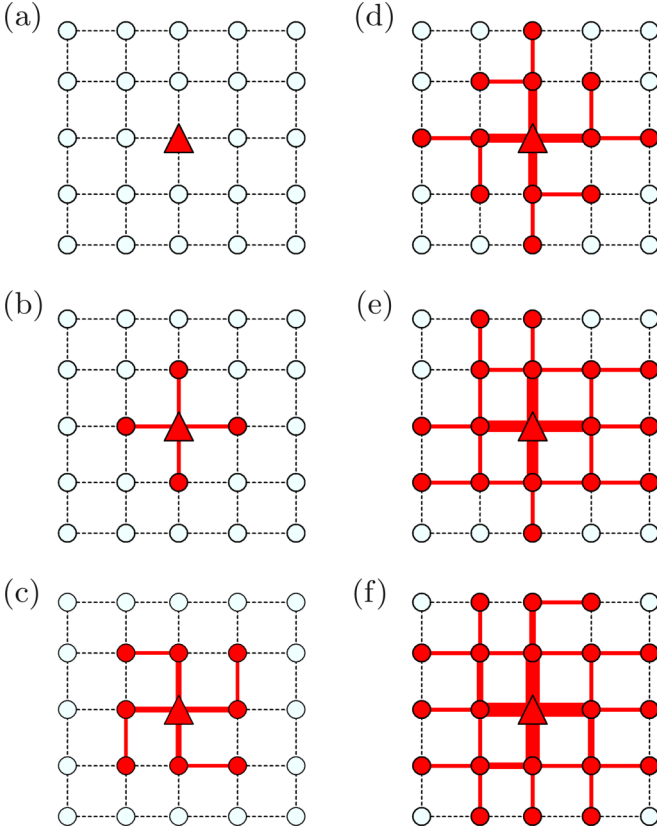


FIG. 2. (Color online) The optimized node configuration on a square lattice of $N = 25$ with $J = 0$ and (a) $U = 0.5$, (b) $U = 2$, (c) $U = 5$, (d) $U = 7$, (e) $U = 10$, and (f) $U = 14$. Other configurations with energy identical to (c) and (e) exist.

identical active node configurations in the range $0 < U < 15$ for systems of all sizes. Differences between $N = 25$ and larger lattices emerge beyond $U = 15$ since larger systems can attain a lower energy state with a diamond-shaped active core instead of an active square with size $N = 25$. Similar

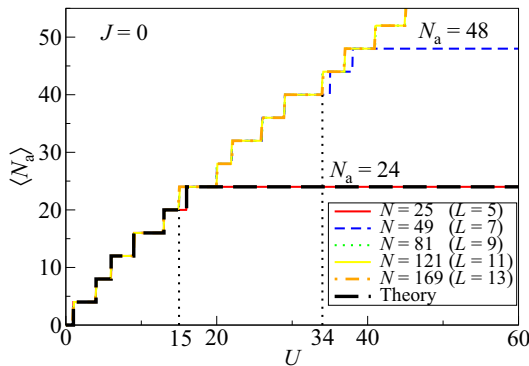


FIG. 3. (Color online) The number of active nodes $\langle N_a \rangle$ as a function of idle penalty U for $N = 25, 49, 81, 121$, and 169 , obtained by the algorithm derived by the cavity approach. Results are averaged over 10 realizations with different sets of random biases. The analytical results for $N = 25$ are obtained by identifying the active node configuration with lowest energy (shown in Fig. 2) at specific values of U .

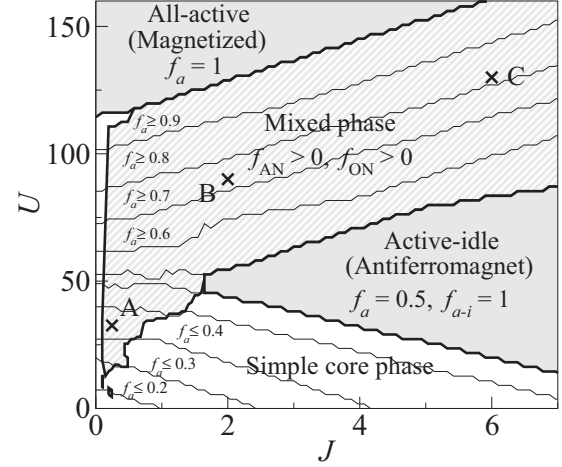


FIG. 4. Phase diagram and the fraction of active nodes as a function of J and U on a square lattice with $N = 121$. The optimal configurations obtained at points A, B, and C are shown in Figs. 5(a)–5(c), respectively.

differences are observed between $N = 49$ and larger lattices beyond $U = 34$. These differences are due to boundary effects of the smaller lattices.

1. Various phases

We show in Fig. 4 the contour plot of the optimized fraction of active nodes as a function of the parameter J and U on square lattices with $N = 121$ nodes. Although these results are obtained on a particular system size only, we will show in the next subsection that the picture is generic to different system sizes by a proper rescaling.

We first discuss the case of $J = 0$, such that the antiferromagnetic interaction in Eq. (1) is absent and the alternating active-idle state is not encouraged. In this case, the fraction of active nodes f_a increases with idle penalty U and becomes saturated at $f_a = 1$ when U is larger than a certain value. It implies that more nodes become active when the idle penalty U increases. This is similar to spin models where the magnetization increases with the magnitude of external field. All the active nodes are found close to the terminal to minimize the supply cost. In other words, there is an active core centered at the terminal which expands with U until the whole network is active. In the context of supply networks, it implies that the optimal locations of outlets are close to the warehouse when demand is low.

Next, we discuss the case of nonzero antiferromagnetic coupling J with $U = 0$. As we can see from Fig. 4, the fraction of active nodes increases with J and becomes saturated at $f_a = 0.5$. It implies that more nodes become active as J increases, but unlike the previous case where there exists an active core, a region of alternating active-idle nodes emerges around the terminal as shown in Fig. 1(b). When J is larger than a certain threshold, the alternating active-idle configuration spans the whole lattice and half of nodes become active; f_a saturates at 0.5. This corresponds to the case where outlets cover the whole network but are installed in every other location.

In the cases with nonzero U and J , we note an *all-active phase* with $f_a = 1$ when U is large, corresponding to a

comprehensive coverage as well as a magnetized phase in a spin system where all spins align with the magnetic field. On the other hand, when J is large we expect an *active-idle phase* emerges where the alternating active-idle pattern spans the whole network, similar to an antiferromagnet. The region with $f_a = 0.5$ does not necessarily identify the active-idle phase since configurations other than the antiferromagnetic pattern may also have $N/2$ active nodes, for instance, a core with $N/2$ active nodes centered at the terminal. We thus measure f_{a-i} , the fraction of edges which connects an active and an idle node, and identify the antiferromagnetic phase to be the region with $f_{a-i} = 1$. As we can see in Fig. 4, the active-idle phase emerges as a triangular regime at large J .

To reveal the configurations outside the magnetized and antiferromagnetic regimes, we measure f_{AN} and f_{ON} , the fraction of active nodes with at least one active neighbor and the fraction of nodes where all neighbors are in the opposite state. The four nodes at the corners are not counted in f_{ON} since they may satisfy its definition even when the antiferromagnetic interaction is absent, for instance, the four idle nodes at the corners of the configuration in Fig. 2(f). We can then identify the *mixed phase* as the regime with nonzero f_{AN} and f_{ON} . In the mixed phase, we generally observe an active core centered at the terminal, surrounded by an area of alternating active-idle nodes, and possibly followed by an idle area depending on the value of U and J . For instance, such a configuration is observed in the optimized states in Figs. 5(a) and 5(b) which are obtained by U and J at point A and B in the mixed phase (see Fig. 4). Other configurations emerge at larger U and J , such as the state in Fig. 5(c) which shows an active-idle ring surrounded by a fully active boundary. The reason for such a configuration is the weaker antiferromagnetic interaction due to fewer neighbors, i.e., less redundancy, on the boundary compared to the interior nodes, which results in a preference of an active boundary over an active-idle chain along it. We remark that this configuration is suitable for surveillance network where cameras are mainly installed more on the boundary than in the interior.

Finally, the unshaded regime outside the magnetized, antiferromagnetic, and the mixed phase in Fig. 4 is characterized by $f_a < 1$ and $f_{a-i} < 1$, as well as either $f_{AN} = 0$ or $f_{ON} = 0$, indicating the presence of either a magnetized or antiferromagnetic core which does not span the whole network. We thus call this regime the *simple core phase*. As we have discussed, the cases of $J = 0$ and $U = 0$ are characterized by an active core and an active-idle core, respectively, which suggests that the vertical unshaded regime near $J = 0$ and to the left of the mixed phase is contributed by configurations with an active core, while the unshaded regime to the right of the mixed phase is contributed by an active-idle core.

With the extensive regions of the all-active and active-idle states in Fig. 4, it may be misleading to conclude that optimal solutions in the mixed phase which have more complex configurations do not result in significant gains over all-active or active-idle state. Figure 6 shows that the energy of the optimal state is significantly lower than the all-active and active-idle states where f_a lies between 0.5 and 1. From the application point of view, the mixed phase is the most interesting regime since the optimal state consists of both the all-active and the active-idle regions, but the determination of their boundary is a nontrivial task.

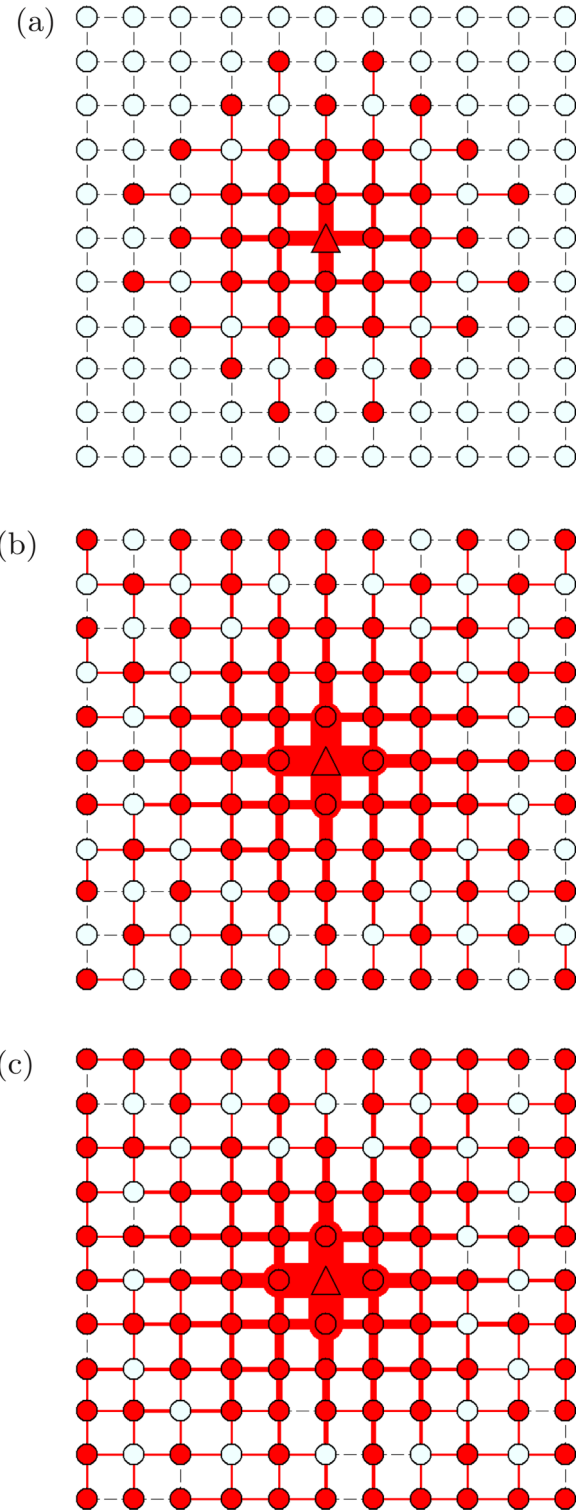


FIG. 5. (Color online) The optimized node configuration in the mixed phase on a square lattice of $N = 121$ with (a) $J = 0.245$ and $U = 32.6$, (b) $J = 2$, $U = 90$, (c) $J = 6$ and $U = 130$.

2. Scaling with system size

We further investigate the scaling behaviors of the system as a function of size N . We will first identify for the cases of $U = 0$ and $J = 0$ a scaling factor for U and J which results in

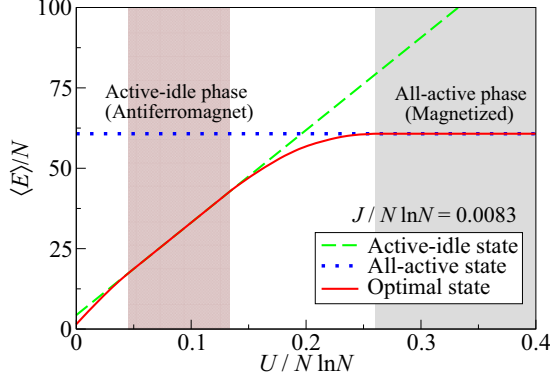


FIG. 6. (Color online) The dependence of the energy on U of the all-active, the active-idle, and the optimal states of a square lattice with $N = 121$ and $J/N \ln N = 0.0083$.

a data collapse of f_a for different N . We will then examine the scaling in the mixed phase when both U and J are nonzero.

To get the correct scaling factor of U when $J = 0$, we adopt the continuum approximation to compute the supply cost. We denote the distance of the furthest active node from the terminal as d_{\max} . Noting that the dominant contribution to the supply cost comes from the region near the center, we can assume that the flows are isotropic and equal to $f_a N / (2\pi r)$, so that the supply cost $\sum_{ij} I_{ij}^2$ can be computed to the leading order as

$$\sum_{(ij)} I_{ij}^2 \approx \int_{\epsilon}^{d_{\max}} \left(\frac{f_a N}{2\pi r} \right)^2 2\pi r dr = \frac{f_a^2 N^2}{2\pi} \ln \frac{d_{\max}}{\epsilon}. \quad (21)$$

The cutoff distance ϵ is of the order 1 and can be approximated as 1 to the leading order. Using $\pi d_{\max}^2 = f_a N$, we obtain

$$E = (1 - f_a)NU + \frac{f_a^2 N^2}{4\pi} \ln(f_a N) - \frac{f_a^2 N^2}{4\pi} \ln \pi, \quad (22)$$

which gives rise, to the leading order,

$$f_a = 2\pi \frac{U}{N \ln N}. \quad (23)$$

This shows that the energy scale for achieving extensive coverage is $U \sim N \ln N$ when $J = 0$. As shown in Fig. 7(a), for individual system size N , the fraction of active nodes f_a increases with U , characterized by functions in the form of staircase due to the discreteness of the lattice. The larger the system, the smaller the step size and the more the steps until f_a saturates at 1. Due to the different step sizes, f_a as a function of $U/(N \ln N)$ for various N do not overlap perfectly on one another, but interestingly there exhibits a data collapse with individual staircases collapsing on a common trend. We remark that the unscaled counterpart of Fig. 7(a), i.e., the number of active nodes N_a as a function of unscaled U , also collapses well for various N as shown in Fig. 3. Nevertheless, the optimized f_a in Fig. 7(a) does not match well with Eq. (23) when f_a approaches 1. This is due to corrections to the supply cost of the order N^2 , which is significant compared with the leading order of $N^2 \ln N$ for the system sizes we simulated. These corrections are worked out in Appendix A and the result of Eq. (A7) is plotted in Fig. 7(a), producing a much improved agreement with the simulation results.

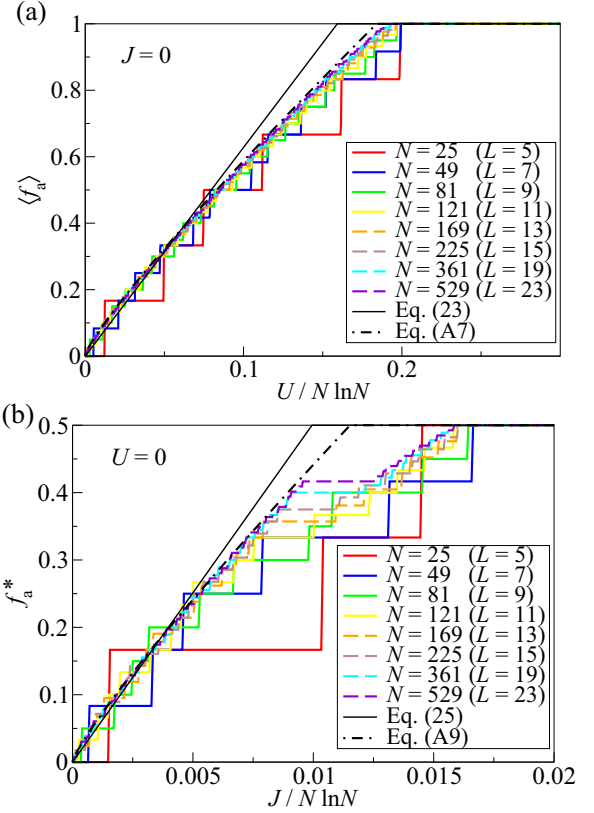


FIG. 7. (Color online) (a) The fraction of active nodes $\langle f_a \rangle$ as a function of $U/N \ln N$ with $J = 0$ and various N from $N = 25$ to 529 . Results are averaged over 10 realizations with different sets of random biases. Analytical approximations (23) and (A7) with $N = 529$ are shown. (b) The fraction of active nodes f_a^* as a function of $J/N \ln N$ with $U = 0$. Results are obtained in the minimum energy instance in 10 different instances. Analytical approximations (25) and (A9) with $N = 529$ are shown.

Next, we examine the case of increasing J with $U = 0$. In this case, we expect a core of alternating active-idle nodes emerges and expands with J , similar to Fig. 2 except that the core is replaced by an alternating active-idle configuration. As shown in Fig. 7(b), f_a with different N collapse well when plotted as a function of the rescaled $J/(N \ln N)$, similar to the case of increasing U with $J = 0$ in Fig. 7(a). The reason for showing f_a^* , the fraction of active nodes obtained in the instance with lowest energy in 10 different realizations, will be explained in the next subsection. These results suggest a continuum approximation similar to Eqs. (22) and (23) is valid even with the alternating active-idle node configuration. Hence, we model the configuration by a circular core of active-idle configuration with radius R , and all nodes beyond are idle. The total energy becomes

$$E = \frac{f_a^2 N^2}{4\pi} \left(\ln(f_a N) - \ln \frac{\pi}{2} + \pi - \frac{3}{2} \right) + 2JN(1 - 4f_a), \quad (24)$$

yielding, to the leading order,

$$f_a = 16\pi \frac{J}{N \ln N}. \quad (25)$$

Alternatively, one can substitute $U = 8J$ in Eq. (23) to obtain Eq. (25) since the energy gain by activating one node is generally $8J$ in the case with $U = 0$. As shown in Fig. 7(b), there is an excellent agreement between the approximation and the simulation results. On the other hand, plateaus of f_a are observed at intermediate values of J . It is indeed an interesting phenomenon related to the lattice boundary; since boundary nodes have fewer neighbors, their benefit from satisfying the antiferromagnetic interaction is smaller, and a larger value of J is required to activate them. This enhances the stability for configurations where the alternating core spans the whole network except the boundary [see Fig. 1(b) for instance], which leads to a plateau in f_a until J is sufficiently large to activate boundary nodes.

To examine the scaling of $N \ln N$ when both U and J are nonzero, we compare the microscopic features of systems with different size N but the same value of $U/(N \ln N)$ and $J/(N \ln N)$. For a quantitative comparison, we measure $f_{a-a}(d)$, $f_{a-i}(d)$, and $f_{i-i}(d)$, respectively, the fraction of edges which connect two active nodes, one active and idle node, and two idle nodes at a Manhattan distance d from the terminal. We obtain these results for systems of different N but at the same values of $U/(N \ln N)$ and $J/(N \ln N)$ indicated by point A in the mixed phase of Fig. 4; its corresponding optimized configuration on the $N = 121$ lattice is shown in Fig. 5(a).

As we can see in Fig. 8, the results of $f_{a-a}(d)$, $f_{a-i}(d)$, and $f_{i-i}(d)$ as a function of $d/(\sqrt{N} - 1)$ for different system sizes collapse well. The results indicate the presence of an all-active core in the neighborhood of the terminal, surrounded by an active-idle region, which is in turn surrounded by a region of idle nodes near the boundary in the various systems examined. It suggests that systems of different sizes behave similarly at the same values of $U/(N \ln N)$ and $J/(N \ln N)$, and the scaling factor $1/(N \ln N)$ is appropriate when both U and J are nonzero. The data collapse shown in Figs. 7 and 8 suggests that phase diagrams similar to Fig. 4 can be identified for other system sizes related by the scaling factor $1/(N \ln N)$.

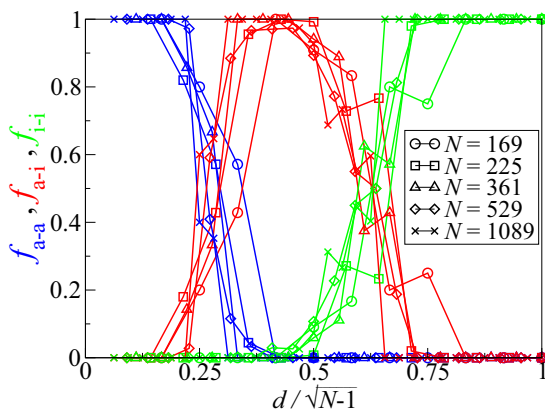


FIG. 8. (Color online) The fraction f_{a-a} , f_{a-i} , and f_{i-i} of edges connecting two active nodes, one active and one idle node, and two idle nodes as a function of rescaled distance $d/(\sqrt{N} - 1)$.

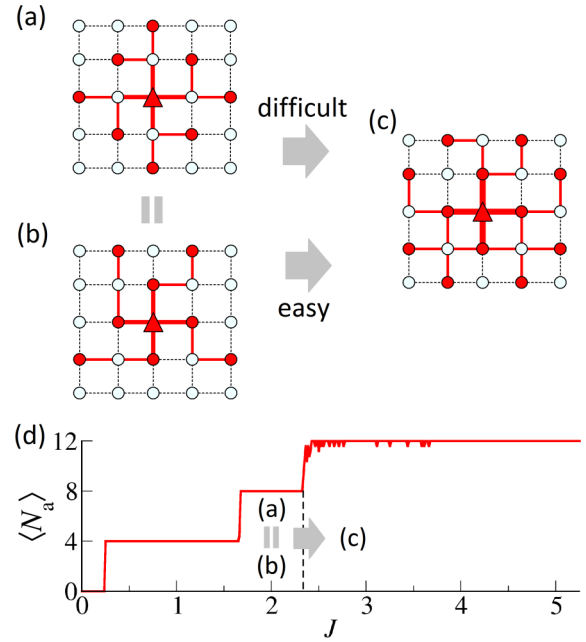


FIG. 9. (Color online) (a), (b) Two degenerate optimal states of a lattice with $N = 25$ in the range $0.083 < J < 1.11$ with $U = 0$; (c) the corresponding optimal state for $J > 1.11$; (d) the number of active nodes $\langle N_a \rangle$ as a function of J averaged over 10 realizations.

3. Suboptimal metastable states

We found that suboptimal solutions are sometimes obtained for some values of J when $U = 0$. This is in contrast to the simpler case of $J = 0$, where the optimal solution is always found by the algorithm (for instance, see Fig. 3). These result in a rough curve of the average number of active nodes N_a with $J > 1.11$, as shown in Fig. 9(d). The reason can be illustrated by the example shown in Figs. 9(a)–9(c). In the range $0.83 < J < 1.11$, both configurations (a) and (b) in Fig. 9 have the lowest energy; when $J > 1.11$, configuration (c) becomes the ground state. The transition from configuration (b) to (c) is easy since one can merely activate the four nodes next to the four corners in (b); the transition from (a) to (c) requires an extensive change of the variables for the whole network in (a), and is difficult. Configuration (a) hence becomes a local minimum separated by high energy barriers when $J > 1.11$, and the algorithm arrives at this suboptimal solution if the initial messages fall in its basin of attraction. The competition between the two antiferromagnetic configurations with opposite staggered magnetizations is crucial for the emergence of these metastable states.

One can easily eliminate this algorithmic deficiency by studying only the minimum energy state in multiple realizations. As shown in Fig. 7(b), the roughness observed in Fig. 9(d) is eliminated by measuring f_a^* , the fraction of active nodes in the instance with the lowest energy in 10 realizations. This is the reason for showing f_a^* instead of $\langle f_a \rangle$ in Fig. 7(b) in the case with $U = 0$.

B. Random regular graph

We continue to examine the optimized configuration of active nodes by running our algorithm on random regular graphs,

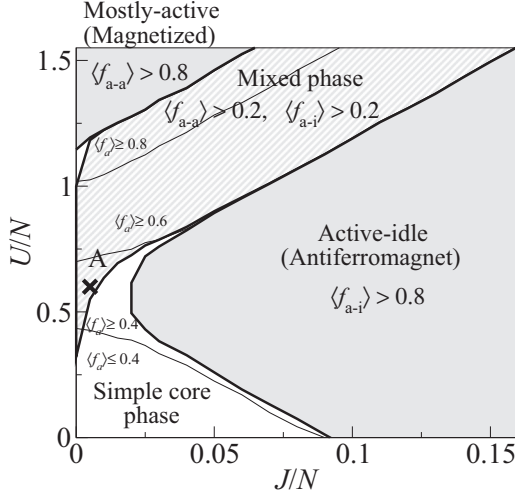


FIG. 10. Phase diagram and the fraction of active nodes as a function of J and U on random regular graphs with $N = 50$ and node degree $K = 3$. Results are averaged over 1000 realizations.

i.e., randomly connected networks with uniform degree K . In general, only loops with size $O(\ln N)$ exist in random regular graphs, which makes the assumption of independent neighbors in Eq. (4) more justified than in the square lattice. We thus expect a better algorithmic convergence, but results show that the algorithm does not converge in a broad parameter regime. For instance, as shown in Fig. 11, the ratio of convergent instances f_{con} significantly decreases when $J > 0.095$ with $U = 0$. This is due to the presence of loops with an odd number of links, which are frustrated for antiferromagnetic interactions, in contrast with square lattices whose loops consist of an even number of links, accounting for the excellent algorithmic convergence.

1. Various phases

Unlike square lattices where a unique optimized energy is found at each specific value of U and J , there is a nonzero variance of optimized energy obtained from an ensemble of finite-size random regular graphs at each U and J . We therefore observed smoothed contour lines of $\langle f_a \rangle$ in Fig. 10 compared to the zigzag contour lines in Fig. 4 of square lattices.

In addition, due to the presence of loops with an odd number of links, $\langle f_{a-i} \rangle$ over realizations is always less than 1 even in the active-idle phase. We have thus chosen a reasonable threshold on $\langle f_{a-i} \rangle$ and $\langle f_{a-a} \rangle$ as an identification criteria of the active-idle and the mixed phases. To identify the magnetized phase, we note that although f_{a-a} approaches 1 for sufficiently large U , it is nevertheless close to 1 for an extensive range of U . We have thus applied similar threshold on $\langle f_{a-a} \rangle$ to identify the magnetized phase on typical instances.

Despite these slight differences in the identification of phases, random regular graphs and square lattices have similar behavior as a function of U and J , as can be seen from the similarity of their corresponding phase diagram. As we can see in Fig. 10, random regular graphs exhibit similar phase diagram as that of square lattices, and are dominated by mostly active magnetized configurations at large value of U characterized by $\langle f_{a-a} \rangle > 0.8$, active-idle antiferromagnetic configurations at large value of J characterized by $\langle f_{a-i} \rangle > 0.8$, as well as a mixed phase for a broad parameter space outside the magnetized and antiferromagnetic regime characterized by $\langle f_{a-a} \rangle > 0.2$ and $\langle f_{a-i} \rangle > 0.2$. The simple core phase is found in the unshaded region outside all these regimes.

2. Hard computation regime

To examine the origin of the nonconvergence, we note that the drop in f_{con} in Fig. 11 is more abrupt when N increases. This suggests a phase transition at $J \approx 0.095$ and the emergence of replica symmetry breaking (RSB) in the range $J > 0.095$. RSB is a phenomenon first observed in spin glass [15] which is characterized by a rugged energy landscape with numerous local minima. When one applies a message-passing algorithm to an RSB system, individual variables may fall into states from different local minima, and leads to conflicting messages and nonconvergence. Algorithms which incorporate the picture of a rugged energy landscape can be derived by drawing further correspondence with spin glasses [30].

To analytically derive the RSB phase boundary, we study two identical replicated systems β and γ with different boundary conditions. If the states of β and γ far away from the boundary are different, long-range correlations exist, suggesting the emergence of RSB. To achieve the goal, we obtain the joint probability distribution $P[E_\beta^V(s, I), E_\gamma^V(s, I)]$ by the following recursive equation:

$$\begin{aligned}
 P[E_\beta^V(s, I), E_\gamma^V(s, I)] &= f_T \delta\{E_\beta^V(s, I) - |I|^\alpha - JC(1+s)\} \delta\{E_\gamma^V(s, I) - |I|^\alpha - JC(1+s)\} \\
 &+ (1 - f_T) \sum_{k=1}^{\infty} \frac{k\rho(k)}{\langle k \rangle} \prod_{j=1}^{k-1} \left[\int dE_{j,\beta}^V(s_j, I_j) dE_{j,\gamma}^V(s_j, I_j) P[E_{j,\beta}^V(s_j, I_j), E_{j,\gamma}^V(s_j, I_j)] \right] \\
 &\times \delta\{E_\beta^V(s, I) - \mathcal{M}[s, I; \{E_{j,\beta}^V(s_j, I_j)\}_{1 \leq j \leq k-1}] + \min_{s', I'} \mathcal{M}[s', I'; \{E_{j,\beta}^V(s_j, I_j)\}_{1 \leq j \leq k-1}]\} \\
 &\times \delta\{E_\gamma^V(s, I) - \mathcal{M}[s, I; \{E_{j,\gamma}^V(s_j, I_j)\}_{1 \leq j \leq k-1}] + \min_{s', I'} \mathcal{M}[s', I'; \{E_{j,\gamma}^V(s_j, I_j)\}_{1 \leq j \leq k-1}]\}, \quad (26)
 \end{aligned}$$

where f_T is the fraction of terminals in the infinite system assumed by the equation. Since we usually assign only one terminal in simulations, f_T should be set to the $1/N_{\text{sim}}$ for

comparison, where N_{sim} is the simulated system size. The last two lines represent two identical operations, one for replica β and the other for replica γ . Hence, the expression characterizes

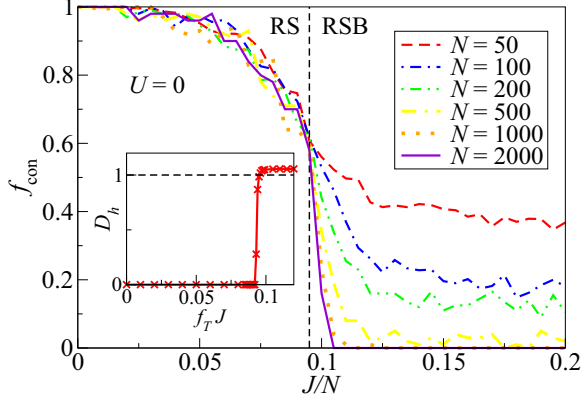


FIG. 11. (Color online) The fraction of convergent instances f_{con} as a function J/N on random regular graph with $N = 50$, $K = 3$, and $U = 0$. Inset: $\langle D_h \rangle$, the fractional change of Hamming distance given by Eq. (28) as a function $f_T J$, with $f_T = 0.02$.

two replicated systems with identical quenched disorders. A simple initial condition of uniform $P[E_\beta^V(s, I), E_\gamma^V(s, I)]$ can already distinguish the RS-RSB behaviors: if one obtains a stable solution of $P[E_\beta^V(s, I), E_\gamma^V(s, I)] = 0$ everywhere except on the diagonal $E_\beta^V(s, I) = E_\gamma^V(s, I)$, then the system has a unique global minimum and is characterized by the RS ansatz; on the other hand, if one obtains a stable solution of $P[E_\beta^V(s, I), E_\gamma^V(s, I)] \neq 0$ for the off-diagonal domain, then the system is likely to be characterized by the RSB ansatz.

Nevertheless, the stable solution of $P[E_\beta^V(s, I), E_\gamma^V(s, I)]$ is crucially dependent on the choice of initial condition in the RSB regime and is difficult to obtain. Instead of computing all the stable solutions of $P[E_\beta^V(s, I), E_\gamma^V(s, I)]$ by starting with all possible initial conditions, we introduce

$$d_\beta(I) = E_\beta^V(1, I) - E_\beta^V(-1, I), \quad (27)$$

which characterizes the energy difference between the cases when node j is active ($s_j = 1$) and idle ($s_j = -1$) in system β . We compute both $d_\beta(I)$ and $d_\gamma(I)$ and monitor the change in Hamming distance between β and γ by

$$D_h = \frac{\sqrt{\sum_I [d_\beta(I) - d_\gamma(I)]^2}}{\frac{1}{k-1} \sum_{j=1}^{k-1} \sqrt{\sum_I [d_{j,\beta}(I) - d_{j,\gamma}(I)]^2}} \quad (28)$$

during the iteration of Eq. (26). The quantity D_h is thus the fractional increase of Hamming distance during each iteration. We then average D_h over all nodes in the infinite tree generated by the population dynamics. If $\langle D_h \rangle < 1$, the difference between β and γ decreases and is eventually washed out; the system is described by the RS ansatz. If $\langle D_h \rangle > 1$, the difference increases and the system is described by the RSB ansatz. The quantity D_h obtained by $f_T = 0.02$ is shown in the inset of Fig. 11, which indicates that $J/N \gtrsim 0.095$ is characterized by the RSB ansatz at $U = 0$. The phase diagram Fig. 12 is drawn by identifying the parameter regime with $\langle D_h \rangle > 1$, which indicates the RS-RSB phase boundary. It shows that the RSB regime covers both the active-idle phase and the mixed phase.

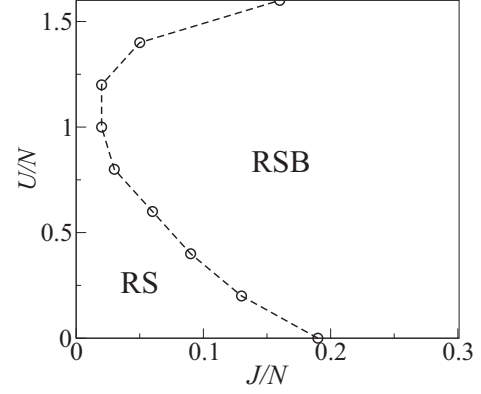


FIG. 12. A phase diagram which shows the RS-RSB boundary obtained by monitoring D_h . Results are obtained with $f_T = 0.02$ and $K = 3$.

3. Scaling with system size

As in the case of square lattices, we examine the scaling of the fraction of active nodes with increasing system size N . To find the correct scaling, we first consider the case of $J = 0$ and obtain an approximation for the supply cost $\sum_{(ij)} I_{ij}^2$ as a function of idle penalty U . We define the distance d between a node and the terminal to be the minimum number of intermediate nodes lying along a connected path from the terminal to the node. The distance between an edge and the terminal is defined in the same manner. In this case, we assume that the random regular graphs can be well described by a tree structure such that the number of edges and nodes at a distance d from the terminal is $K(K-1)^d$, the total number of nodes within a distance d is $K[1 + (K-1) + \dots + (K-1)^{d-1}] = K[(K-1)^d - 1]/[K-2]$. Again, we define d_{max} to be the distance of the farthest active node from the terminal, such that

$$f_a N \approx \frac{K[(K-1)^{d_{\text{max}}} - 1]}{K-2}. \quad (29)$$

The supply cost can be approximated by

$$\begin{aligned} \sum_{(ij)} I_{ij}^2 &\approx \sum_{d=1}^{d_{\text{max}}} \frac{(f_a N - \frac{K[(K-1)^d - 1]}{K-2})^2}{K(K-1)^d} \\ &\approx \frac{(f_a N)^2}{K} \sum_{d=1}^{d_{\text{max}}} \frac{1}{(K-1)^d}, \\ &\approx \frac{(f_a N)^2 (K-1)}{K(K-2)}. \end{aligned} \quad (30)$$

To arrive at the last two lines, we have used the fact that $K[(K-1)^d - 1]/[K-2] \ll f_a N$ when d is small, and the contribution from terms with large d is negligible. Using the above results, we can approximate the Hamiltonian in Eq. (1) by

$$E = (1 - f_a)NU + \frac{(f_a N)^2 (K-1)}{K(K-2)}. \quad (31)$$

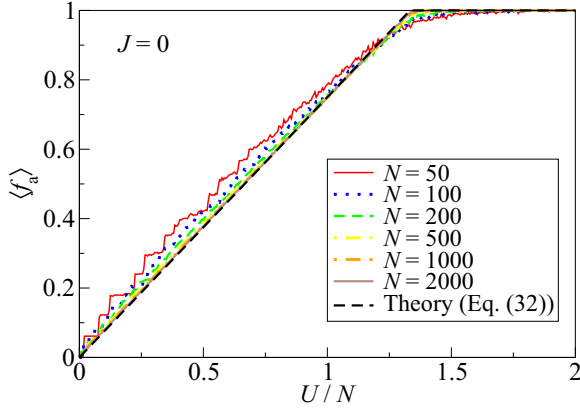


FIG. 13. (Color online) The fraction of active nodes f_a on random regular graphs as a function of U/N with $J = 0$, $K = 3$, and various N from $N = 50$ to 2000. The theoretical predictions are obtained from Eq. (32).

Minimizing E with respect to f_a , we obtain

$$f_a = \frac{U K(K-2)}{N 2(K-1)}. \quad (32)$$

As shown in Fig. 13, the simulated results of f_a for different system sizes collapse well when plotted as a function of U/N . Compared with square lattices in which f_a scales as $U/(N \ln N)$, this reveals a crucial difference between two-dimensional lattices and random graphs. This result also suggests that in the large N limit, covering a random graph is more efficient than covering a square lattice since the costs U and J scale linearly with N . From Eq. (1), we may interpret U to be the cost increase when one additional node is activated. This implies that it is less costly to set up retail networks with the same fraction of outlets in random graphs.

On the other hand, discrete jumps in f_a are only observed at small values of U in Fig. 13, unlike Fig. 7 of square lattices where steps are observed for the whole range of U and $f_a < 1$. The smoothed f_a at larger values of U is due to

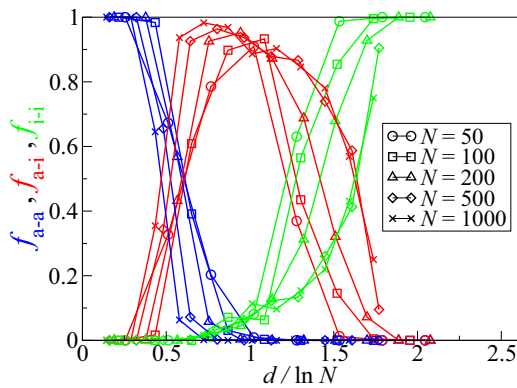


FIG. 14. (Color online) The fraction f_{a-a} , f_{a-i} , and f_{i-i} of edges connecting two active nodes, one active and one idle node, and two idle nodes as a function of rescaled distance $d / \ln N$. Results are averaged over more than 300 converged instances for $N = 50, 100, 200$ and a smaller number of instances for $N = 500, 1000$ due to lower convergence rate.

the average over realizations with different network topology. Nevertheless, the simulated f_a approaches the analytical result (32) as N increases. Data collapse of f_a is also observed in the case of increasing J with $U = 0$.

As in the case of square lattices, we examine the scaling in the mixed phase by comparing $f_{a-a}(d)$, $f_{a-i}(d)$, and $f_{i-i}(d)$ obtained from systems of different N but at the same value of U/N and J/N . As shown in Fig. 14, the results obtained by the values of U/N and J/N at point A of the phase diagram Fig. 10 indicate the presence of an active core near the terminal, followed by an active-idle region, and then by a region with higher density of idle nodes. Similar configurations are found in square lattices (see Fig. 8) at a similar point A in the phase diagram Fig. 4. Although the data do not collapse well in Fig. 14 for small system sizes, we observe a better collapse for larger system sizes, which suggests that the scaling factor $1/N$ is appropriate for cases with both U and J nonzero and at large values of N .

V. CONCLUSION

We have studied the problem of optimal facility location by mapping it to a model of interacting spins, and applied statistical physics from the studies of disordered systems to elucidate the effects of competing considerations on the system behavior. We found that in a large parameter region, a magnetized domain exists in the core region, surrounded by an antiferromagnetic domain, which is in turn surrounded by an idle domain. Aside from the coexistence of these three patterns, phases dominated by each pattern can be obtained by tuning the idle penalty and the antiferromagnetic coupling. The scaling of parameters with system size agrees with the continuum approximation in two dimensions and the tree approximation in random graphs.

The above results agree with the intuition that transportation costs favor the selection of active nodes in the vicinity of the center. However, they also show that by introducing antiferromagnetic couplings with appropriate strengths, the active nodes can spread over the arena, enhancing the coverage. Combined with the message-passing algorithm, this provides a simple decentralized control measure in such applications as sensor networks in which it is desirable to collect information spread over the entire arena.

To optimize the cost function, we derived a readily applicable algorithm which balances the need for coverage and energy saving. Compared with traditional mixed integer programming, this algorithm scales favorably with system size. In addition to the conventional task of locating facilities, our algorithm identifies the optimal quantity of demand and optimizes individual paths. The local expansion approach suggested in the message updates also sheds light on similar simplifications in other problems. Compared with traditional approaches of global optimization, the message-passing algorithm is useful in applications to large networks and dynamically evolving networks. In large networks, decentralized approaches have advantages over global optimization, which scales up rapidly with system size and requires heavy overhead. In dynamically evolving networks such as mobile sensor networks, the changing topology implies that global optimization approaches are not temporally relevant. Rather,

locally optimal solutions should be able to cope with the evolving situations.

On the other hand, due to frustrations caused by the competition between coverage and supply cost, a transition between easy and hard computation regimes is observed. The hard computation regime coincides with the antiferromagnetic phase and the mixed phase, where metastable states are prevalent. Convergence of the message-passing approach is affected. Further computational techniques such as decimation [30] or reinforcement [33] may be introduced to alleviate the problem

ACKNOWLEDGMENTS

We thank D. Saad, J. Raymond, A. Fung, H. Wang, and P. Choi for fruitful discussions. This work is supported by Research Grants Council of Hong Kong (Grants No. 605010, No. 604512, and No. 605813), EU FET project STAMINA (Project No. FP7-265496) and Royal Society Exchange Grant No. IE110151.

APPENDIX: TOTAL COST IN THE CONTINUUM APPROXIMATION FOR $J = 0$

We first calculate the supply cost to the order N^2 . Assuming that the flows are isotropic, the flux (i.e., the flow per unit cross sectional length) at distance r from the center is given by

$$J(r) = \frac{\pi(d_{\max}^2 - r^2)}{2\pi r}. \quad (\text{A1})$$

Hence, the supply cost is calculated to be

$$E_{\text{supply}} = \int_{\epsilon}^{d_{\max}} J^2(r) 2\pi r dr = \frac{\pi d_{\max}^4}{2} \left(\ln \frac{d_{\max}}{\epsilon} - \frac{3}{4} \right), \quad (\text{A2})$$

where we have assumed $\epsilon \ll d_{\max}$. The result is rather sensitive to the cutoff distance ϵ in two-dimensional lattices. We derive its value by noting that the flux can be considered as the

gradient of the chemical potential. By requiring the potential difference between the center and its neighbor to be equal to the flow in their link, we can obtain a self-consistent condition for the cutoff. Specifically, we let $J(r) = -d\mu(r)/dr$, so that

$$\mu(r) = - \int J(r) dr = - \frac{d_{\max}^2}{2} \ln r + \frac{r^2}{4}. \quad (\text{A3})$$

The flow in a link crossing the boundary is given by

$$\mu(0) - \mu(1) = - \frac{d_{\max}^2}{2} \ln \epsilon - \frac{1}{4} = \frac{f_a N}{4}. \quad (\text{A4})$$

Using $\pi d_{\max}^2 = f_a N$, we arrive at $\epsilon = \exp(-\pi/2)$, and Eq. (A2) becomes

$$E_{\text{supply}} = \frac{f_a^2 N^2}{4\pi} \left(\ln(f_a N) - \ln \pi + \pi - \frac{3}{2} \right). \quad (\text{A5})$$

The total cost becomes

$$E = (1 - f_a) N U + \frac{f_a^2 N^2}{4\pi} \left(\ln(f_a N) - \ln \pi + \pi - \frac{3}{2} \right) + \frac{f_a^2 N^2}{4\pi} \ln f_a. \quad (\text{A6})$$

The optimal fraction of active nodes satisfies the condition

$$f_a (\ln N + \ln f_a - \ln \pi + \pi - 1) = \frac{2\pi U}{N}. \quad (\text{A7})$$

Similarly, one can derive the fraction of active nodes as a function of J when $U = 0$. Since only half of the nodes are active within the distance d_{\max} in this case, $\pi d_{\max}^2 = 2f_a N$ and the flux is given by

$$J(r) = \frac{1}{2} \frac{\pi(d_{\max}^2 - r^2)}{2\pi r}. \quad (\text{A8})$$

Following the line of the above calculation, we arrive at

$$f_a \left(\ln N + \ln f_a - \ln \frac{\pi}{2} + \pi - 1 \right) = \frac{16\pi J}{N}. \quad (\text{A9})$$

-
- [1] S. Rangwala, R. Gummadi, R. Govindan, and K. Psounis, *Comput. Commun. Rev.* **36**, 63 (2006).
- [2] N. Megiddo, E. Zemel, and S. L. Hakimi, *SIAM J. Algebraic Discrete Methods* **4**, 253 (1981).
- [3] A. Vasani and S. P. Simonovic, *J. Water Res. Planning Manage.* **136**, 279 (2010).
- [4] A. Ghosh and C. S. Craig, *J. Marketing* **47**, 56 (1983).
- [5] O. Berman, R. C. Larson, and N. Fouska, MIT Operations Research Center Working Paper OR 231-90, 1990 (unpublished).
- [6] C. ReVelle *et al.*, *Health Services Res.* **12**, 129 (1977).
- [7] J. N. Al-Karaki and A. E. Kamal, *IEEE Wireless Commun.* **11**, 6 (2004).
- [8] H. Frey, S. Rührup, and I. Stojmenović, in *Guide to Wireless Sensor Networks*, edited by S. Misra, S. C. Misra, and I. Woungang (Springer, London, 2009).
- [9] D. Seward, *Napoleon and Hitler* (Viking, New York, 1988).
- [10] T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan, and K. K. Saluja, in *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications, WSNA'02, 2002* (ACM, New York, 2002), pp. 42–48.
- [11] Y. Zou and K. Chakrabarty, in *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, INFOCOM 2003, San Francisco, CA* (IEEE, Piscataway, NJ, 2003), Vol. 2, pp. 1293–1303.
- [12] F. Plaustria, *Eur. J. Operat. Res.* **129**, 461 (2001).
- [13] K. Y. M. Wong and D. Saad, *J. Phys. A: Math. Theor.* **41**, 324023 (2008).
- [14] J. Vannimenus and G. Toulouse, *J. Phys. C: Solid State Phys.* **10**, L537 (1977).
- [15] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [16] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, UK, 2001).
- [17] K. Y. Michael Wong and D. Saad, *Phys. Rev. E* **74**, 010104(R) (2006).
- [18] C. H. Yeung and D. Saad, *Phys. Rev. Lett.* **108**, 208701 (2012).

- [19] W. Selke, *Phys. Rep.* **170**, 213 (1988).
- [20] C. Domb and R. B. Potts, *Proc. R. Soc. London, Ser. A* **210**, 125 (1951).
- [21] J. Stephenson and D. D. Betts, *Phys. Rev. B* **2**, 2702 (1970).
- [22] P. Bak and J. von Boehm, *Phys. Rev. B* **21**, 5297 (1980).
- [23] J. Vannimenus, *Z. Phys. B: Condens. Matter Quanta* **43**, 141 (1981).
- [24] S. Inawashiro, C. J. Thompson, and G. Honda, *J. Stat. Phys.* **33**, 419 (1983).
- [25] N. N. Ganikhodjaev, C. H. Pah, and M. R. B. Wahiddin, *J. Phys. A: Math. Gen.* **36**, 4283 (2003).
- [26] H. Pirkul and V. Jayaraman, *Comput. Operat. Res.* **25**, 869 (1998).
- [27] In typical facility location problems, the demands of the clients at different fixed locations are constrained to be satisfied and the locations of the facilities are selected to minimize a cost function [26]. Alternatively, the cost of installing the facilities is fixed and the locations of the clients or facilities are selected to minimize the population deprived of the facilities [28]. This brings the mathematical formulation comparable to our model, except for the presence of antiferromagnetic couplings and the quadratic transportation cost.
- [28] G. C. Moore and C. ReVelle, *Manage. Sci.* **28**, 775 (1982).
- [29] IBM ILOG CPLEX optimization studio, <http://www-03.ibm.com/software/products/en/ibmilogcpleoptistud/>.
- [30] M. Mézard and R. Zecchina, *Phys. Rev. E* **66**, 056126 (2002).
- [31] C. H. Yeung, in *IEEE International Conference on Communications Workshops (ICC), Budapest* (IEEE, Piscataway, NJ, 2013), pp. 1420–1424.
- [32] C. H. Yeung, D. Saad, and K. Y. M. Wong, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13717 (2013).
- [33] J. Chavas, C. Furtlehner, M. Mézard, and R. Zecchina, *J. Stat. Mech.* (2005) P11016.