Distributed Optimization in Transportation and Logistics Networks

K. Y. Michael WONG[1], David SAAD[2], and Chi Ho YEUNG[3]
[1]Hong Kong University of Science and Technology, Hong Kong
[2]Aston University, Birmingham, B4 7ET, UK
[3]Education University of Hong Kong, Hong Kong
Email: phkywong@ust.hk

Abstract

Many important problems in communication networks, transportation networks, and logistics networks are solved by the minimization of cost functions. In general, these can be complex optimization problems involving many variables. However, physicists noted that in a network, a node variable (such as the amount of resources of the nodes) is connected to a set of link variables (such as the flow connecting the node), and similarly each link variable is connected to a number of (usually two) node variables. This enables one to break the problem into local components, often arriving at distributive algorithms to solve the problems. Compared with centralized algorithms, distributed algorithms have the advantages of lower computational complexity, and lower communication overhead. Since they have a faster response to local changes of the environment, they are especially useful for networks with evolving conditions. This review will cover message-passing algorithms in applications such as resource allocation, transportation networks, facility location, traffic routing, and stability of power grids.

1. Introduction

Optimization of network flows is one of the most important problems in science with many areas of application [13]. It has found wide application in circuits transporting electric currents, transportation networks, communications networks, hydraulic networks, mammalian circulatory systems and vascular systems in plants [1,7,14,27,30]. A unified approach to these problems is facilitated by the minimization of cost functions. For example, the cost functions may represent the dissipation energy (via Thomson's principle for electric currents) [13] or time delays in communications networks. There is a close relation between the flow patterns and the cost functions.

Traditionally, network resource allocation and routing problems have been solved using global optimization techniques, such as linear or quadratic programming [4]. However, with the increasing sizes of fixed networks and the evolving configuration of wireless networks, centralized control becomes increasingly costly and infeasible. Distributed control in networks involves a group of independent controllers which make locally optimal decisions. Compared with the traditional centralized approach, this has the advantage of less computational load and communication overhead, and robustness against network breakdown. The Dynamic Alternative Routing of British Telecom was an early successful example [19]. Also, in computer science, many algorithmic solutions have been proposed to distribute computational load among computers connected in a network, but they are usually more heuristic. Some of them may tend to optimize the benefit to an individual node or task, without considering the impact it makes to the rest of the network [5].

Message-passing algorithms originated from two threads. In the physics literature, the microscopic description of disordered systems is derived from the Thouless-Anderson-Palmer (TAP) equations [31], which were subsequently generalized to become the cavity method [22], and resulted in many computationally efficient schemes. These approaches have laid the foundation for the study of complex systems. Applications can be found in associative memories, perceptron learning, error-correcting codes, image restoration, CDMA multiuser demodulation [26], data compression [23] and compressed sensing [21]. In complex optimization, the theory has been applied to problems with discrete variables, such as graph partitioning [26], travelling salesman [22], number partitioning [26], *K*-satisfiability [20], graph coloring [25], and coloring diversity [8].

Parallel to the development in the physics community, the informatics community applied graphical models to probabilistic information processing [15]. By mapping the probabilistic relations between the parameters to links in graphs, the problems become factorized and can be solved by iterating the message-passing equations that relate the conditional probabilities on neighboring nodes. This technique has been fruitfully applied to pattern classification, image restoration, error-correcting codes and data compression. For example, the famous belief propagation (BP) algorithm has been successfully applied to error-correcting codes.

In this review, we consider the application of message-passing algorithms to resource and flow allocation problems in transportation and logistics networks. Our contributions have been presented in several previous publications [32,33,36,37,35,38,18, 39,16,17].

2. The Resource Allocation Problem

We start with a typical model of resource allocation on networks [32,33]. Consider a network of N nodes, labeled $i = 1, \ldots, N$. The set of neighbors of node i is given by ∂i. Let Λ_i be the capacity of node i. Positive Λ_i represents the provision of resources, and negative Λ_i their consumption. The objective of optimization is to transport the resources along the links so that the total transportation cost is minimized, while the final quantity of resources of each node becomes non-negative.

Let $y_{ij} \equiv -y_{ji}$ be the flow on the link from j to i, and an even function $\phi(y_{ij})$ the transportation cost along link (ij). Depending on the context, $\phi(y)$ can be a convex or concave function of y_{ij}. If $\phi(y)$ is convex, for example, $\phi(y) \propto y^\gamma$ for $\gamma > 1$, it tends to homogenize the flow. This is useful when traffic is heavy and one aims to avoid congestion. On the other hand, if $\phi(y)$ is concave, for example, $\phi(y) \propto y^\gamma$ for $\gamma < 1$, it tends to concentrate the flows in a few links. This is useful when traffic is light or one aims at consolidating the resources to utilize fewer links [34].

Hence the optimization problem becomes the minimization of $E = \Sigma_{(ij)}\phi(y_{ij})$ subject to

$$\sum_{j\in\partial i} y_{ij} + \Lambda_i \geq 0 \text{ for all } i. \tag{1}$$

Introducing Lagrange multipliers, the function to be minimized becomes

$$L = \sum_{(ij)} \phi(y_{ij}) + \sum_i \mu_i \left(\sum_{j\in\partial i} y_{ij} + \Lambda_i \right) \text{ and } \mu_i \leq 0 \ \forall i. \tag{2}$$

The non-positivity of μ_i arises from the minimization of L. Optimizing L with respect to y_{ij}, one obtains

$$y_{ij} = \phi'^{-1}(\mu_j - \mu_i), \tag{3}$$

where μ_i is referred to as the chemical potential of node i, and ϕ' is the derivative of ϕ with respect to its argument. This can be interpreted as the current being driven by the potential difference.

The chemical potentials can be obtained by solving Eq. (1) together with the non-positivity of μ_i. In particular, for the quadratic cost $\phi(y) = y^2/2$,

$$\mu_i = \min\left[\frac{1}{|\partial i|} \left(\sum_{j\in\partial i} \mu_j + \Lambda_i \right), 0 \right]. \tag{3}$$

Alternatively, we introduce the cavity method. This method is known to be exact in sparse networks. Since the probability of finding loops of finite lengths is vanishing in large sparse networks, the structure of a sparse network can be approximated by a tree locally, and the correlations among the branches of a tree are neglected. In each branch, nodes are arranged in generations. A node is connected to an ancestor node of the previous generation, and node i is connected to $|\partial i| - 1$ descendant nodes of the next generation.

Suppose the ancestor of node j is node i, and its descendants are labeled by $k \in \partial j \backslash i$. Then the total energy $E_{j\to i}(y_{ij})$ of the tree terminated at node j can be expressed as the energies $E_{k\to j}(y_{jk})$ of its descendants that branch out from node j

$$E_{j\to i}(y_{ij}) = \min_{\{y_{jk} | \sum_{k\in\partial j\backslash i} y_{jk} - y_{ij} + \Lambda_j \geq 0\}} \left[\sum_{k\in\partial j\backslash i} E_{k\to j}(y_{jk}) + \phi(y_{ij}) \right] \tag{4}$$

In the framework of the cavity method, these energies are *cavity* energies, since the effects of the ancestor nodes are not accounted for. The local nature of their recursion relation points to the possibility that the network optimization can be solved by message-passing approaches. However, in contrast to other message-passing algorithms that pass conditional probability estimates of discrete values to neighboring nodes, the messages in the present context are more complex, since they are functions $E_{j\to i}(y_{ij})$ of the current y_{ij}.

We simplify the messages to two parameters, namely, the first and second derivatives of the vertex energies. Let

$$(A_{ij}, B_{ij}) = \left(\frac{\partial E_{j\to i}(y_{ij})}{\partial y_{ij}}, \frac{\partial^2 E_{j\to i}(y_{ij})}{\partial y_{ij}^2} \right)$$

be the message passed from node j to i. Based on the messages received from the descendants $k \neq i$, the vertex energy from j to i can be obtained by minimizing the energy in the space of the current adjustments ε_{jk} drawn from the descendants. The optimal solution is given by

$$E_{j\to i}(y_{ij}) = \phi(y_{ij}) + \sum_{k\in\partial j\backslash i} \frac{\mu_{ij}^2 - A_{jk}^2}{2B_{jk}}, \tag{5}$$

where

$$\mu_{ij} = \min\left[\frac{\sum_{k\in\partial j\setminus i}(y_{jk} - B_{jk}^{-1}A_{jk}) - y_{ij} + \Lambda_j}{\sum_{k\in\partial j\setminus i} B_{jk}^{-1}}, 0\right]. \quad (6)$$

The first and second derivatives of the optimal solution lead to the forward message

$$A_{ij} = \phi'(y_{ij}) - \mu_{ij}, \quad B_{ij} = \phi''(y_{ij}) + \frac{\Theta(-\mu_{ij} - \varepsilon)}{\sum_{k\in\partial j\setminus i} B_{jk}^{-1}}, \quad (7)$$

To calculate the flow y_{ij} on a link, we can consider the link as the bridge between two trees, one with vertex i and the other j, with flows y and $-y$ from the vertices respectively. Taking into account the double-counting of the transportation cost on the bridge, the current is given by

$$y_{ij} = \arg\min_{\{y\}}\left[E_{i\to j}(y) + E_{j\to i}(-y) - \phi(y)\right] \quad (8)$$

The average transportation cost per link can also be calculated from

$$\left\langle \Delta E_{\text{link}} \right\rangle = \left\langle \phi(y)\Big|_{y=\arg\min_{\{y'\}}\left[E_{i\to j}(y') + E_{j\to i}(-y') - \phi(y')\right]} \right\rangle. \quad (9)$$

The message-passing algorithm achieves global optimization by randomly selecting links and passing messages to neighbors to calculate the optimal flows.

It can be verified that the message-passing algorithm, in the two-parameter approximation, yield solutions identical to the chemical potential algorithm, which is exact even when loops are present, as long as the algorithms converge [33].

3. Networks with Finite Bandwidths

The message-passing algorithm can be extended to consider the effects of bandwidths of the transportation links [36,37]. In communication networks, connections usually have assigned bandwidths. Bandwidths limit the flows in the links. However, in these networks, nodes with resource demand can still experience shortage even though their neighbors have adequate supply of resources, since the provision of resources can be limited by the bandwidths of the links. Hence the cost function is generalized to include the cost of shortage of resources, and the problem becomes the minimization of the cost function

$$E = \sum_i \psi(\xi_i) + \sum_{(ij)} R\phi(y_{ij}) \quad (10)$$

subject to constraints

$$\sum_{j\in\partial i} y_{ij} + \Lambda_i + \xi_i \geq 0, \quad \xi_i \geq 0, \quad -W \leq y_{ij} \leq W. \quad (11)$$

For quadratic costs $\phi(y) = y^2/2$ and $\psi(\xi) = \xi^2/2$, the flow is related to the potential difference by $y_{ij} = Y(\mu_j - \mu_i)$, where

$$Y(x) = \max\left\{-W, \min\left[W, [\phi']^{-1}\left(\frac{x}{R}\right)\right]\right\}. \quad (12)$$

The message-passing algorithm can also be worked out. Here we describe some interesting results due to the bandwidth constraints.

Links of three types can be identified. Those links with $|y_{ij}| = W$ are referred to as *saturated* links. Those with $0 < |y_{ij}| < W$ and $|y_{ij}| = 0$ are referred to as *unsaturated* and *idle* links respectively. From figure 1, we note that as the bandwidth decreases, the fraction of saturated links increases since for a link to transport the same flow, it is easier to saturate a link with low bandwidth than one with higher bandwidth.

To provide sufficient resources to the consumer nodes, one would anticipate a decreasing fraction of idle links as the bandwidth decreases, since more links should participate in the task of resource allocation. Surprisingly, we notice an increasing fraction of idle links as the bandwidth decreases, in contrast with our anticipation.

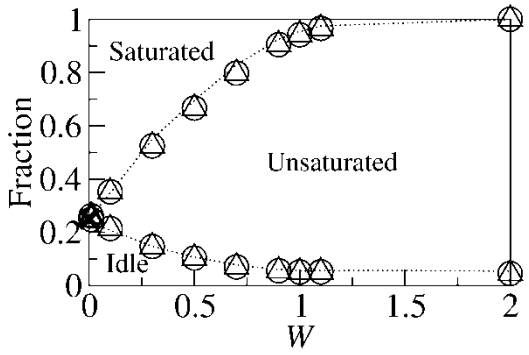

Figure 1. The fraction of idle, unsaturated and saturated links as a function of bandwidth for $\langle\Lambda\rangle = 0$ [36].

This is a consequence of the *bottleneck effect*, as illustrated in figure 2. When the bandwidth decreases, resources transferred from the secondary neighbors may become redundant since resources from nearest neighbors already saturate the link to the unsatisfied node, which can therefore be considered as a

bottleneck in transportation. The existence of bottlenecks is common in many real networks. Among the most common examples are bottlenecks occurring in traffic congestion.

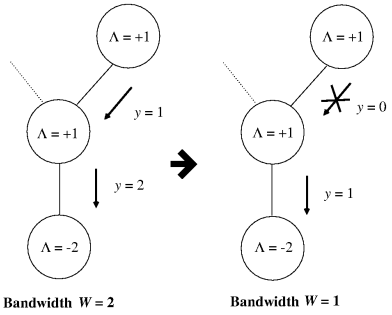

Figure 2. An example of a bottleneck effect [36].

Next, we consider the high connectivity limit. In this limit, the transportation is so efficient that no nodes suffer from shortage for positive $\langle\Lambda\rangle$. For negative $\langle\Lambda\rangle$, all resources from the providers can be distributed to the consumers. For the quadratic cost $\psi(\xi) = \xi^2/2$, $\mu = -\xi$, and the chemical potential is equal to the final resource (or the minus of the shortage). In the high connectivity limit, the chemical potential μ of a node becomes a well-defined function of the initial resource Λ.

For quadratic costs $\phi(y) = y^2/2$ and $\psi(\xi) = \xi^2/2$, and a Gaussian distribution of Λ with variance 1, this function is monotonic as long as $W \le \sqrt{\pi/2}/c$, where c is the connectivity. However, when $W > \sqrt{\pi/2}/c$, turning points exist as shown in figure 3(a). This creates a thermodynamically unstable scenario, since in the region with negative slope, nodes with lower capacities have higher chemical potentials than their neighbors with higher capacities. This implies that resource flow from poorer nodes to richer ones. Nevertheless, there exists another stable solution in which the unstable region is replaced by a range of constant μ as shown in figure 3(b), analogous to Maxwell's construction in thermodynamics. The position of this construction can be determined by the conservation of resources, which implies that the areas A and B in figure 7(b), weighted by the distribution of Λ, should be equal.

Nodes with uniform chemical potentials represent clusters of nodes interconnected by an extensive fraction of unsaturated links, which provides the freedom to fine-tune their flows so that the shortages among the nodes are uniform. They are referred to as *balanced* nodes. Their quantity is a measure of the efficiency of resource allocation.

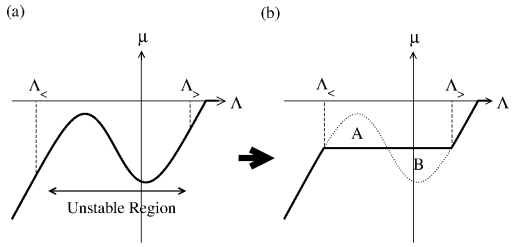

Figure 3. Maxwell's construction on $\mu(\Lambda)$ [36].

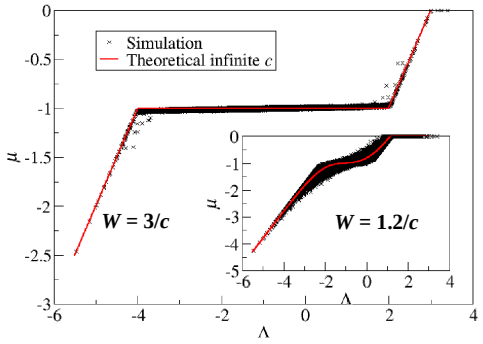

Figure 4. The simulation results for $\mu(\Lambda)$ for N = 10,000, c = 15, R = 0.1, $\langle\Lambda\rangle = -1$ and $W = 3/c$ with 70,000 data points, compared with the theoretical prediction. Inset: the corresponding results for $W = 1.2/c$ [36].

As shown in figure 4, the simulation results agree with the analytical results. The presence of the balanced nodes at high bandwidth, and the absence at low bandwidth, are evident.

It is well known that communication networks have scale-free structures [2]. Their connectivity distribution obeys a power law, and is characterized by the presence of hubs, which can modify the network behavior significantly. Figure 5 shows the simulation results for nodes with connectivity 3 and 10 in a scale-free network. The data points are consistent with the analytical results for both sets of nodes. This implies that the previous argument of increasing efficiency by increasing connectivity also holds for scale-free networks, as a smaller gradient is found for nodes with higher connectivity.

More important, nodes with low connectivity benefit from the presence of hubs in the networks. To see these benefits, the simulation results for nodes in scale-free networks are compared with nodes in regular networks of the same connectivity. As shown in figure 5(b), the data points from regular networks are more scattered away from the Maxwell's construction, when compared with those from scale-free networks. This shows that the presence of hubs increases the efficiency of the entire network, especially for nodes with low connectivity. This provides support for scale-free networks being better candidates for resource allocation than regular networks.

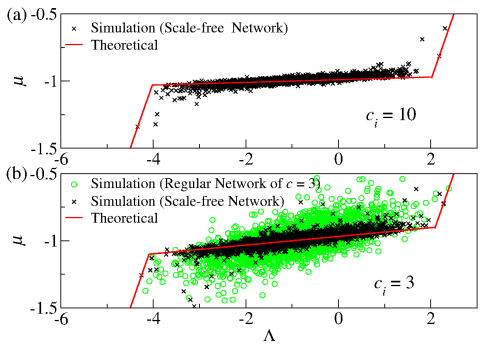

Figure 5. Simulation results for (Λ, μ) for $N = 2 \times 10^5$, $R = 0.1$, $W_{ij} = 3/\max(c_i, c_j)$ and $\langle\Lambda\rangle = -1$ as compared with theoretical results, for (a) nodes with $c_i = 10$ in scale-free networks with $P(c_i) \sim c_i^{-3}$, (b) nodes i with $c_i = 3$ in scale-free networks and nodes in regular networks with $c = 3$. Each data set contains 2,500 data points [36].

4. Optimal Source Location

So far we have considered the issue of optimizing transportation costs by adjusting the flow of resources. A further issue in network design and optimization involves selecting additional locations to install source nodes. The optimal source location problem has wide applications [35,38]. For example, to determine the optimal locations of access points in wireless networks, one needs to balance the installation cost of the access points and the power transmission cost of the channels linking the subscribers.

With network applications in mind, we consider the case in which a fraction ϕ_s of nodes are *surplus* nodes with $\Lambda_i = A$ (>> 1), and a fraction of $\phi_d \equiv 1 - \phi_s$ of nodes are *deficient* nodes with $\Lambda_i = -1$. Formally, we introduce the state variables $s_i = \pm 1$. For a surplus node i, s_i is fixed at +1, whereas for a deficient node i, $s_i = -1$ when it is converted to a source node with $\Lambda_i = A$, and $s_i = +1$ when it remains a consumer node. The cost function is then

$$E = \frac{u^2}{4}\sum_i (1 - s_i) + \frac{1}{2}\sum_{(ij)} y_{ij}^2, \qquad (13)$$

subject to $s_i\left(\Lambda_i + \sum_{j\in\partial i} y_{ij}\right) \geq 0$. $u^2/2$ is the installation cost of converting an initially deficient node to a source node.

The optimal flow is given by $y_{ij} = \mu_j s_j - \mu_i s_i$ and $\mu_i = \min\left[0, \left(\Lambda_i + \sum_{j\in\partial i}\mu_j s_j\right)\Big/(s_i \mid \partial i \mid)\right]$. The set of optimal $\{s_i\}$ can be found by an approach similar to the GSAT algorithm [28].

As shown in figure 6 for networks with regular connectivity $c = 3$, two phases can be identified: (1) the *all-source* phase for $u^{-1} \geq \sqrt{3}$, in which all nodes are assigned as source nodes; (2) the *partial-source* phase for $0 < u^{-1} < \sqrt{3}$, in which only some nodes are assigned as source nodes. The fraction of source nodes is a discontinuous function of u^{-1}, showing abrupt jumps at threshold values of u^{-1}. The step size of the curve decreases as u^{-1} decreases, and gradually becomes unresolvable in the numerical experiments. This resembles the Devil's staircase observed in the circle map and other dynamical systems [11]. These thresholds mark the positions at which certain configurations of the source and consumer nodes become energetically stable. The regime $\sqrt{3/2} < u^{-1} < \sqrt{3}$ with isolated consumer nodes is the *singlet* regime, and $\sqrt{21/25} < u^{-1} < \sqrt{3/2}$ is the *doublet* regime.

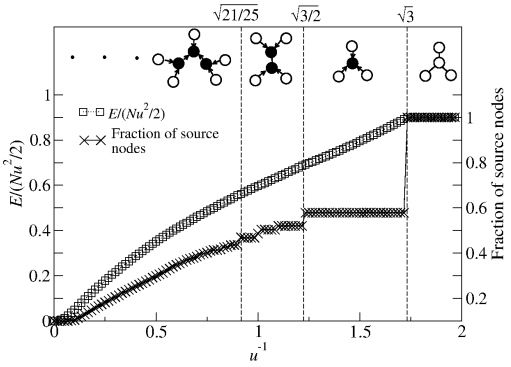

Figure 6. Simulation results of average energy per node and the fraction of network nodes acting as source nodes. Parameters: $c = 3$, $N = 100$, and 90% of the nodes are deficient. New clusters formed on increasing u^{-1} are sketched on top, with filled and unfilled circles representing consumer and source nodes respectively [38].

Using the cavity method, we can show that in the singlet regime, only 3 cavity states are relevant. They form a closed set under recursion: source (*S*), consumer (*C*), and bistable (*B*). The recursion relations reduce to

$$S/B + \cdots + S/B \rightarrow C, \qquad (14a)$$

$$S/B + \cdots + S/B + C \rightarrow B, \qquad (14b)$$

$$\text{all other combinations} \rightarrow S. \qquad (14c)$$

In equations (14a) and (14b), the states of $c - 1$ and $c - 2$ descendants are either *S* or *B* respectively. Similarly, the *full* states of a node are denoted as **C**, **B** and **S**, representing respectively the consuming, bistable and source states. They are obtained via

$$\mathbf{S}/\mathbf{B} + \cdots + \mathbf{S}/\mathbf{B} \rightarrow \mathbf{C}, \qquad (15a)$$

$$\mathbf{S}/\mathbf{B} + \cdots + \mathbf{S}/\mathbf{B} + \mathbf{C} \rightarrow \mathbf{B}, \qquad (15b)$$

$$\text{all other combinations} \rightarrow \mathbf{S}. \qquad (15c)$$

However, discrepancy exists between the simulated and predicted results of the average energy when the fraction of initially deficient nodes is high. We thus examine the stability of the assumption that the optimal state is unique (the so-called replica symmetric (RS) ansatz). We define $\eta_s^{j\to i}=1$ if the cavity state of node j excluding i is uniquely S (that is, the probability of the cavity state to be S is exactly 1), and 0 otherwise. Similarly, $\eta_c^{j\to i}=1$ if the cavity state is uniquely C, and 0 otherwise. We further introduce $\eta_g^{j\to i}=1-\eta_s^{j\to i}-\eta_c^{j\to i}$. $\eta_g^{j\to i}>0$ indicates the occurrence of *glassy* behavior.

As shown in figure 7, the picture of a unique optimal solution breaks down when the fraction of initially deficient nodes is greater than 0.75 for $c = 3$. Similar to the point where the RS solution in disordered systems becomes unstable, this transition is called the Almeida-Thouless (AT) transition [10].

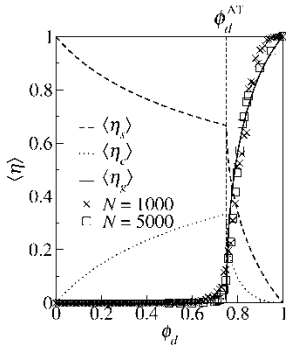

Figure 7. The site averages of the variables $\eta_s^{j\to i}$, $\eta_s^{j\to i}$ and $\eta_s^{j\to i}$ as a function of ϕ_d, the fraction of initially deficient nodes for $c = 3$. The symbols represent the simulated fraction of non-converging messages. ϕ_d^{AT} locates the AT transition [38].

To improve the analysis, entropic effects have to be considered since bistable states exist [18]. Recursion relations similar to the minimal vertex problem were introduced [40]. The message passed from node j to node i consists of two cavity variables: the probability of node j being in the S state (when node i is excluded), and the entropy change when node j and its adjacent links are added, except (ij). Extending the analysis to the picture that multiple clusters of optimal solutions exist (the so-called one-step replica symmetry-breaking solution), the predicted fraction of source nodes is consistent with the asymptotic limit obtained by extremal optimization [6].

Accompanying entropic effects is the appearance of frozen nodes, which are those initially deficient nodes taking the same state (source or consumer) in all optimal solutions. In networks where the positions of the initially deficient nodes are randomly distributed, it is sufficient to consider freezing on the 2-core subgraph, that is, the graph that remains after recursively removing the initially deficient nodes of connectivity one or lower. If the thermodynamics of the 2-core is simple, so is the entire graph, since the initially deficient nodes outside the 2-core are in tree-like structures and their states can be determined accordingly.

We classify the deficient nodes in the network into 2-core nodes and peripheral nodes (those that are not on 2-core). Figure 8(a) and (d) show that the fraction deviates from the RS prediction when ϕ_s is small, whereas the fraction of frozen peripheral nodes is almost consistent with the RS prediction in the entire range of ϕ_s. Among the 2-core nodes for $c = 3$, we further classify the nodes into hubs (those connected to 3 other 2-core nodes) and chain nodes (non-hubs connected to only 2 other 2-core nodes). Comparing figures 10(b) and (c), we conclude that the deviation from the RS prediction is primarily due to the hubs rather than other substructures.

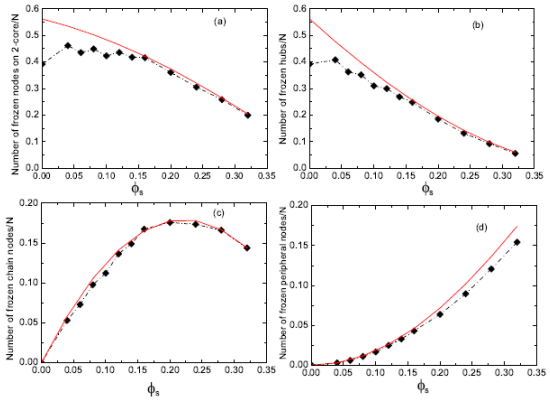

Figure 8. The dependence of the fraction of frozen nodes on the fraction of surplus nodes ϕ_s. (a) All frozen 2-core nodes considered. (b) Frozen hubs. (c) Frozen chain nodes. (d) Frozen peripheral nodes. Lines: replica symmetric prediction. Symbols: Asymptotic results of the extremal optimization algorithm [18].

5. Facility Location Problem

So far we have considered the optimization of transportation costs on networks, but there are many communication and logistics networks in which the coverage of a geographical region is equally crucial. Examples include networks of fire sensors, surveillance video cameras, local weather monitors, locations of supermarket branches, teller machines, chain store outlets, and public facilities such as such as schools and clinics.

As an example illustrating the need to balance coverage and transportation cost, consider a

population of sensors, each with simple structure, used to collect local information in a region. The sensors form a network sending the collected information to a central base station [6,9]. Due to the limited power available to each sensor, the active lifetime of the network may be short. To prolong the lifetime of the network, an alternative is to render some sensors inactive. If the transportation cost is to be minimized, then only those sensors in the neighborhood of the base station will be activated. However, to reduce the amount of information loss, the active sensors should be spread out.

Hence we consider a transportation network and introduce $S_i = 1, -1$ when node i is active or idle, respectively [23]. The cost function is given by

$$E = \sum_{(ij)} \phi(y_{ij}) + \frac{U}{2} \sum_i (1 - S_i) + J \sum_{(ij)} S_i S_j . \qquad (16)$$

Compared with previous models, there are two extra terms. The second term represents the loss in value due to idling a node, and U is the *turn-off cost*. The last term tends to spread the active nodes and hence increase the coverage, and J is the *redundancy cost*.

In the following example, $\phi(x) = x^2$, the flow originating from each node is $(1 + S_i)/2$, and all flows terminate at the base station.

As shown in figure 9(a), a typical configuration consists of an active core around the center, surrounded by an area of alternating active and idle nodes, and an outer inactive band.

Figure 10 shows the phase diagram in the space of U and J. The examples in figure 9 belong to the mixed phase at the center of the diagram. When U increases, the active core expands until it covers the entire lattice, and results in the all-active phase. When J increases, the active-idle band expands until it covers the entire lattice, resulting in the active-idle phase. When U or J decreases, we have an active-idle core or an all-active core surrounded by an all-idle region.

With the extensive regions of the all-active and active-idle states in the phase diagram, 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. However, figure 11 shows that the energy of the optimal state is significantly lower than the all-active and active-idle states when the fraction of active nodes lie between 0.5 and 1.

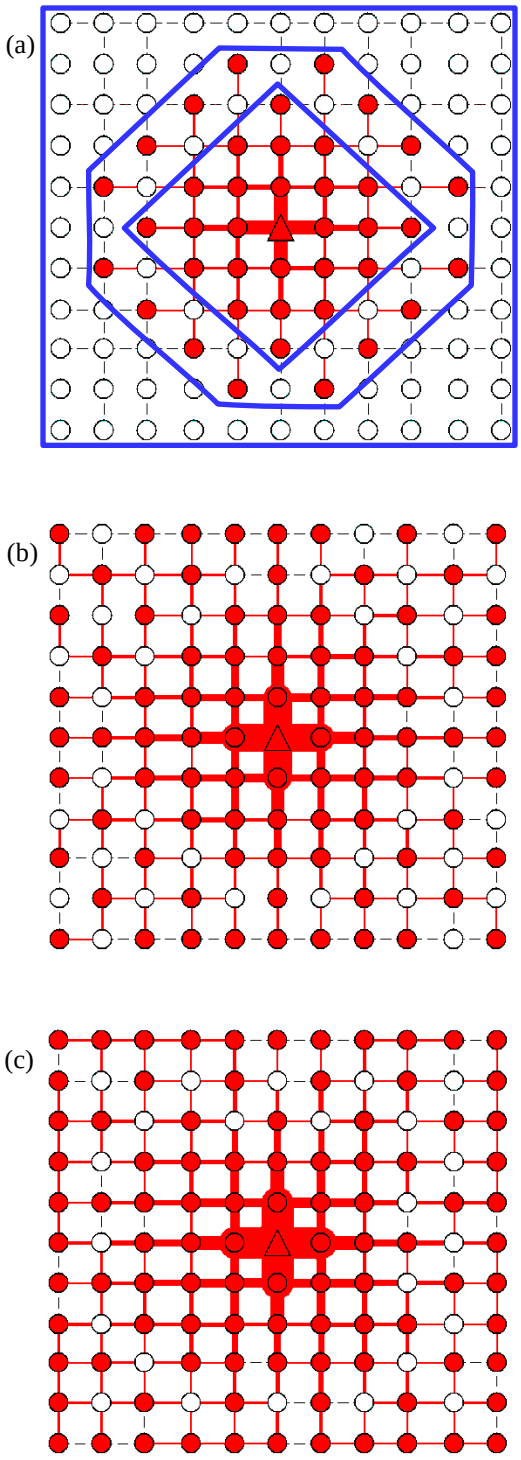

Figure 9. Three examples of optimized node configuration on a square lattice of $N = 121$ at different values of U and J. The blue frames in (a) correspond to the active core (innermost), the active-idle band (in-between) and the inactive band (outermost) [39].

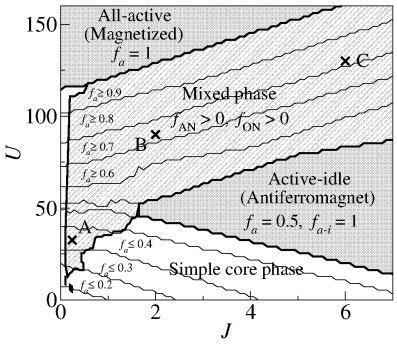

Figure 10. Phase diagram and the fraction of active nodes as a function of J and U on a square lattice with N = 121.The optimal configurations obtained at points A, B and C are shown in figures 9(a) – 9(c) respectively. f_a is the fraction of active nodes. f_{AN} is the fraction of active nodes with at least one active neighbor. f_{ON} is the fraction of nodes where all neighbors are in the opposite state. $f_{a\text{-}i}$ is the fraction of links which connect an active and an idle node [39].

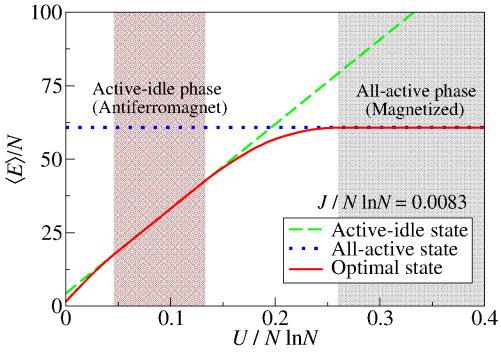

Figure 11. The dependence of energy on U of the all-active, the active-idle, and the optimal states of a square lattice with N = 121 and $J/N \ln N$ = 0.0083 [39].

6. Traffic Routing

In the family of network optimization problems, multipath optimization is among those with the broadest applications, ranging from public transport to Internet traffic, sensor networks, military convoy movements and journey planners. However, the difficulty lies in the requirement to simultaneously assigning the individual paths that affect each other while optimizing the global cost. This is similar to minimizing the energy of interacting polymers, and was recently used to derive a message-passing algorithm for routing [34]. The same result can also be derived from the cavity method.

Consider the routing of M passengers, each labeled with given source and destination on the network, and the cost function is $E = \Sigma_i \phi(\lambda_i)$, where λ_i is the fraction of passengers passing through node i.

Let the cavity energies be $A^{\nu}_{j\to i}$ when passenger ν routes from node j to i, $B^{\nu}_{j\to i}$ when she routes from node i to j, and $N^{\nu}_{j\to i}$ when she does not route between i and j. As schematically shown in figure 12, the recursion relations can be written as

$$A^{\nu}_{j\to i} = \min_{\{k|k\in\partial j\setminus i\}}\left[\phi\left(\lambda_j + \frac{1}{M}\right) + A^{\nu}_{k\to j} + \sum_{l\in\partial j\setminus ik} N^{\nu}_{l\to j}\right], \tag{17a}$$

$$B^{\nu}_{j\to i} = \min_{\{k|k\in\partial j\setminus i\}}\left[\phi\left(\lambda_j + \frac{1}{M}\right) + B^{\nu}_{k\to j} + \sum_{l\in\partial j\setminus ik} N^{\nu}_{l\to j}\right], \tag{17b}$$

$$N^{\nu}_{j\to i} = \min\left[\min_{\{k,l|k,l\in\partial j\setminus i\}}\left(\phi\left(\lambda_j + \frac{1}{M}\right) + A^{\nu}_{k\to j} + B^{\nu}_{l\to j} + \sum_{m\in\partial j\setminus ikl} N^{\nu}_{m\to j}\right), \phi(\lambda_j) + \sum_{k\in\partial j\setminus i} N^{\nu}_{k\to j}\right]. \tag{17c}$$

To determine the route of passenger ν, it suffices to provide the values of the cavity energies relative to $N^{\nu}_{j\to i}$. Hence the two messages to be passed from node j to i for passenger ν are

$$a^{\nu}_{j\to i} = \min_{\{k|k\in\partial j\setminus i\}}\left(a^{\nu}_{k\to j}\right) - c^{\nu}_{j\to i}, \tag{18a}$$

$$b^{\nu}_{j\to i} = \min_{\{k|k\in\partial j\setminus i\}}\left(b^{\nu}_{k\to j}\right) - c^{\nu}_{j\to i}, \tag{18b}$$

where $c^{\nu}_{j\to i} = \min\left[\min_{\{k,l|k,l\in\partial j\setminus i\}}\left(a^{\nu}_{k\to j} + b^{\nu}_{k\to j}\right), -\phi'(\lambda_j)\right]$

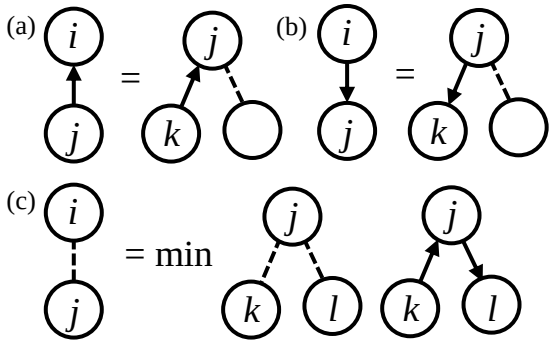

Figure 12. The recursion relations for $A^{\nu}_{j\to i}$, $B^{\nu}_{j\to i}$ and $N^{\nu}_{j\to i}$.

We apply the algorithm on the London underground network based on real passenger source-destination data obtained from the Oyster card system. The cost function is $E = \sum_i \lambda_i^2$, considered to be a measure of congestion, since it is proportional to the average crowd size a passenger will meet along her journey. Comparing with the commonly used Dijkstra algorithm [12], the cost is reduced by 20%, with only a slight increase in the average path length by 5.8%. Comparing with the state-of-the-art

congestion-aware algorithm, the cost is reduced by 0.7%, with an increase in the average path length by 0.7% [29].

7. Power Grids

The stability and robustness of power grids received renewed interest with the advent of renewable energy such as wind and solar power. These energy sources are much more intermittent and volatile, and the power grid needs to cope with fluctuations in supplies and demand. If these fluctuations lead to node or link failures, they can cascade throughout a large area of the network [24,3].

In a typical situation, the power distributor needs to consider the pattern of distribution in the next 15 to 60 minutes. Hence distribution algorithms need to make decisions based on probabilistic predictions of future supply and demand. Recently, a message-passing approach was proposed to deal with the probabilistic nature of the problem [17]. This can be done by considering equation (5), but noting that Λ_j is a fluctuating quantity. Averaging over Λ_j, we have

$$E_{j\to i}(y_{ij}) = \phi(y_{ij}) + \sum_{k\in\partial j\setminus i} \frac{\left\langle \mu_{ij}^2 \right\rangle_{\Lambda_j} - A_{jk}^2}{2B_{jk}}, \qquad (19)$$

where

$$\mu_{ij} = \left\langle \min\left[\frac{\sum_{k\in\partial j\setminus i}(y_{jk} - B_{jk}^{-1}A_{jk}) - y_{ij} + \Lambda_j}{\sum_{k\in\partial j\setminus i} B_{jk}^{-1}}, 0 \right] \right\rangle_{\Lambda_j}. \qquad (20)$$

For Λ_j obeying a Gaussian distribution with mean $\langle\Lambda_j\rangle$ and σ_j^2, analytical expressions can be obtained, albeit a bit tedious to be presented in this review. Nevertheless, the method can efficiently allocate extra resources to cope with fluctuating demands. Ongoing research work will continue to address more issues in this family of problems.

8. Conclusion

We have considered how message-passing algorithms can be applied to different useful problems in transportation and logistics networks. Starting from the fundamental and typical problem of optimizing transportation costs during the allocation of resources from source nodes to consumer nodes, the algorithm can be extended to various applications. In dealing with bandwidth constraints, we found interesting effects such as the bottleneck effect and the appearance of clusters of balance nodes. The problem can also be combined with the decision of optimal source locations, arriving at a problem involving both continuous and discrete variables. This leads to pictures of complex energy landscapes that involve multiple optimal solutions, corresponding to the so-called replica symmetry-breaking solutions. To deal with the facility location problem, one can also include redundancy costs that take into account the need to increase coverage. Generalizing from unlabeled traffic to multi-class labeled traffic (that is, traffic with individual components specifying source and consumer nodes), the algorithm can be applied to multi-passenger routing. By including uncertainties in supply and demand, it can be used in pre-emptive control in power grids.

These studies show that the message-passing technique has the advantages of being decentralized, having lower computational complexity, lower communication overhead, and faster response to local changes. Although it is derived by assuming a sparse network structure, it yields exact results when the algorithm converges. The cavity method also enables us to introduce interesting analysis.

Another distributed approach that we have introduced is the chemical potential algorithm. Chemical potentials arise from the Lagrange multipliers of the conservation of resources. Since they can be interpreted as storage prices of resources at the nodes [19], they have the potential to be applied to soft control of traffic and logistics through pricing policies.

The above decentralized approaches are expected to be applicable to many problems that can be formulated in terms of network structures. Besides transportation and logistics networks, they can also be useful in areas such as engineering, biology, economics and social science.

Acknowledgements

This work is supported by the EPSRC (Industrial CASE Studentship 12330048) and the Research Grants Council of Hong Kong (grant number 605813).

References

[1] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Topology of the Fittest Transportation Network, Phys. Rev. Lett. **84**, 4745 (2000).

[2] A. L. Barabási and R. Albert, Science **286**, 509 (1999).

[3] A. Bernstein, D, Bienstock, D, Hay, M. Uzunoglu, and G. Zussman, Clumbia University,

Electrical Engineering Technical Report #2011-05-06 (2011).
[4] D. Bertsekas, Linear Network Optimization (MIT Press, Cambridge MA, 1991).
[5] D. Bertsekas and R. Gallager, Data Networks (Prentice Hall, Englewood Cliffs, NJ, 1992).
[6] S. Boettcher and A. G. Percus, Phys. Rev. Lett. **86**, 5211 (2001).
[7] S. Bohn and M. O. Magnasco, Structure, Scaling, and Phase transition in the Optimal Transport Network, Phys. Rev. Lett. **98**, 088702 (2007).
[8] S. Bounkong, J. van Mourik, and D. Saad, Phys. Rev. E **74**, 057101 (2006).
[9] 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.
[10] J. R. L. de Almeida and D. J. Thouless, J. Phys. A: Math. Gen. **11**, 983 (1978).
[11] R. L. Devaney, An Introduction to Chaotic Dynamical Systems (Redwood City, CA: Addison-Wesley, 1989).
[12] E. W. Dijkstra, Numerische Mathematik **1**, 269 (1959).
[13] P. G. Doyle and J. L. Snell, Random Walks and Electric Networks (Mathematical Association of America, 1984).
[14] M. Durand, Structure of Optimal Transport Networks Subject to a Global Constraint, Phys. Rev. Lett. **98**, 088701 (2007).
[15] B. J. Frey, Graphical Models for Machine Learning and Digital Communication (MIT Press, Cambridge, MA, 1998).
[16] E. Harrison, D. Saad, and K. Y. M. Wong, 2nd International Conference on Intelligent Green Building and Smart Grid (IGBSG 2016), Prague, Czech Republic, 27-29 June 2016 (2016).
[17] E. Harrison, D. Saad, and K. Y. M. Wong, Int. J. of Smart Grid and Clean Energy, accepted (2016).
[18] H. Huang, J. Raymond and K. Y. M. Wong, J. Stat. Phys. **156**, 301 (2014).
[19] F. P. Kelly, Phil. Trans. R. Soc. Lond. A **337**, 343 (1991).
[20] S. Kirkpatrick and B. Selman, Science **264**, 1297 (1994).
[21] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, Phys. Rev. X **2**, 021005 (2012).
[22] M. Mézard, G. Parisi, and M. A. Virasoro, Spin Glass Theory and Beyond (World Scientific, Singapore, 1987).
[23] M. Mézard and R. Zecchina, Phys. Rev. E **66**, 056126 (2002).
[24] A. E. Motter and Y.-C. Lai, Phys. Rev. E **66**, 065102 (2002).
[25] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **89**, 268701 (2002).
[26] H. Nishimori, Statistical Physics of Spin Glasses and Information Processing (Oxford University Press, Oxford, 2001).
[27] R. T. Rockafellar, Network Flows and Monotropic Optimization (Wiley, 1984).
[28] B. Selman, H. Kautz, and B. Cohen, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 26 (Providence, RI: American Mathematical Society), 521 (1996).
[29] F. Shahrokhi and D. W. Matula, J. of ACM **37**, 318 (1990).
[30] Z. Shao and H. Zhou, Optimal Transportation Network with Concave Cost Functions: Loop Analysis and Algorithms, Phys. Rev. E **75**, 066112 (2007).
[31] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Phil. Mag. **35**, 593 (1977).
[32] K. Y. M. Wong and D. Saad, Phys. Rev. E. **74**, 010104(R) (2006).
[33] K. Y. M. Wong and D. Saad, Phys. Rev. E **76**, 011115 (2007).
[34] C. H. Yeung, D. Saad and K. Y. M. Wong, Proc. Natl. Acad. Sci. USA **110**, 13717-13722 (2013).
[35] C. H. Yeung and K. Y. M. Wong, Phys. Rev. E **80**, 021102 (2009).
[36] C. H. Yeung and K. Y. M. Wong, Stat. Mech., P03029 (2009).
[37] C. H. Yeung and K. Y. M. Wong, Europhys. J. B **74**, 227 (2010).
[38] C. H. Yeung and K. Y. M. Wong, J. Stat. Mech., P04017 (2010).
[39] C. H. Yeung, K. Y. M. Wong, and B. Li, Phys. Rev. E **89**, 062805 (2014).
[40] J. Zhou and H. Zhou, Phys. Rev. E **79**, 020103(R) (2009).
[41] 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.