Statistical Mechanics of Steiner trees

Statistical Mechanics of Steiner Trees

M. Bayati,[1] C. Borgs,[1] A. Braunstein,[2] J. Chayes,[1] A. Ramezanpour,[3] and R. Zecchina[2]

[1]*Microsoft Research, One Microsoft Way, 98052 Redmond, Washington, USA*
[2]*Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*
[3]*ICTP, Strada Costiera 11, I-34100 Trieste, Italy*

(Received 5 May 2008; published 18 July 2008)

The minimum weight Steiner tree (MST) is an important combinatorial optimization problem over networks that has applications in a wide range of fields. Here we discuss a general technique to translate the imposed global connectivity constrain into many local ones that can be analyzed with cavity equation techniques. This approach leads to a new optimization algorithm for MST and allows us to analyze the statistical mechanics properties of MST on random graphs of various types.

DOI: 10.1103/PhysRevLett.101.037208 PACS numbers: 64.60.aq, 05.20.−y, 87.18.Vf, 89.70.Eg

Given a graph or a lattice, finding a subgraph that optimizes some global cost function is an important problem in many fields. One of the most basic versions of this is known as the minimum weight Steiner tree (MST) problem.

Given an undirected graph with positive weights on the edges, the MST problem consists in finding a connected subgraph of minimum weight that contains a selected set of ''terminal'' vertices. Such construction may require the inclusion of some nonterminal nodes which are called Steiner nodes. Clearly, an optimal subgraph must be a tree. Solving MST is a key component of many optimization problems involving real networks. Concrete examples are network reconstruction in biology (phylogenetic trees and regulatory subnetworks), Internet multicasting, circuit design, and power or water distribution networks design, just to mention few famous ones. MST is also a beautiful mathematical problem in itself which lies at the root of computer science being both NP complete [1] and difficult to approximate [2]. In physics the Steiner tree problem has similarities with many basic models such as polymers, self-avoiding walks, or transport networks (e.g., [3]) with a nontrivial interplay between local an global frustration.

Here we show that the cavity approach of statistical physics can be used to both analyze and solve this problem on random graphs (as, e.g., [4–6]) once an appropriate representation is chosen. We actually study the even more general (and eventually harder) D-MST problem in which we consider the depth of the tree from a root terminal node to be bounded by D. Unfortunately the traditional techniques for studying topologically connected structures, as for instance the so-called $O(n)$ model, are incompatible with the cavity method. We provide here instead an arborescent representation of the Steiner problem which allows us to implement explicitly global connectivity constraints in terms of local ones.

In recent years many algorithmic results have appeared showing the efficacy of the cavity approach for optimization and inference problems defined over both sparse and dense random networks of constraints [4–9]. These performances are understood in terms of factorization properties of the Gibbs measure over ground states, which can be also seen as the onset of correlation decay along the iterations of the cavity equations [10]. Here we make a step further by presenting evidence for the exactness of the cavity approach for a qualitatively different class of models, namely, problems which are subject to rigid global constraints that couple all variables. Quite often this type of global constraint is of topological origin and is common to many problems across disciplines [e.g., the traveling salesman problem (TSP) in computer science or self-avioding walks in physics].

Our work addresses two questions: by analyzing the distributional equations we provide the phase diagrams of the problem in the control parameters α and D, where αN is the number of terminals in a graph of N vertices and D is the allowed depth of the tree from a randomly chosen root. We compute quantities like the behavior of the minimum cost as a function of D for a given fraction α of terminals, or the number of Steiner nodes cN^s where both c and the exponent s depend on D and α. Such quantities are of extreme interest in that they are directly connected with the topology of the tree. For instance, for the case of complete graphs with random weights we find that an extremely small depth D_N is sufficient for reaching costs which are close to optimal ones for the unbounded trees [e.g., for the complete graph with random weights we find that $D_N \sim \log\log N$ is sufficient to reach asymptotically a cost close to the optimal one $\zeta(3)$ [11,12] of the minimum spanning tree which has depth $\Theta(N^{1/3})$ [13]]. For finite D the results of the cavity approach can be compared with rigorous upper and lower bounds [14] making us conjecture that the cavity approach is exact, as it happens for random matchings [15]. Similar results hold for other classes of random graphs. Here we give results for fixed degree and scale-free graphs, for which some nontrivial patterns of solutions for optimal Steiner trees appear.

On the algorithmic side, the arborescent representation of the problem leads to cavity equations that can be turned into an algorithm for solving single instances.

Very few results are known on the Steiner problem on random graphs in the regime in which α is finite. For the complete graph with random weights some upper and lower bounds for the minimum cost have been derived [16], which are compatible with those predicted by the cavity method. For finite degree random graphs (e.g., Erds-Rnyi, fixed degree or scale-free graphs) much less is known.

The model.—We model the Steiner tree problem as a rooted tree (such a construction is often associated with the term ''arborescence''). Each node i is endowed with a pair of variables (p_i, d_i), a pointer p_i to some other node in the neighborhood $V(i)$ of i and a depth $d_i \in \{1, \ldots, D\}$ defined as the distance from the root. Terminal nodes must point to some other node in the final tree and hence $p_i \in V(i)$. The root node conventionally points to itself. Nonroot nodes either point to some other node in $V(i)$ if they are part of the tree (*Steiner* and *terminal* nodes) or just do not point to any node if they are not part of the tree (allowed only for nonterminals), a fact that we represent by allowing for an extra state for the pointer $p_i \in V(i) \cup \varnothing$. The depth of the root is set to zero, $d_i = 0$ while for the other nodes in the tree the depths measure the distance from the root along the unique oriented path from the node to the root.

To impose the global connectivity constraint for the tree we need to impose the condition that if $p_i = j$ then $p_j \neq \varnothing$ and $d_j = d_i - 1$. This condition forbids loops and guarantees that the pointers describe a tree. In building the cavity equations (or the belief propagation equations), we need to introduce the characteristic functions f_{ij} which impose such constraints over configurations of the independent variables (p_i, d_i). For any edge (i, j) we have the indicator function $f_{ij} = g_{ij} g_{ji}$ where $g_{jk} = [1 - \delta_{p_k,j}(1 - \delta_{d_j,d_k-1})](1 - \delta_{p_k,j}\delta_{p_j,\varnothing})$.

Cavity equations.—The cavity equations take the form

$$P_{j\to i}(d_j, p_j) \propto e^{-\beta c_{jp_j}} \prod_{k\in j\backslash i} Q_{k\to j}(d_j, p_j) \tag{1}$$

$$Q_{k\to j}(d_j, p_j) \propto \sum_{d_k p_k} P_{k\to j}(d_k, p_k) f_{jk}(d_k, p_k, d_j, p_j) \tag{2}$$

where c_{ij} is the weight of the link (i, j), with $c_{i\varnothing} = \infty$ if i is a terminal. The $\propto$ symbol accounts for a multiplicative normalization constant. Allowed configurations are weighted by $e^{-\beta c_{ij}}$ where β^{-1} is a temperature fixing the energy level. The zero temperature limit is taken by considering the following change of variables: $\psi_{j\to i}(d_j, p_j) = \beta^{-1} \log P_{j\to i}(d_j, p_j)$ and $\phi_{k\to j}(d_j, p_j) = \beta^{-1} \log Q_{k\to j}(d_j, p_j)$. In the $\beta \to \infty$ limit Eqs. (1) and (2) reduce to

$$\psi_{j\to i}(d_j, p_j) = -c_{jp_j} + \sum_{k\in j\backslash i} \phi_{k\to j}(d_j, p_j), \tag{3}$$

$$\phi_{k\to j}(d_j, p_j) = \max_{d_k, p_k:\, f_{jk}(d_k,p_k,d_j,p_j)\neq 0} \psi_{k\to j}(d_k, p_k). \tag{4}$$

The previous two equalities must be understood to hold except for an additive constant. Equations (3) and (4) are in the so-called ''max sum'' form.

On a fixed point, one can compute *marginals* ψ_j:

$$\psi_j(d_j, p_j) = -c_{jp_j} + \sum_{k\in j} \phi_{k\to j}(d_j, p_j) \tag{5}$$

and the optimum tree should be given by $\arg\max \psi_j$.

If the starting graph is a tree $\psi_{j\to i}(d_j, p_j)$ can be interpreted as the minimum cost change of removing a vertex j with forced configuration d_j, p_j from the subgraph with link (i, j) already removed. We introduce the variables $A^d_{k\to j} \equiv \max_{p_k\neq j,\varnothing} \psi_{k\to j}(d, p_k)$, $B^d_{k\to j} \equiv \psi_{k\to j}(d, \varnothing)$, $C^d_{k\to j} \equiv \psi_{k\to j}(d, j)$, $D_{k\to j} = \max_d \max\{A^d_{k\to j}, B^d_{k\to j}\}$, and $E^d_{k\to j} = \max\{C^{d+1}_{k\to j}, D_{k\to j}\}$. This is enough to compute $\phi_{k\to j}(d_j, p_j) = A^{d_j-1}_{k\to j}$, $D_{k\to j}$, $E^{d_j}_{k\to j}$ for $p_j = k$, $p_j = \varnothing$, and $p_j \neq k, \varnothing$ respectively. Equations (3) and (4) can then be solved by repeated iteration of the following set of equations:

$$A^d_{j\to i}(t+1) = \sum_{k\in j\backslash i} E^d_{k\to j}(t) + \max_{k\in j\backslash i}\{A^{d-1}_{k\to j}(t) - E^d_{k\to j}(t) - c_{jk}\}, \tag{6}$$

$$B_{j\to i}(t+1) = -c_{j\varnothing} + \sum_{k\in j\backslash i} D_{k\to j}(t), \tag{7}$$

$$C^d_{j\to i}(t+1) = -c_{ij} + \sum_{k\in j\backslash i} E^d_{k\to j}(t), \tag{8}$$

$$D_{j\to i}(t) = \max[\max_d A^d_{j\to i}(t), B_{j\to i}(t)], \tag{9}$$

$$E^d_{j\to i}(t) = \max[C^{d+1}_{j\to i}(t), D_{j\to i}(t)]. \tag{10}$$

For graphs without cycles the above equations are guaranteed to converge to the optimal solution. In graphs with cycles, these equations may instead fail to converge in some cases. For the classes of random graphs studied in this work, this appears not to be due to a replica symmetry breaking instability but rather to the effect of local structures in the underlying graph (as it is known to happen in simpler problems such as random matchings [17]). This observation is corroborated by the analysis of the distributional cavity equations discussed later. While more work is needed to understand this point, from the algorithmic viewpoint the problem can be overcome by applying a small perturbation [6]. The term $\psi_j(d_j, p_j)$ of Eq. (5) multiplied by a (small) constant ρ is added to the right-hand side of Eq. (3). This leads to a set of equations which show good convergence properties for vanishing ρ.

An equivalent formulation of the problem can be constructed by introducing a link representation of the pointer variables (one may introduce link variables $x_{ij} = 0, \pm 1, 0$ if i does not point j, 1 if i points j, and -1 if j points i). In this representation, the number of states of the independent

variables is just $3D$, which can be kept finite for complete graphs or at most of order $\log N$ for sparse graphs.

Distributional equations and average case analysis.— Population dynamics (or density evolution) is a powerful tool to solve distributional equations that deal with a large number of random variables. In the physics community the method was introduced in [18] for the study of spin glass models on diluted random graphs. Population dynamics is useful especially when the equations involve sums over many states of the variables. The underlying idea is to represent probability distributions with a population of random variables and use the equations to update such populations. After a suitably large number of updates the histogram of variables in the population will converge to a stable distribution.

To obtain results on the $N \to \infty$ limit one would need to rescale simultaneously all d-dependent quantities in order to eliminate their direct dependence on N in Eqs. (6)–(10). We limited, however, ourselves here for all cases analyzed to large but finite N, in particular, because the obviously needed dependence of D on N for finite degree graphs makes this task even more involved.

We will apply the population dynamics method to find the statistical properties of the cavity fields $M_{i\to j} = (A^d_{i\to j}, B_{i\to j}, C^d_{i\to j}, D_{i\to j}, E^d_{i\to j})$ in Eqs. (6)–(10). Given an ensemble of random graphs we will find the probability distribution of these fields from which we will derive the quantities of interest, namely, the average minimum cost and average number of Steiner nodes as a function of N, in the so-called Bethe approximation which is implicit in the cavity approach. The method proceeds by initializing at random a population of field vectors $M_a = (A^d_a, B_a, C^d_a, D_a, E^d_a)$ with $a \in [0, N_p]$ and $d \in [0, D]$. The first member M_0 represents messages sent by root. Members with label $a = 1, \ldots, N_t$ represent messages sent by terminal nodes. Here $N_t = \alpha N_p$ where $\alpha = K/N$ is the fraction of terminal nodes. Then the population dynamics algorithm works by updating the population using Eqs. (6)–(10) until convergence is reached. For brevity, we omit the details of this procedure. Once convergence is reached, marginals $\psi_a(d, p)$ can be computed using Eq. (5). The state (d^*, p^*) that maximizes the local marginal gives the energy contribution of the ath member. If $p^* \neq \varnothing$ and $N_t < a$, then a is a Steiner member. Finally the minimum cost reads $E = Ke_t + (N - K)e_s$ where e_t and e_s are the average energy of terminal and Steiner members. The fraction of Steiner members in the population will give the fraction of Steiner nodes in the ensemble of random graphs.

In Figs. 1–3 we display numerical results for three classes of random graphs, namely, complete graphs, finite connectivity random graphs, and scale-free graphs. We first verify a quite remarkable agreement between the output of the algorithm which finds Steiner trees on given random instances with the outcomes of the population dynamics averaged over the randomness. In Figs. 1 and 2, we estimate the dependence on the depth D of the minimum cost and of the size of the Steiner set nodes. For complete graph with random weights we are able to provide an accurate estimate of the scaling exponents which for $\alpha = 1$ are compatible with rational exponents predicted by rigorous analysis [14]. Moreover, we observe a very rapid decrease of the minimum cost with D, compatible with $N^{1/(2^D-1)}$. This suggests that very few "hops" ($\sim \log \log N$) are indeed sufficient to reach optimal costs. From a qualitative point of view we observe a nontrivial dependence on N and α of the size of the Steiner set. The size itself turns out to be sublinear, with a rational exponent that depends on D.

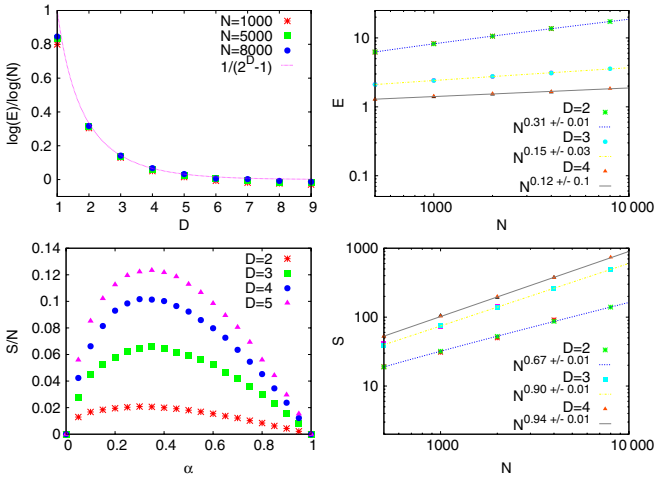

FIG. 1 (color online). D-MST on complete graphs. Left: Minimum cost (at $\alpha = 0.5$) and fraction of Steiner nodes (for $N = 8000$) as a function of D. Right: Comparison of population dynamics with the algorithm on single samples for various values of N at $\alpha = 0.5$. Fits are in very good agreement with known bounds.

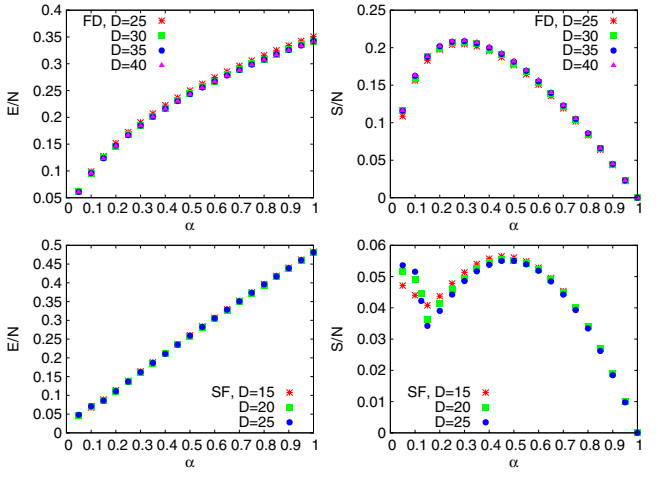

FIG. 2 (color online). Fixed degree (FD) and scale-free (SF) graphs. Left: Minimum cost as function of α for different values of D. Right: Fraction of Steiner nodes as a function of α. The FD graphs have degree $C = 3$ and size $N = 10^6$. The SF graphs have exponent $\gamma = 3$ and size $N = 10^4$.

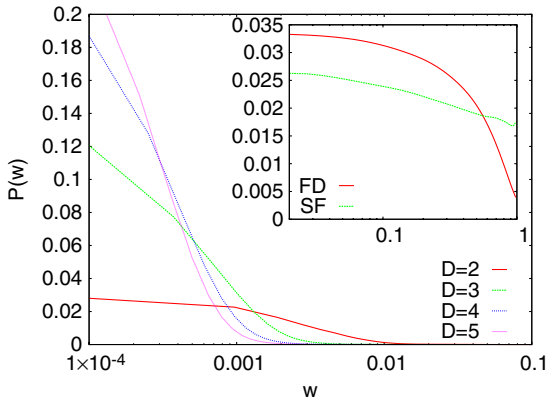

FIG. 3 (color online). Weight distribution of the MST for complete graphs of size $N = 8000$ at $\alpha = 0.5$. Inset: For FD graphs of degree $C = 3$ ($N = 10^6$) and SF graphs of exponent $\gamma = 3$ ($N = 10^4$) with parameters $D = 25$, $\alpha = 0.5$.

For fixed N there appears a maximum for relatively small values of α. For the scale-free graphs there appears an additional cuspidlike minimum. Finally, in Fig. 3 we provide the probability distribution of optimal weights for all classes.

We conclude this Letter by mentioning the connection with rigorous results. For the case of bounded depth trees on complete graphs our numerical results show that the cavity equations are indeed consistent with known bounds. As discussed in [14], the analysis of a simple greedy algorithm and a Chernoff-type bound lead to upper and lower bounds for the minimum cost that are able to identify the exact scaling exponent and to give bounds for the prefactors. More precisely, it can be shown that the average minimum E_D grows with the size as $N^{1/(2^D-1)}$. The case $D = 2$ and $\alpha = 1$ is particularly easy to understand: the greedy algorithm amounts to choosing a first set of N_1 nodes at depth 1 by selecting the N_1 links with smallest weights. Successively the remaining $N - N_1$ nodes at depth 2 are connected to the first layer by choosing the smallest weight for each node. By optimizing over the size of N_1 one finds for the average minimum cost $E_2 = \frac{3}{2}N^{1/3}$ (a naive guess may give an exponent $1/2$ instead of $1/3$). Comparisons with the cavity approach for small D show that indeed the exponent is $1/(2^D - 1)$ as it should and that there exist a constant additional (negative) term to the minimum cost which improves over the greedy algorithm. Table I shows the results of a power law fit to our data for the average minimum cost and number of Steiner nodes as a function of N. For $D = N - 1$ and $\alpha = 1$ it is possible to prove using techniques based on the computation tree that if the BP equations converge, then the result is optimal. Details about these results and hopefully about their extensions to the $\alpha < 1$ case will be given elsewhere. Work is in progress to apply the algorithmic scheme we have presented to clustering, network reconstruction, and protein pathways identification problems.

TABLE I. Comparing the exponents and prefactors for complete graphs. The parameters have been obtained by fitting data to $a + bx^c$. In all the data $N \leq 8000$. Values in the parenthesis are known analytical results.

	D	α	a	b	c
E	2	0.5	-1.07 ± 0.07	0.92 ± 0.01	0.31 ± 0.01
S	2	0.5	-3.62 ± 0.13	0.35 ± 0.01	0.67 ± 0.01
E	3	0.5	-0.83 ± 0.05	1.21 ± 0.02	0.15 ± 0.03
S	3	0.5	0	0.14 ± 0.01	0.90 ± 0.01
E	2	1	-1.46 ± 0.25	$1.47 \pm 0.03(3/2)$	$0.35 \pm 0.01(1/3)$
E	3	1	-0.95 ± 0.05	1.75 ± 0.02	$0.15 \pm 0.02(1/7)$

[1] R. Karp, *Complexity of Computer Computations*, IBM Research Symposia Series Vol. 43, edited by R. E. Miller and J. W. Thatcher (Plenum Press, New York, 1972), p. 85.
[2] G. Robins and A. Zelikovsky, in *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms* (SIAM, San Francisco, 2000), pp. 770–779.
[3] M. Durand Phys. Rev. Lett. **98**, 088701 (2007).
[4] A. Braunstein, M. Mzard, and R. Zecchina, Random Struct. Algorithms **27**, 201 (2005).
[5] M. Mezard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[6] A. Braunstein and R. Zecchina, Phys. Rev. Lett. **96**, 030201 (2006).
[7] B. Frey and D. Dueck, Science **315**, 972 (2007).
[8] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. E **68**, 036702 (2003).
[9] C. Di, A. Montanari, and R. Urbanke, in *Proceedings of the International Symposium on Information Theory* (ISIT, Chicago, 2004).
[10] F. Krzakal, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborova, Proc. Natl. Acad. Sci. U.S.A. **104**, 10 318 (2007).
[11] A. Frieze, Discrete Appl. Math. **10**, 47 (1985).
[12] D. Gamarnik, in *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms* (SIAM, Vancouver, 2005), pp. 700–704.
[13] L. Addario-Berry, N. Broutin, and B. Reed, in *Proceedings, Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities* (DMTCS, Nancy, 2006).
[14] O. Angel, A. Flaxman, D. Wilson, and R. Zecchina, "Bounded Depth Minimum Steiner Trees" (unpublished).
[15] D. Aldous, Random Struct. Algorithms **18**, 381 (2001).
[16] B. Bollobs, D. Gamarnik, O. Riordan, and B. Sudakov, Combinatorica **24**, 187 (2004).
[17] M. Mezard and G. Parisi, Europhys. Lett. **2**, 913 (1986).
[18] M. Mezard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).