Statistical Mechanics of Steiner trees

Original

Availability:
This version is available at: 11583/1841213 since:

Publisher:
American Physical Society (APS)

Published
DOI:10.1103/PhysRevLett.101.037208

Terms of use:
openAccess
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)
Statistical Mechanics of Steiner Trees

M. Bayati,1 C. Borgs,1 A. Braunstein,2 J. Chayes,1 A. Ramezanpour,3 and R. Zecchina2

1Microsoft Research, One Microsoft Way, 98052 Redmond, Washington, USA
2Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
3ICTP, 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-avoiding 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 \( \alpha \) and \( D \), where \( \alpha N \) 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 \( \alpha \) of terminals, or the number of Steiner nodes \( c N^s \) where both \( c \) and the exponent \( s \) depend on \( D \) and \( \alpha \). 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 \( 3(3) \) [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 \( \alpha \) 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.,
Erdős-Rényi, 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 \( p_i \in V(i) \cup \emptyset \). 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 \( p_j \neq \emptyset \)
and \( d_j = d_i - 1 \). This condition forbids loops and guar-
antees 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
\( g_{jk} = (1 - \delta_{p_i,j}(1 - \delta_{d_i,d_j - 1}))(1 - \delta_{p_j,i}\delta_{p_j,\emptyset}) \).

Cavity equations.—The cavity equations take the form
\[
P_{j\rightarrow i}(d_j, p_j) \propto e^{-\beta c_{ij}} \prod_{k \in \partial i} Q_{k\rightarrow j}(d_j, p_j)
\]  
(1)
and
\[
Q_{k\rightarrow j}(d_j, p_j) \propto \sum_{d_k, p_k} P_{k\rightarrow j}(d_k, p_k) f_{jk}(d_k, p_k, d_j, p_j)
\]  
(2)
where \( c_{ij} \) is the weight of the link \((i, j)\), with \( c_{i\emptyset} = \infty \)
if \( i \) is a terminal. The \( \propto \) symbol accounts for a mul-
tiplicative normalization constant. Allowed conﬁgurations
are weighted by \( e^{-\beta c_{ij}} \) where \( \beta^{-1} \) is a temperature
ﬁxing the energy level. The zero temperature limit is
taken by considering the following change of variables:
\[ \psi_{j\rightarrow i}(d_j, p_j) = \beta^{-1} \log P_{j\rightarrow i}(d_j, p_j) \]
and
\[ \phi_{k\rightarrow j}(d_j, p_j) = \beta^{-1} \log Q_{k\rightarrow j}(d_j, p_j). \]
In the \( \beta \to \infty \) limit Eqs. (1) and (2) reduce to
\[ \psi_{j\rightarrow i}(d_j, p_j) = -c_{ij} + \sum_{k \in \partial i} \phi_{k\rightarrow j}(d_k, p_k), \]
(3)
and
\[ \phi_{k\rightarrow j}(d_j, p_j) = \max_{d_k, p_k; f_{ji}(d_k, p_k, d_j, p_j) \neq 0} \psi_{k\rightarrow j}(d_k, p_k). \]
(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 \( \psi_j \):
\[ \psi_j(d_j, p_j) = -c_{jp_j} + \sum_{k \in \partial j} \phi_{k\rightarrow j}(d_j, p_j) \]
(5)
and the optimum tree should be given by \( \arg \max \psi_j \).

If the starting graph is a tree \( \psi_{j\rightarrow i}(d_j, p_j) \) can be inter-
preted 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
\[ A_{k\rightarrow j}^d = \max_{p_i \neq j, \emptyset} \psi_{k\rightarrow j}(d, p_k) \]
\[ B_{k\rightarrow j}^d = \psi_{k\rightarrow j}(d, \emptyset) \]
\[ C_{k\rightarrow j}^d = \psi_{k\rightarrow j}(d, j) \]
\[ D_{k\rightarrow j} = \max_{d} \max_{j} \{A_{k\rightarrow j}^d, B_{k\rightarrow j}^d\} \]
\[ E_{k\rightarrow j}^d = \max \{C_{k\rightarrow j}^d, D_{k\rightarrow j}\}. \]
This is enough to compute \( \phi_{k\rightarrow j}(d_j, p_j) = A_{k\rightarrow j}^d, D_{k\rightarrow j} \) for
\( p_j = k, p_j = \emptyset \), and \( p_j \neq k, \emptyset \) respectively. Equations (3) and (4)
can then be solved by repeated iteration of the following set
of equations:
\[ A_{j\rightarrow i}^d(t + 1) = \sum_{k \in \partial j} E_{k\rightarrow j}^d(t) + \max_{d} \{A_{k\rightarrow j}^{d-1}(t) - E_{k\rightarrow j}^d(t) - c_{jk}\} \]
(6)
\[ B_{j\rightarrow i}(t + 1) = -c_{j\emptyset} + \sum_{k \in \partial j} D_{k\rightarrow j}(t) \]
(7)
\[ C_{j\rightarrow i}(t + 1) = -c_{ij} + \sum_{k \in \partial j} E_{k\rightarrow j}^d(t) \]
(8)
\[ D_{j\rightarrow i}(t) = \max_{d} \{A_{d\rightarrow j}^d(t), B_{j\rightarrow i}(t)\} \]
(9)
\[ E_{j\rightarrow i}^d(t) = \max \{C_{j\rightarrow i}^d(t), D_{j\rightarrow i}(t)\}. \]
(10)

For graphs without cycles the above equations are guar-
anteed 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 struc-
tures 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 distri-
butional cavity equations discussed later. While more work
is needed to understand this point, from the algorithmic view-
point 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 \( \rho \) is added to the right-hand side of
Eq. (3). This leads to a set of equations which show good
convergence properties for vanishing \( \rho \).

An equivalent formulation of the problem can be con-
structed 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 $M_{i \to j} = (A_{i \to j}, B_{i \to j}, C_{i \to j}^d, D_{i \to j}, E_{i \to j}^d)$ 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 ensemble of random graphs we will find the probability distribution from which we will derive the ensemble of random graphs. For complete graphs this means that the cavity fields are sublinear, with a rational exponent that depends on $\alpha$.

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 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 $\alpha$ for different values of $D$. Right: Fraction of Steiner nodes as a function of $\alpha$. The FD graphs have degree $C = 3$ and size $N = 10^6$. The SF graphs have exponent $\gamma = 3$ and size $N = 10^4$. 

For fixed $N$ there appears a maximum for relatively small values of $\alpha$. 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 $E_2 = \frac{2}{3}N^{1/3} \alpha$ (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>
<thead>
<tr>
<th>$D$ $\alpha$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$ 2 0.5</td>
<td>$-1.07 \pm 0.07$</td>
<td>$0.92 \pm 0.01$</td>
<td>$0.31 \pm 0.01$</td>
</tr>
<tr>
<td>$S$ 2 0.5</td>
<td>$-3.62 \pm 0.13$</td>
<td>$0.35 \pm 0.01$</td>
<td>$0.67 \pm 0.01$</td>
</tr>
<tr>
<td>$E$ 3 0.5</td>
<td>$-0.83 \pm 0.05$</td>
<td>$1.21 \pm 0.02$</td>
<td>$0.15 \pm 0.03$</td>
</tr>
<tr>
<td>$S$ 3 0.5</td>
<td>0</td>
<td>$0.14 \pm 0.01$</td>
<td>$0.90 \pm 0.01$</td>
</tr>
<tr>
<td>$E$ 2 1</td>
<td>$-1.46 \pm 0.25$</td>
<td>$1.47 \pm 0.03(3/2)$</td>
<td>$0.35 \pm 0.01(1/3)$</td>
</tr>
<tr>
<td>$E$ 3 1</td>
<td>$-0.95 \pm 0.05$</td>
<td>$1.75 \pm 0.02$</td>
<td>$0.15 \pm 0.02(1/7)$</td>
</tr>
</tbody>
</table>

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.