Statistical estimation for optimization problems on graphs

Mikhail Langovoy and Suvrit Sra
Max Planck Institute for Intelligent Systems
72076 Tübingen, Germany
{langovoy, suvrit.sra}@tuebingen.mpg.de

Abstract

Large graphs abound in machine learning, data mining, and several related areas. A useful step towards analyzing such graphs is that of obtaining certain summary statistics—e.g., or the expected length of a shortest path between two nodes, or the expected weight of a minimum spanning tree of the graph, etc. These statistics provide insight into the structure of a graph, and they can help predict global properties of a graph. Motivated thus, we propose to study statistical properties of structured subgraphs (of a given graph), in particular, to estimate the expected objective function value of a combinatorial optimization problem over these subgraphs. The general task is very difficult, if not unsolvable; so for concreteness we describe a more specific statistical estimation problem based on spanning trees. We hope that our position paper encourages others to also study other types of graphical structures for which one can prove nontrivial statistical estimates.

1 Introduction

A cornucopia of applications in machine learning and related areas involve large-scale graphs. Towards analyzing such graphs a basic step is that of obtain certain summary statistics. For example, we might want to know what might be the expected length of a shortest path between two nodes, or what is the expected weight of an associated minimum spanning tree, etc. Such statistics provide insight into the global structure of a graph; and estimating them helps predict properties of the entire graph without having to actually look at the whole graph, a very practical scenario.

Our considerations stem from a classic paper of Frieze [1985], who studied the expected value of the weight of a minimum spanning tree (MST) of a complete graph on n nodes, with edge-weights distributed according to a common distribution function. For such graphs, Frieze obtained an explicit value for the expected weight of an MST as n tends to infinity. In subsequent years, his analysis has been refined and extended to cover more general graphs, and under different assumptions—see [Steele, 2002] and the references therein.

This precedent suggests that with increasing sizes, one can estimate statistical properties of various combinatorial structures on graphs. This statement brings us to the key challenge of this paper.

Problem 1 (**Statistics on graphical structures**). *Let $G_n = (V_n, E)$ be a graph with n vertices, and let $\mathcal{G}$ be a collection of certain "structured" subgraphs of G_n. Let $\varphi : \mathcal{G} \to \mathbb{R}_+$ be function that measures the "cost" of a subgraph in $\mathcal{G}$. As n tends to infinity, what can we say about the expected value $\mathbb{E}[\min_{g \in \mathcal{G}} \varphi(g)]$ of the minimum cost structure, and under what restrictions on the structures $\mathcal{G}$ and on the cost function φ?*

Our current paper is a position paper that advances Problem 1 as a key research question worthy of careful investigation. Admittedly, in general this problem is very difficult; perhaps too broad to be useful. But our ultimate aim is less to tackle the general problem and more to identify special classes

of *structures* and *cost functions*, for which we can make nontrivial statistical statements. We hope that this workshop paper stimulates discussion and also encourages others to study this problem.

2   Formulation

Let us now move onto a somewhat more formal treatment of Problem 1.

Let $G_n = (V_n, E_n)$ be a graph with n vertices and $|E_n|$ edges. For any vertex $v \in V_n$ we observe an integer-valued random variable $X_v : \Omega \to \{1, 2, \dots, k\}$ (on an appropriate probability space Ω). Here k is assumed to be finite, but otherwise unknown. We call the collection $\{X_v \,|\, v \in V_n\}$ a *random coloring* of the graph G_n.[1]

Assume that $\{X_v \,|\, v \in V_n\}$ form a collection of completely independent random variables and that they are identically distributed according to a distribution function F_X.

Further, suppose that for any pair of vertices $v_i, v_j \in V_n$ we observe a real-valued random variable $Y_{v_i, v_j} : \quad \Omega \to \mathbb{R}_+$, where $\{Y_{v_i, v_j} \,|\, v_i, v_j \in V_n\}$ form a collection of completely independent random variables that are identically distributed according to a distribution function F. It is also assumed for now that the X's and Y's are completely independent of each other. We will add extra assumptions on F and F_X below, when necessary.

Remark: The above assumption on edge weights corresponds to the case when G_n is a complete graph on n vertices. To study general graphs on n vertices, one has to consider only the reduced collection of random variables $\{Y_{v_i, v_j} \,|\, (v_i, v_j) \in E_n\}$.

Denote by $\mathcal{G} = \{G \,|\, G \subseteq G_n\}$, a collection of (structured) subgraphs of G_n. Fix a *cost function* $\varphi : \mathcal{G} \to \mathbb{R}_+$. Ultimately, we will be interested in the case when φ is a set function (over the sets of vertices or edges participating in the subgraphs characterized by $\mathcal{G}$).

Often, a more convenient and specific form of φ might be assumed, namely that for any $G = (V, E) \subseteq G_n$ there exists a decomposition

$$\varphi(G) = \varphi(\{(X_v, Y_e) \,|\, v \in V, e \in E\}) = \varphi_1(\{Y_e \,|\, e \in E\}) + \varphi_2(\{X_v \,|\, v \in V\}), \qquad (1)$$

where φ_1 is a cost function that depends only on the edges, and φ_2 depends only on the vertices.

For concreteness, we now focus on the following statistical goal: *estimate the average cost* (per vertex) of a minimal cost spanning tree of G_n—henceforth, φ-MST.[2] We want a computationally efficient procedure for this estimate, and the estimator itself must be consistent. Additionally, we also care about the corresponding rates of convergence.

Since the original graph G may be very large and a direct computation of the (minimal cost) spanning tree may be infeasible, we suggest computing an estimator based on a suitably constructed auxiliary graph that is much smaller than G, but exhibits similar statistical properties. To this end, we propose the following generic method.

Method

1. On the basis of the coloring $\{X_v \,|\, v \in V\}$, construct a suitable (possibly problem-dependent) $\sqrt{n}$-consistent estimate $\widehat{F_X}$ of the distribution function F_X.

2. Using the collection $\{Y_e \,|\, e \in E\}$, construct a suitable $\sqrt{|E|}$-consistent estimate $\widehat{F}$ of the edge weight distribution function F. Note that when G_n is a complete graph, already the standard empirical distribution function gives an n-consistent estimate.

3. Generate an auxiliary graph $G'_{d(n)} = (V'_{d(n)}, E'_{d(n)})$, having $d(n)$ vertices, where $d(n)$ is suitably chosen and satisfies the growth conditions

$$\lim_{n \to \infty} d(n) = \infty, \qquad \lim_{n \to \infty} \frac{d(n)}{n} = 0. \qquad (2)$$

---

[1]We can also consider infinite colorings based on $X_v : \Omega \to \mathbb{N}$; but for simplicity we study only finite ones.
[2]A similar discussion also applies to other problems such as shortest paths, cuts, etc.

4. Simulate i.i.d. random variables $\{X_v' \mid v' \in V_{d(n)}'\}$ and $\{Y_e' \mid e \in E_{d(n)}'\}$ from the distribution function estimates constructed at Steps 1 and 2 correspondingly. (Remark: for a complete graph G_n, we generate $G_{d(n)}'$ to be complete as well.)

5. Find the minimum φ-cost spanning tree $\mathrm{ST}(G_{d(n)}')$; compute $\varphi(\mathrm{ST}(G_{d(n)}'))$. (Remark: This step requires solution of a potentially hard discrete optimization problem.)

Based on the above generic method, we introduce the estimate:

$$\widehat{\mathrm{Avcost}}(G_n) \; := \; \frac{\varphi(\mathrm{ST}(G_{d(n)}'))}{d(n)} \; . \tag{3}$$

Processing the reduced graph $G_{d(n)}'$ is obviously much faster than processing G_n itself. But we need to theoretically characterize to what extent it is acceptable to process $G_{d(n)}'$. To that end, we attempt to investigate the following main questions:

- When is the above method consistent;
- What can be the rate of convergence of the estimator (3); and
- What is the computational complexity of the new method.

We show that the above estimation procedure has a highly nontrivial behavior. Statistical analysis remains nonetheless possible, but requires delicate results from discrete probability as well as novel statistical methods. We provide below theoretical justification of our approach for some basic cases.

3   Applications to special cases

Consider the case when $\varphi_2 \equiv 0$, and the spanning tree weight depends only on edge weights.[3] As before, let $\mathcal{G} = \{G \mid G \subseteq G_n\}$ be the chosen collection of (structured) subgraphs of G_n. Next, assume that for an arbitrary member $G = (V, E) \in \mathcal{G}$ the edge-cost set function φ_1, defined in (1), satisfies additionally

$$\varphi_1\big(\{Y_e \mid e \in E\}\big) \; = \; \varphi_1\Big(\sum_{e \in E} Y_e\Big), \tag{4}$$

and that φ_1 is continuous and nondecreasing. This includes for example the important class of submodular functions that can be expressed as nondecreasing concave functions of sums (see e.g., [Stobbe and Krause, 2010, Goel et al., 2010]).

Based on the assumptions (1) and (4), we can prove the following.

Proposition 2. *Let F be a distribution function that is continuously differentiable at 0, having $F(0) = 0$ and $F'(0) > 0$. Suppose that F has finite mean and variance. Assume that the cost function φ satisfies (1) and (4) with $\varphi_2 \equiv 0$. Then for a minimum spanning tree of the complete graph G_n it holds that*

$$\lim_{n \to \infty} \mathbb{E}_F \, \varphi(\mathrm{ST}(G_n)) \; = \; \varphi_1(\zeta(3)/F'(0)), \tag{5}$$

where ζ is the Riemann Zeta function. Moreover, for any $\varepsilon > 0$,

$$\lim_{n \to \infty} Pr(|\varphi(\mathrm{ST}(G_n)) - \zeta(3)/F'(0)| > \varepsilon) \; = \; 0 \,. \tag{6}$$

The proof uses results from [Frieze, 1985] and [Steele, 2002]. Using this proposition, we will prove a consistency theorem for our estimator (3) for the case of complete graphs[3] and a wide class of edge-dependent weight functions.

First, we need to introduce a special class of estimators.

---

[3]We alert the reader to the fact that analysis of just the expected weight of an ordinary (linear) MST for general graphs is a difficult problem [Steele, 2002, Frieze et al., 2000].

Definition 1 (**Boundary respecting estimators**). As above, we assume that F is a distribution function that is continuously differentiable at 0, having $F(0) = 0$ and $F'(0) > 0$. Let $\mathcal{F}$ be some class of real-valued distribution functions that contains F. Suppose that we have a sequence of functions $\{\Psi_{\mathcal{F}}^{(n)}\}_{n \geq 1}$ such that for each n it holds that $\Psi_{\mathcal{F}}^{(n)} = (\widehat{F}^{(n)}, \psi_0^{(n)})$, where $\widehat{F}^{(n)}$ maps $\mathbb{R}^n \to \mathcal{F}$, and $\widehat{F}^{(n)}$ is differentiable at 0 with the derivative $\psi_0^{(n)}$. Here $\psi_0^{(n)}(X_1, X_2, \ldots, X_n)$ is a real-valued random variable itself.

Assume that there exists a real sequence $\{r_n\}$ such that for any i.i.d. sample $X_1, X_2, \ldots, X_n$ generated from a distribution $F \in \mathcal{F}$, and any $\varepsilon > 0$, there exists a constant $C(\varepsilon, F) > 0$ such that for

$$\Pr\big(|\psi_0^{(n)}(X_1, X_2, \ldots, X_n) - F'(0)| > \varepsilon \big) \leq \frac{C(\varepsilon, F)}{r_n} \,. \tag{7}$$

If the sequence r_n satisfies

$$\lim_{n \to \infty} r_n = \infty \,, \tag{8}$$

we say that the estimator $\{\Psi_{\mathcal{F}}^{(n)}\}_{n \geq 1}$ *respects the boundary* of distribution F from the class $\mathcal{F}$. In case the constant $C(F)$ above can be chosen independently of $F \in \mathcal{F}$, we say that the estimator $\{\Psi_{\mathcal{F}}^{(n)}\}_{n \geq 1}$ *respects the boundary uniformly* for distributions from the class $\mathcal{F}$. $\qquad\square$

Such estimators actually exist—see [Balabdaoui, 2007] or [Alberts and Karunamuni, 2003] for examples. It is necessary to remark here that a statistical question of constructing estimates that are consistent at boundary points can be tricky, and is certainly a **nonstandard** task. Blind use of standard methods can lead to incorrect results: many of the well-established estimation methods are consistent in integral norms such as L_1- or L_2-norms, or within the interior of the parameter spaces. Behavior of estimators at boundary points is substantially less studied, and the estimators that behave well at the boundary are usually not governed by conventional statistical results.

An an illustration, we note that in the setup of the Definition 1, the usual kernel density estimator gives a *biased* estimate of $F'(0)$, even if one substantially restricts the space $\mathcal{F}$. Instead, Alberts and Karunamuni [2003] proposes a modified kernel density estimator that has a correction for the bias on the boundary.

On the other hand, it is important to observe that Definition 1 only requires that F and F' are consistently estimated at the single boundary point 0; at other points $\widehat{F}^{(n)}$ may even be *inconsistent*! This leaves a lot of opportunities for nonstandard constructions of estimators. Surprisingly enough, even inconsistent estimators are useful in our problem, as long as they respect the boundary.

Theorem 1 (Consistency). *Let G_n be a complete graph on n vertices with random edge weights and let the cost function φ satisfy (1) and (4) with $\varphi_2 \equiv 0$. Consider the problem of estimating the expected per vertex cost of an MST (using cost function φ) of G_n. Generate a complete auxiliary graph $G'_{d(n)}$ on $d(n)$ vertices, via sampling the new edge weights $\{Y'_e \,|\, e \in E'_{d(n)}\}$ from the distribution function $\widehat{F}^{(n(n-1)/2)}(\{Y_e \,|\, e \in E_n\})$. Suppose that $\{\Psi_{\mathcal{F}}^{(n)}\}_{n \geq 1}$ respects the boundary for F.*

1) Then, $\widehat{\mathrm{Avcost}}(G_n)$ is a consistent estimate, in the sense that for any $\varepsilon > 0$

$$\lim_{n \to \infty} Pr\bigg(\big| \widehat{\mathrm{Avcost}}(G_n) - \frac{1}{n}\varphi_1\big(\mathrm{ST}(G_n)\big) \big| > \varepsilon \bigg) = 0 \,. \tag{9}$$

2) Much more than that, our auxiliary sample allows estimating the weight of the MST itself consistently in probability, i.e., for any $\varepsilon > 0$

$$\lim_{n \to \infty} Pr\big(\big| d(n) \cdot \widehat{\mathrm{Avcost}}(G_n) - \varphi_1(\mathrm{ST}(G_n)) \big| > \varepsilon \big) = 0 \,. \tag{10}$$

The meaning of this theorem is that, for example, in the case of random complete graphs, one can consistently estimate some of their important characteristics by using just a small (but properly

constructed) model of the initial large graph. In the particular case of spanning trees, one can have the number of vertices $d(n)$ grow to infinity *arbitrarily slowly*, but still obtain asymptotically consistent estimates. This observation could be of much help in problems that require optimization on huge networks that would be practically intractable to treat as a whole.

4   Related work and open problems

In this section we first summarize some related work, and then discuss a list of open problems and challenges arising from this paper.

4.1   Related work

Random graph theory is a mature subject (see [Bollobás, 2001]); but our interest is more specific. In particular, we draw upon work on estimating weights of (ordinary) MSTs dating back to [Frieze, 1985]. For a good summary, and additional references we refer to the paper of Steele [2002]. Bertsimas [1990] studies a closely related but very different formulation, wherein he assumes that nodes may be present (or absent) with a certain probability. Based on this model, he studies what the expected weight of an MST might be. In contrast, we assume that the edge weights are random (according to specific law), and we study the expected value under a cost function strictly more general than the ordinary linear cost used for MSTs. Also note that in our framework one tends to build auxiliary graphs on $d(n) \ll n$ vertices, so our method is intended to works for graphs that are incomparably smaller than the original graph, while Bertsimas [1990] studies graphs on $O(n)$ nodes.

To make our method practical, we depend on availability of an algorithm to solve the φ-MST problem on the auxiliary graph. For appropriate choices of the cost function φ, recent algorithms such as those of [Stobbe and Krause, 2010] or [Jegelka and Bilmes, 2011b,a] might offer practical methods for tackling the subproblem on the auxiliary graphs. Additionally, there is a well-developed body on submodular optimization that we could tap into; see for instance [Chudak and Nagano, 2007, Fujishige, 2005, Iwata and Nagano, 2009]. We note, however, that submodular set functions offer only one class of possible cost functions—if algorithms (or approximation algorithms) are available for other type of cost functions, we could benefit from those too—e.g., those in [Murota, 2003].

4.2   Open Problems

Since this is a position paper that also advances a new set of research problems, there are numerous aspects that remain to be studied. We highlight some of the important questions below.

An important open problem is to determine the types of deterministic or random graphs for which we can ensure consistency of the estimator from Theorem 1. There are fine probabilistic results on MSTs for several classes of random graphs, both asymptotic and finite sample (see [Steele, 2002, Frieze et al., 2000] and references therein). Most notably, a lot is known about MSTs of cubes, and some other "regular" graphs. And, as some reflection shows, for such graphs, it is easy to check whether the estimator from Theorem 1 is consistent or not. But more generally, even if there is no hope to get closed form probabilistic results about the weight of the φ-MSTs, its proposed estimator may be expected to be consistent in many more interesting cases.

As shown in [Steele, 2002], the expected weight of a (linear) MST of an arbitrary connected graph G can be represented as an integral of a function that depends on the Tutte polynomial of G. This observation leads us to conjecture that the expected weight of the φ-MST for submodular φ might be representable as a Choquet integral involving Tutte polynomials. If this is the case, our estimators will also be randomized approximations of certain Choquet integrals, a curious byproduct.

As usual, it would be valuable to study rates of convergence of our estimators, as well as some basic properties such as asymptotic variance. The fact that these estimators can be consistent even when they are based on a "small" graph (with $d(n) \ll n$ vertices), is promising since it provides theoretical grounds for replacing processing on giant networks by processing suitably constructed, smaller networks. Results on variance and rates of convergence of the estimators will contribute towards judging actual accuracy of such replacements.

Since we expect our estimation to work on large graphs, it is crucial that we be able to minimize the cost function φ efficiently, at least on the auxiliary graph $G'_{d(n)}$. This raises the cornerstone question: for which types of cost functions φ (submodular, monotone, etc.) does there exist an efficient optimization method for finding (at least approximately) the desired minimum cost structure (spanning tree, path, etc.) that simultaneously also respects our statistical estimation procedure. The present short paper suggests that this class of cost functions is rich (at least infinite-dimensional).

Finally, we close by mentioning that even though we illustrated only spanning trees, the same argument extends to obtaining estimators for any other graphical structures such as paths, cuts, etc., as long as suitable estimators are available for corresponding linear cost functions. More challengingly, we wish to consider deriving conditions on φ_1 and φ_2 in the decomposition (1), under which one obtains consistent estimators.

References

T. Alberts and R. J. Karunamuni. A semiparametric method of boundary correction for kernel density estimation. *Statistics and Probability Letters*, 61(3):287–298, 2003.

F. Balabdaoui. Consistent estimation of a convex density at the origin. *Mathematical Methods of Statistics*, 16:77–95, 2007. ISSN 1066-5307.

D. J. Bertsimas. The probabilistic minimum spanning tree problem. *Networks*, 20(3):245–275, 1990.

B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

F. A. Chudak and K. Nagano. Efficient solutions to relaxations of combinatorial problems with submodular penalties via the Lovász extension and nonsmooth convex optimization. In *SODA*, 2007.

A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, 10:47–56, 1985.

A. M. Frieze, M. Ruszinkó, and L. Thoma. A note on random minimum length spanning trees. *Electronic Journal of Combinatorics*, 2000.

S. Fujishige. *Submodular functions and optimization*, volume 58 of *Annals of Discrete Mathematics*. Elsevier Science, 2005.

G. Goel, P. Tripathi, and L. Wang. Optimal Approximation Algorithms for Multi-agent Combinatorial Problems with Discounted Price Functions. In *Foundations of Software Technology and Theoretical Computer Science*, 2010.

S. Iwata and K. Nagano. Submodular function minimization under covering constraints. In *FOCS*, 2009.

S. Jegelka and J. A. Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *Computer Vision and Pattern Recognition (CVPR)*, June 2011a.

S. Jegelka and J. A. Bilmes. Approximation bounds for inference using cooperative cuts. In *International Conference on Machine Learning (ICML)*, 2011b.

K. Murota. *Discrete Convex Analysis*. SIAM, 2003.

J. M. Steele. Minimum spanning trees for graphs with random edge lengths. In *In Mathematics and Computer Science II: Algorithms, Trees, Combinatorics and Probabilities, Birkhäuser*, pages 223–245, 2002.

P. Stobbe and A. Krause. Efficient minimization of decomposable submodular functions. In *NIPS*, 2010.