Counting Vertex Covers in General Graphs

Ad Ridder (VU Amsterdam)
Zdravko Botev (UNSW Sydney)
Radislav Vaisman (UQ Brisbane)

APS, 6 July, 2015
What is a Vertex Cover in a Graph?

- A set of vertices such that each edge of the graph is incident to at least one vertex of the set.

Example

Finding a minimum vertex cover is one of the classical NP-complete decision problems.

\{v_1, v_3\} and \{v_2, v_3\} are minimal vc’s. All supersets of these are vc.
**Associated Counting Problem**

- How many vertex covers are there for a given graph?
- #P-complete counting problem.
- Related to propositional model counting.
- Efficient model counting algorithms are of interest for Bayesian inference problems or combinatorial design problems.
Randomized Approximation Algorithms

- We will consider simple undirected graphs $G = G(V, E)$.
- Let $c_G(n)$ be the exact (but unknown) number of vertex covers in an instance graph $G$ with $n = |V|$ vertices.
- A randomized algorithm produces a random output $\hat{c}_G(n)$ as estimate.
- A randomized algorithm is a fully polynomial randomized approximation scheme (FPRAS) if for every triple $(n, \epsilon, \delta)$ the output satisfies
  \[ \mathbb{P}\left( (1 - \epsilon)c_G(n) < \hat{c}_G(n, \epsilon, \delta) < (1 + \epsilon)c_G(n) \right) > 1 - \delta \]
  in a running time that is polynomial in $\epsilon^{-1}, \log \delta^{-1}$ and $n$.
- Note that $\epsilon$ and $\delta$ may be part of the input of the estimator.
FPRAS Successes

- Other combinatorial counting problems.
- Generally hard to construct FPRAS.
- Some (not exhausted!) are
  - Karp et al. (1989) for counting the number of satisfying assignments to a boolean formula in disjunctive normal form.
  - Jerrum and Sinclair (1996) for counting the number of matchings (of all sizes) in a graph.
  - Cryan and Dyer (2003) for the number of contingency tables when the number of rows is constant.
  - Dyer (2003) for counting the number of solutions to a 0-1 knapsack problem.
  - Jerrum et al. (2004) for counting the permanent of a matrix with nonnegative entries.
FPRAS for Counting Vertex Covers in a Graph

- Not (yet?) developed.
- But ...
FPRAS for Counting Vertex Covers in Random Graphs

We have constructed an algorithm that shows FPRAS for random graphs. This means

1. Let $\mathcal{S}(n)$ be the set of all (simple undirected) graphs with $n$ vertices.

2. Then
   $$\Pr_{\text{EG}}(\text{algorithm is FPRAS for } G \in \mathcal{S}(n)) \to 1,$$
   as $n \to \infty$, when $G$ is drawn randomly from $\mathcal{S}(n)$ according to the Edgar Gilbert model.

3. This means that each edge from the $\binom{n}{2}$ possible edges is present with probability $1/2$.  

The Algorithm

Importance sampling.

- Given an undirected simple graph $G = G(V, E)$ with $n = |V|$ vertices.
- Consider binary vectors $\mathbf{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$.
- Any binary vector corresponds one-to-one with a vertex set $V(\mathbf{x}) \subset V$ by
  \[ v_i \in V(\mathbf{x}) \iff x_i = 1. \]
- Let $f$ be a proposal PMF on $\{0, 1\}^n$ such that
  \[ V(\mathbf{x}) \text{ is vertex cover in } G \implies f(\mathbf{x}) > 0. \]
- Then
  \[ c_G(n) = \mathbb{E}_f \left[ \frac{\mathbb{I}\{V(\mathbf{x}) \text{ is vertex cover in } G\}}{f(\mathbf{X})} \right]. \]
Sequential Importance Sampling (SIS)

- Decomposition by conditional PMF’s:

\[ f(x) = \prod_{i=1}^{n} f_i(x_i|x_1, \ldots, x_{i-1}). \]

- Given a proposal \( f \),
  - easy to generate \( x_1, x_2, \ldots \) iteratively from the conditional PMF’s;
  - hence, easy to get binary vector \( x \sim f \);
  - finally, easy to check vertex cover property of associated vertex set \( V(x) \).

- Repeat \( N \) times to get unbiased estimator

\[ \hat{c}_G(n) = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbb{I}\{V(X_i) \text{ is vertex cover in } G\}}{f(X_i)} \cdot \]

- For what proposal \( f \) is SIS algorithm FPRAS for random graphs?
The Zero-variance Proposal PMF

- Define
  \[ f^*(x) = \frac{1}{c_G(n)} \mathbb{1}\{V(x) \text{ is vertex cover in } G\}. \]

- Then \[ \text{Var}_{f^*}(\hat{c}_G(n)) = 0. \]

- This is optimal importance sampling (and certainly FPRAS).

- Unfortunately, not implementable.

- But ...
Deomposition of Zero-variance PMF

We can show that

\[ f^*(x) = \prod_{i=1}^{n} f^*_i(x_i|x_1, \ldots, x_{i-1}). \]

Where

\[ f^*_i(1|x_1, \ldots, x_{i-1}) = \frac{c_{G[i]}}{c_{G[i]} + c_{G[-i]}} \]

\[ f^*_i(0|x_1, \ldots, x_{i-1}) = 1 - f^*_i(1|x_1, \ldots, x_{i-1}) \]

Where

- \( G[i] \) and \( G[-i] \) are specific (known) subgraphs of \( G \), given by the values of \( x_1, \ldots, x_{i-1} \);
- \( c_{G[i]} \) is the associated number of vertex covers in subgraph \( G[i] \) (exact but unknown).
An Implementable Proposal PMF

- Approximate the conditional zero-variance PMF’s:

\[ f_i(1|x_1, \ldots, x_{i-1}) = \frac{A[i]}{A[i] + A[-i]} \cdot \]

- Where \( A[i] \) and \( A[-i] \) are computable approximations of \( c_{G[i]} \) and \( c_{G[-i]} \), respectively.

- As follows (for \( c_{G[i]} \)):
  - Given \( x_1, \ldots, x_{i-1} \), determine subgraph \( G[i] \);
  - Say \( G[i] \) has \( k \) vertices;
  - Let \( G \) be a random graph of \( k \) vertices according to the Edgar Gilbert model;
  - Then set \( A[i] = \mathbb{E}_{EG}[c_{G}(k)] \);

- Easy to compute

\[ \mathbb{E}_{EG}[c_{G}(k)] = \sum_{i=0}^{k} \binom{k}{i} 2^{-i(i)} \]
Main Result

**Theorem**

The SIS algorithm with the approximated conditional zero-variance PMF’s is FPRAS for counting vertex covers in random graphs.

The proof is based on a similar result for counting cliques (Rasmussen 1997) and the relation between vertex covers in a graph and cliques in the complement graph.
Improved Algorithm

- Again approximate the vertex cover numbers $c_{G[i]}$ and $c_{G[-i]}$ that pop up in the expression of the conditional zero-variance PMF's:

$$\tilde{f}_i(1|x_1, \ldots, x_{i-1}) = \frac{B[i]}{B[i] + B[-i]}.$$ 

- Approximation is based on a vertex cover relaxation.
Vertex Cover Relaxation

- Consider the subgraph $G[i] = (V[i], E[i])$.
- Suppose $k$ vertices.
- Label the vertices in some order $v_1, \ldots, v_k$.
- Define probabilities $p_i = d_i/(k - 1)$;
- where $d_i$ = the number of downstream (i.e., $j > i$) neighbours of $v_i$.
- Define a probability space $\Omega_G$ of all graphs $G' = (V[i], E')$ with the same vertex set $V[i]$;
- but where each possible edge $(v_i, v_j)$, $j > i$ is present in $E'$ with probability $p_i$.
- Let $G$ be a random graph in this probability space.
- Then set $B[i] = \mathbb{E}_{\Omega_G}[c_G(k)]$.
- It can be shown that the computation of $B[i]$ has polynomial complexity $O(k^2)$.
- Similarly for the subgraph $G[-i]$. 
Comparison

Conjecture

The SIS estimators of counting vertex covers satisfy

$$\mathbb{V}ar_f(\hat{c}_G(n)) \leq \mathbb{V}ar_f(\hat{c}_G(n)).$$
Experiments

- Our SIS algorithms denoted Alg. A and Alg. B.
- *Cachet* is exact model counting software introduced by Sang et al. (2004); based on a SAT solver.
- *SampleSearch* is a probabilistic model counting technique by Gogate and Dechter (2006, 2007); based on sampling from the search space of a Boolean formula.
- No randomized algorithms have been developed dedicated to the vertex cover counting problem.
Random Graphs

- 40 random graphs for each \( n = 5, 10, \ldots, 100 \).

- Plot of the estimated coefficients of variation of the SIS estimators (ratio of variance and square mean).
A Small Model

- $n = |V| = 100$ vertices and $|E| = 2432$ edges.
- Exact (Cachet): $c_G(n) = 244941$.
- Alg. B: estimate $2.444e+05$ with (numerical) relative error $1.28e-02$.
- SampleSearch: estimate 196277!
A Large Model

- \( n = |V| = 1000 \) vertices and \( |E| = 249870 \) edges.
- Alg. B estimate \( 2.773e+11 \) with (statistical) relative error \( 1.579e-02 \).
- Cachet and SampleSearch failed.
References