

8-13-2020

Research of Approximation Algorithm of Vertex Cover

Wenyu Gao

School of information science, Guangdong University of Finance and Economics, Guangzhou 510320, China;

Li Hua

School of information science, Guangdong University of Finance and Economics, Guangzhou 510320, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Research of Approximation Algorithm of Vertex Cover

Abstract

Abstract: Vertex cover is one of the most important NP complete problems, and it is also one of the most studied problems in parameterized algorithm design in recent years. *According to some deficiencies of existing vertex cover approximation algorithms, a novel NT theorem based 2-approximation algorithm was proposed based on advances in parameterized algorithm. The new algorithm first used kernelization technology to reduce the origin graph, then used greedy strategy to guide the vertex selection in the remaining graph, kernelization provided effective guarantee for the approximation algorithm.* In order to test the performance of the new algorithm, simulations on different types of random graph were carried to compare the new algorithm and the classical matching-based 2-approximation algorithm. Simulation results show that the new algorithm outperforms the matching-based 2-approximation algorithm.

Keywords

vertex cover, NP complete, approximation algorithm, parameterized algorithm, NT theorem

Recommended Citation

Gao Wenyu, Li Hua. Research of Approximation Algorithm of Vertex Cover[J]. Journal of System Simulation, 2016, 28(11): 2784-2789.

点覆盖问题的近似算法研究

高文宇, 李华

(广东财经大学信息学院, 广州 510320)

摘要: 点覆盖问题是最重要的 NP 完全问题之一, 也是近年来参数算法设计中研究得最多的问题之一。针对现有点覆盖近似算法的一些不足, 基于点覆盖问题参数算法的进展, 提出了该问题一个基于 NT 定理规约的 2-近似算法。利用了参数算法中的核化技术对图进行化简, 在剩余图上采用贪心策略来指导节点的选择, 核化技术为算法提供了有效的近似度保障。为检验新算法性能, 在不同类型的随机图上通过仿真实验比较了新算法和经典的基于匹配的 2-近似算法。仿真实验结果表明新算法较基于匹配的 2-近似算法有着明显的优势。

关键词: 点覆盖; NP 完全; 近似算法; 参数算法; NT 定理

中图分类号: TP301.6 文献标识码: A 文章编号: 1004-731X (2016) 11-2784-06

DOI: 10.16182/j.issn1004731x.joss.201611020

Research of Approximation Algorithm of Vertex Cover

Gao Wenyu, Li Hua

(School of information science, Guangdong University of Finance and Economics, Guangzhou 510320, China)

Abstract: Vertex cover is one of the most important NP complete problems, and it is also one of the most studied problems in parameterized algorithm design in recent years. According to some deficiencies of existing vertex cover approximation algorithms, a novel NT theorem based 2-approximation algorithm was proposed based on advances in parameterized algorithm. The new algorithm first used kernelization technology to reduce the origin graph, then used greedy strategy to guide the vertex selection in the remaining graph, kernelization provided effective guarantee for the approximation algorithm. In order to test the performance of the new algorithm, simulations on different types of random graph were carried to compare the new algorithm and the classical matching-based 2-approximation algorithm. Simulation results show that the new algorithm outperforms the matching-based 2-approximation algorithm.

Keywords: vertex cover; NP complete; approximation algorithm; parameterized algorithm; NT theorem

引言

图 $G(V, E)$ 的点覆盖是图的节点集的一个子集 $C \subseteq V$, 使得对任意的边 $(u, v) \in E$, 有 $u \in C$ 或 $v \in C$ 。即子集 C 中的节点覆盖了图 G 所有的边。最小点覆盖问题就是在一个给定的图中, 找出最小

规模的点覆盖。将这一最优化问题表述为一个判断问题就是, 给定图 $G(V, E)$ 中, 是否存在一个点覆盖集 $C \subseteq V$, 使得 $|C| \leq k$ (其中 k 为一个常数)。图的点覆盖问题是最重要的 NP 完全问题之一, 其在无线传感器网络、生物信息学等领域都有着重要的应用价值。点覆盖问题存在简单的 2-近似算法, 一些特殊图中的点覆盖问题是多项式时间可解的, 著名的 Konig 定理指出在二分图中, 最小点覆盖的大小等于最大匹配的大小, 它们都是多项式时间可解的^[1]。多年来, 研究者从不同角度对点覆盖问题



收稿日期: 2015-03-19 修回日期: 2015-11-28;
基金项目: 广东省教育厅科技创新项目(2013KJCX0-084), 广东省自然科学基金(8151032001000013);
作者简介: 高文宇(1972-), 男, 湖南永州, 博士, 教授, 研究方向为算法理论及应用。

<http://www.china-simulation.com>

• 2784 •

展开了深入研究,特别是随着近年来参数计算理论^[2-3]的发展,点覆盖成为了参数计算中被研究得最多的问题。Buss 和 Goldsmith 最先给出求解点覆盖问题的复杂度为 $O(kn + 2^k k^{2k+2})$ 的参数算法^[4],不久,Downey 和 Fellows 将这一结果提高到 $O(kn + 2^k k^2)$ ^[5]。之后,Balasubramanian 等人将底数降到了 2 以下,给出了一个 $O(kn + 1.324718^k k^2)$ 的算法^[6],随后 Chen 又给出时间复杂度为 $O(kn + 1.285^k)$ 的算法^[7]。目前最好结果是 Chen 设计的时间复杂度为 $O(1.2738^k)$ 的参数算法^[8]。本文在点覆盖核化算法的基础上设计了一个 2-近似算法,然后通过随机仿真实验比较了不同的近似算法的性能。

1 点覆盖问题的近似算法

1.1 基于匹配的 2-近似算法

对于点覆盖问题而言,有一个非常直观的贪心算法。由于点覆盖是要选取尽量少的点来覆盖图中所有的边,因此通过在图中反复选取度最大的点即可得到一个较小的点覆盖。但是该近似算法存在明显的弱点,并且无法提供确定的近似比保障。如图 1 所示的图 G ,若按照度最大原则选择点覆盖,则首先会选定节点 8 和 9,然后还需在节点 0~7 中选至少 4 个点,因此得到的点覆盖集大小为 6。而该图的最小点覆盖集可以由 $\{4,5,6,7\}$ 这 4 个节点构成。

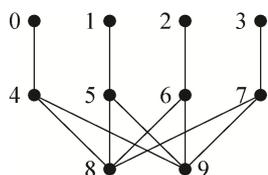


图 1 图 $G(V, E)$

为了获得有近似比保证的近似算法,经典的做法是采用基于匹配的方法,即每次从图中选择一条边(即一个匹配),将其关联的两个点加入点覆盖集,同时将该条边及其相邻边从图中删除,随后在剩余图中重复此过程,直至剩余图的边集为空。详细算法描述如下。

基于匹配的 2-近似算法:

- 1 $C \leftarrow \emptyset$
- 2 while $E \neq \emptyset$
- 3 从 E 中任取一条边 uv
- 4 $C \leftarrow C \cup \{u, v\}$
- 5 从 E 中删除所有与节点 u 或 v 相连的边
- 6 endwhile
- 7 return C

由于所有被选出的边在原图中构成一个极大匹配,因此哪怕仅仅覆盖选出的这 k 条边也需至少 k 个点,因此原图的最小点覆盖不可能小于 k 个点,而该算法得到的点覆盖的大小为 $2k$,因此这是一个 2-近似算法。

基于匹配的 2-近似算法有着很好的理论上的近似比保障,但是从实际情况来看,有时可能会得到较大的点覆盖集。例如在图 1 所示的图上运行该算法,得到的点覆盖集大小为 8。

1.2 基于 NT 定理规约的 2-近似算法

鉴于上述近似算法的不足,我们根据点覆盖的参数化核化算法设计了一个新的近似算法。核化^[3,9]是参数算法设计中最重要技术之一,所谓核化是指如果存在一个多项式时间算法 K 和一个递归函数 g ,使得对于参数问题 Q 的任意一个实例 (x, k) ,应用算法 K 将 Q 转化为一个新实例 (x', k') ,使得 $|x'| \leq g(k)$ 和 $k' \leq k$,并满足当且仅当 (x', k') 是参数问题 Q 的一个真实例, (x, k) 也是 Q 的一个真实例,则我们说参数问题 Q 是可核化的,算法 K 称为核化算法,新实例 (x', k') 称为核化后的问题核。核化后问题规模极大地降低。核的大小可能是参数 k 的线性式(线性核),也可能是参数 k 的多项式(多项式核),甚至是参数 k 的非多项式。一个规模极大降低的核不仅可以用于设计参数算法,也可用于近似算法的设计。NP 完全问题的参数化核化算法研究是当前一个研究热点,许多问题都得到了多项式甚至线性核,然而这些核化算法在实际应用中的效果如何却很少见到这方面的研究。我们根据点覆盖问题的参数化核化算法的一些结论设计

了一个 2-近似算法,并通过随机仿真实验对算法的实际性能进行了分析。

目前点覆盖问题最好的核化算法可以得到 $2k$ 大小的线性核^[7],其关键在于利用 NT 定理^[10-11]对图进行了规约。主要过程如下。

(1) 根据图 $G(V, E)$ 定义一个二分图 $B(V', E_B)$, 其中 $|V'| = |V|$, 二分图的边集 E_B 定义如下: $E_B = \{(x, y') | (x, y) \in E\}$ 。

(2) 求出二分图 $B(V', E_B)$ 的最小点覆盖 C_B , 并根据 C_B 将图 G 的节点集 V 划分为如下 3 个子集(根据 Konig 定理, 二分图的最小点覆盖大小等于其最大匹配的大小, 有多种方法求二分图的最小点覆盖):

$$V_1 \leftarrow \{x | x \in C_B \text{ AND } x' \in C_B\};$$

$$V_{1/2} \leftarrow \{x | x \in C_B \text{ XOR } x' \in C_B\}; V_0 \leftarrow V - (V_1 \cup V_{1/2})。$$

(3) 在 NT 定理中证明了集合 V_1 必定属于图 G 的某个最小点覆盖集, 集合 V_1 加上诱导子图 $G(V_{1/2})$ 的最小点覆盖集可以构成图 G 的最小点覆盖集, 且诱导子图 $G(V_{1/2})$ 的最小点覆盖集的大小不少于 $V_{1/2}$ 包含的节点数的一半。

因此, 根据 NT 定理的结果可以设计出如下的具有近似比保障的 2-近似算法。

基于 NT 定理规约的 2-近似算法:

- 1 $C \leftarrow \emptyset$
- 2 根据图 G 生成对应的对称二分图 B
- 3 找出二分图 B 的最小点覆盖集 C_B
- 4 根据 C_B 将图 G 的节点集 V 划分为必选点集 V_1 、可选点集 $V_{1/2}$ 以及不选点集 V_0
- 5 $C \leftarrow C \cup V_1$
- 6 对可选点集 $V_{1/2}$ 的诱导子图 $G(V_{1/2})$ 执行基于节点度的贪心策略, 通过反复选取度最大的节点求出诱导子图 $G(V_{1/2})$ 的点覆盖集 $CV_{1/2}$
- 7 $C \leftarrow C \cup CV_{1/2}$
- 8 return C

由于集合 $V_{1/2}$ 中不少于一半的点属于图 G 的最小点覆盖集, 而上述算法中得到的点覆盖集 C 由必选点集合 V_1 和诱导子图 $G(V_{1/2})$ 的点覆盖集构

成, 因此所得的点覆盖集 C 小于等于最小点覆盖的 2 倍。所以, 上述算法也是一个 2-近似算法。

相比于基于匹配的 2-近似算法, 新算法不是盲目地选择相邻的点作为点覆盖集中的候选节点, 而是借助于 NT 定理的结果确定了一些必选节点, 同时将不必要的点删除, 将剩余节点控制在一定范围之内, 然后在剩余节点上应用贪心策略来指导候选节点的选择。基于 NT 定理的新算法虽然在理论上仍然是 2-近似算法, 但是由于有效地避免了一些冗余的选择, 因此有理由预计新算法具有更好的效果。

时间复杂度方面, 算法中根据图 G 生成对应的二分图 B 所需时间为 $|E|$ (将图 G 的 $|E|$ 条边转换成 $2|E|$ 条边插入到二分图 B 中, 二分图 B 的规模和图 G 的规模是线性相关的, 即二分图 B 的节点数和边数分别是图 G 的节点数和边数的两倍); 找出二分图 B 的最小点覆盖集可以利用二分图的最大匹配算法, 而借助于网络流的二分图最大匹配算法可以仅用 $O(|E| \sqrt{|V|})$ 时间求出^[12]; 根据二分图 B 的最小点覆盖集划分图 G 的节点集 V 所需时间为 $|V|$ 。最后使用贪心策略求出诱导子图 $G(V_{1/2})$ 的点覆盖集所需时间为 $|V_{1/2}|^2$ 。因此整个算法能以较高效率运行。

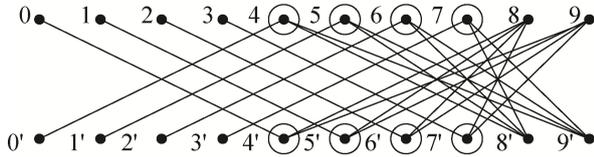
基于 NT 定理规约的 2-近似算法不仅在理论上有着严格的近似比, 而且由于其在处理过程中提前去掉了一些不必要的点, 极大地降低了图的规模, 因此在实际应用中也有着很好的效果。例如对于图 1 中所示的图运行基于 NT 定理规约的 2-近似算法可得结果如下:

首先根据图 1 生成对应的二分图 B (见图 2)。

图 1 中的图 G 包含 10 个节点(序号为 0~9), 因此二分图 B 的左集和右集分别包含 10 个节点(序号分别为 0~9 和 0'~9'); 图 G 中节点 0 和 4 之间有边相连, 则二分图 B 中节点 0 和 4' 以及 0' 和 4 之间有边相连, 依此类推;

然后可以使用确定性算法求出二分图 B 的最小点覆盖, 在此例中最小点覆盖 $C_B = \{4, 5, 6, 7, 4', 5', 6', 7'\}$ (图 2 中用圆圈圈住的 8 个节点), 所以依

据 NT 定理的结论可以将图 G 的节点集 V 划分为 3 个子集: $V_1 = \{4, 5, 6, 7\}$, $V_{1/2} = \emptyset$, $V_0 = \{0, 1, 2, 3, 8, 9\}$ 。因此, 图 G 的最小点覆盖集就由 V_1 和诱导子图 $G(V_{1/2})$ 的点覆盖集构成, 由于 $V_{1/2} = \emptyset$, 所以本例中图 G 的点覆盖集就是 $C = \{4, 5, 6, 7\}$, 其大小为 4。

图 2 图 $G(V, E)$ 对应的二分图 B

2 近似算法的实验分析

2.1 近似算法在普通随机图上的实验

对于近似算法来说, 随机仿真实验是验证算法性能的有效方法^[13-15]。因此我们也通过随机图上的仿真实验来比较新算法和经典的基于匹配的 2-近似算法的性能。通过指定图的节点数 $|V|$ 和边数 $|E|$, 首先生成一个 $|V|$ 个节点 0 条边的空图, 然后每次随机产生两个表示节点的随机数 u, w , 将 uw 作为一条边添加到图中, 重复该过程 E 次, 这样就得到一个包含 $|V|$ 个节点和 $|E|$ 条边的随机图。通过变更参数 $|V|$ 和 $|E|$ 可以很容易地调整图的规模和稀疏度, 以满足不同实验需求。当然, 在生成随机图的过程中需要去掉自环和平行边以保证得到的是简单图。

实验在不同节点数、不同稀疏度的随机图上进行多次, 每个随机图分别运行两个不同的近似算法。在实验中进行了 4 组随机实验, 第一组将随机图的节点数设定为 100, 边数由 0 逐步变化到 4 900(步长为 100), 在每一个边数不同的图上分别运行两个不同的近似算法, 所得点覆盖集大小的情况如图 3(a)。当节点数 $|V|=100$, 边数 $|E|=100$ 时, 该图包含的边较为稀疏, 因此点覆盖集的也较小, 当节点数 $|V|=100$, 边数 $|E|=4 900$ 时, 该图所有节点接近构成一个团(包含 100 个节点的团含的边数为 $100 \times 99 / 2 = 4 950$), 因此运行近似算法所得的点覆盖集的大小也接近 100。

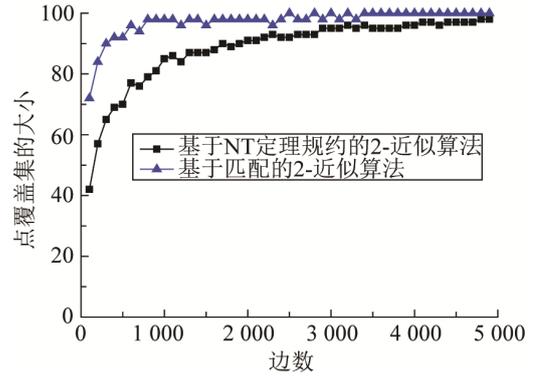
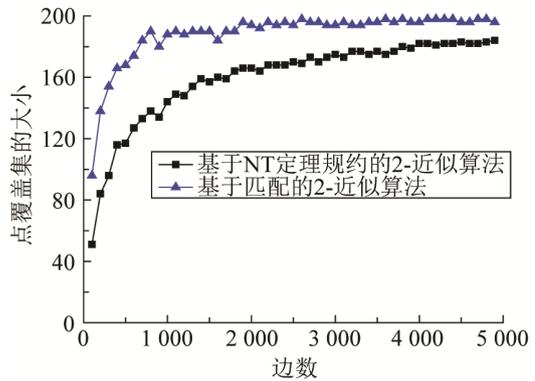
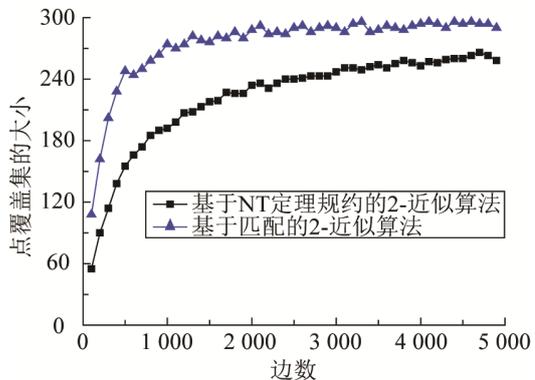
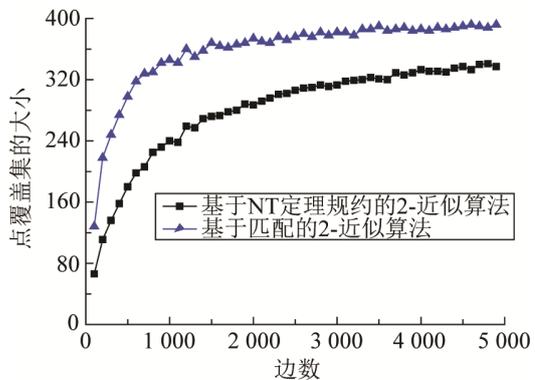
(a) $|V|=100$ (b) $|V|=200$ (c) $|V|=300$ (d) $|V|=400$

图 3 随机图上的实验结果

从图的整体趋势来看, 基于 NT 定理规约的 2-近似算法得到的点覆盖集明显小于基于匹配的 2-近似算法得到的点覆盖集。

图 3(b), 3(c), 3(d) 分别展示的是节点数固定为 200、300、400 时, 边数由 100 逐步变化到 4 900, 分别运行两个不同的近似算法所得点覆盖集大小的情况。图中趋势与图 3(a) 类似。

将图 3 中的 4 幅图比较一下还可看出随着仿真实验中节点数的增加, 基于 NT 定理规约的 2-近似算法的优势表现得更为明显。这是因为用于仿真实验的随机图的边数是固定的。随着随机图节点的增加, 而边数不变, 那么图中的边越来越稀疏。这也说明, 在边较为稀疏的图中, 新算法有着更为优越的性能。究其原因, 是因为基于匹配的 2-近似算法每次都选择两个相邻的点, 因此这两个相邻的点之间的那条边被最终的点覆盖集覆盖了两次, 在较为稀疏的图中, 这种冗余的覆盖导致了最终的点覆盖集可能较大; 而在基于 NT 定理规约的 2-近似算法中, 首先通过 NT 定理规约得到的是必选点, 其后通过贪心策略每次选择一个局部最优的点, 从而使得在有近似比保证的情况下又尽量减少了冗余覆盖。

2.2 近似算法在随机二分图上的实验

为了进一步比较算法在不同场景下的性能, 我们还采用了随机二分图来检验近似算法的效果。因为二分图的最小点覆盖是在多项式时间求出的, 因此可以对近似算法的评价提供一个基准。

随机二分图的生成方法如下。在每一个二分图的生成过程中, 首先指定左集 M 和右集 N 的大小, 以及边数 $|E|$ 。然后每次随机产生两个表示节点的随机数 $u, w (0 \leq u < |M|, 0 \leq w < |N|)$, 将 uw 作为一条边添加到二分图中, 重复该过程 E 次(期间去掉平行边), 这样就得到一个左集大小为 $|M|$, 右集大小为 $|N|$, 边数为 $|E|$ 的二分图。

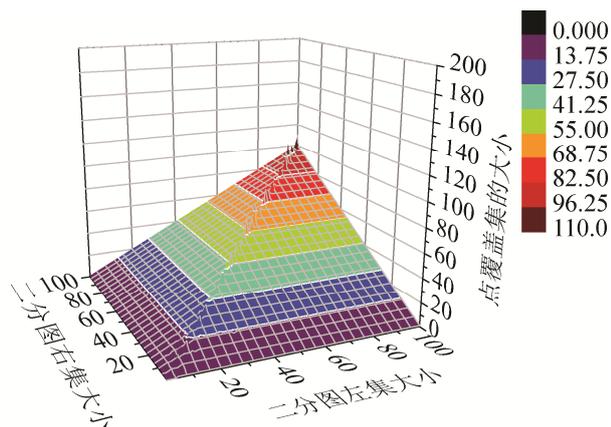
在整个实验中, 二分图的左集 M 的大小由 1 增长到 100, 然后其右集 N 的大小也从 1 增长到

100, 因此生成的二分图的数目为 10 000 个, 实验在这 10 000 个二分图上进行了 10 000 次, 设定随机二分图的边的稀疏度为 0.2, 即每个二分图包含的边数为 $|E| = |M| \times |N| \times 0.2$ 。在随机生成的 10 000 个二分图上分别运行和基于 NT 定理规约的 2-近似算法和基于匹配的 2-近似算法, 以及求二分图最小点覆盖的确定性算法。结果发现基于 NT 定理规约的 2-近似算法所得结果和确定性算法所得结果在很多时候都是一致的, 远远优于基于匹配的 2-近似算法所得结果。

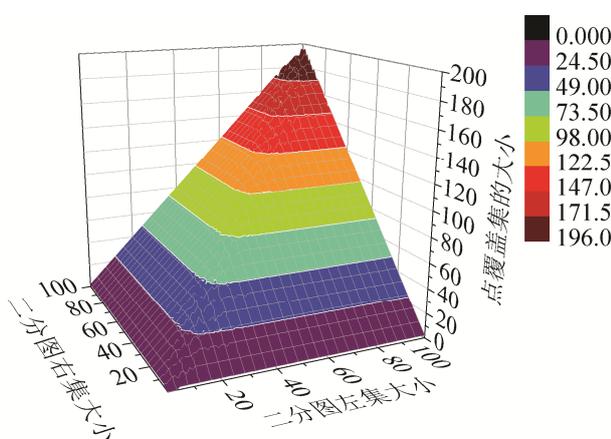
图 4(a) 是在这 10 000 个二分图上运行基于 NT 定理规约的 2-近似算法得到的点覆盖集大小的三维效果展示, 其中 X 轴为左集 M 的大小, Y 轴为右集 N 的大小, Z 轴为所得点覆盖集的大小。图 4(b) 是在这 10 000 个二分图上运行基于匹配的 2-近似算法得到的点覆盖集大小的三维效果展示。可以看出, 基于 NT 定理规约的 2-近似算法明显优于基于匹配的 2-近似算法。图 4(a) 展示的新算法得到的点覆盖集大小在多数时候都只有图 4(b) 展示的基于匹配的 2-近似算法所得点覆盖集大小的一半。这是因为在针对二分图的仿真实验中, 新算法在 NT 定理的指导下更“聪明”地倾向于在二分图的一个部集中选择节点, 而基于匹配的 2-近似算法则在二分图的两个部集中同时选择节点, 因此选择了更多的冗余节点, 从而导致最终的点覆盖集较大。

为了进一步检验边的稀疏度对于算法效果的影响, 在随机二分图中将边的稀疏度分别设定为 0.4, 0.6, 0.8, 然后重复实验, 所得结果与前面类似, 在边较为稠密的情况下, 基于 NT 定理规约的 2-近似算法依然明显优于基于匹配的 2-近似算法。

二分图上的实验更清晰地展示了基于匹配的 2-近似算法中由于选择节点时存在的冗余导致了最终结果不是太理想; 而基于 NT 定理规约的 2-近似算法则很好地解决了这个问题。



(a) 基于 NT 定理规约的 2-近似算法所得结果



(b) 基于匹配的 2-近似算法所得结果

图 4 随机二分图上的实验结果

3 结论

图的点覆盖问题是最重要的 NP 完全问题之一, 由点覆盖问题研究而得到的 NT 定理对近似算法和参数算法设计都有着重要的影响。本文提出了一个基于 NT 定理规约的 2-近似算法, 新算法利用 NT 定理的相关结论对算法的近似度提供了有效保证, 较低的算法时间复杂度也让算法在实际应用中能高效运行。文中通过在随机图和随机二分图上的算法仿真实验比较了不同近似算法的性能。实验结果表明, 基于 NT 定理规约的 2-近似算法较经典的基于匹配的 2-近似算法有着明显的优势, 这主要得益于新算法在处理过程中去掉了一些不必要的节点, 极大地降低了图的规模, 从而优化了最终结果。

参考文献:

- [1] West D B. Introduction to graph theory [M]. Second edition ed. USA: Prentice Hall, 2001.
- [2] Downey R G, Fellows M R. Parameterized Complexity [M]// Monographs in Computer Science. New York, USA: Springer-Verlag New York Inc., 1999: 548.
- [3] Flum J, Grohe M. Parameterized Complexity Theory [M]// ed. by W Brauer, G Rozenberg, A Salomaa, Texts in Theoretical Computer Science. An EATCS Series Berlin, Germany: SpringerVerlag, 2006: 495.
- [4] Buss J F, Goldsmith J. Nondeterminism within P [J]. Siam J. Comput. (S1064-8275), 1993, 22(3): 560-572.
- [5] Downey R G, Fellows M R. Parameterized Computational Feasibility [M]// Feasible Mathematics II Progress in Computer Science and Applied Logic. USA: Birkhäuser Boston, 1995: 219-244.
- [6] Balasubramanian R, Fellows M R, Raman V. An improved fixed-parameter algorithm for vertex cover [J]. Information Processing Letters (S0020-0190), 1998, 65(3): 163-168.
- [7] Chen J, Kanj I A, Jia W. Vertex Cover: Further Observations and Further Improvements [J]. Journal of Algorithms (S0196-6774), 2001, 41(2): 280-301.
- [8] Chen J, Kanj I A, Xia G. Improved upper bounds for vertex cover [J]. Theoretical Computer Science (S0304-3975), 2010, 411(40/42): 3736-3756.
- [9] 李绍华, 王建新, 冯启龙, 等. 参数计算中核心化技术及其应用 [J]. 软件学报, 2009, 20(9): 2307-2319.
- [10] Nemhauser G L, Trotter L E. Vertex packing: structural properties and algorithms [J]. Mathematical Programming (S0025-5610), 1975, 8(1): 232-248.
- [11] Bar-yehuda R, Even S. A local-ratio theorem for approximating the weighted vertex cover problem[J]. North-Holland Mathematics Studies (S0356-2670), 1985, 109(1): 27-45.
- [12] Dinic E A. Algorithm for solution of a problem of maximum flows in networks with power estimation[J]. Soviet Mathematics Doklady (S1064-5624), 1970, 11(11): 1277-1280.
- [13] 李军侠, 吴卫山. 加权快速聚类距离解模糊算法 [J]. 系统仿真学报, 2012, 24(4): 868-872.
- [14] 郭晓军, 胡云安, 张雷. 单变量多单元半全局极值搜索算法仿真研究 [J]. 系统仿真学报, 2013, 25(8): 1901-1906.
- [15] 王强, 张安, 宋志蛟. UAV 协同任务分配的改进 DPSO 算法仿真研究 [J]. 系统仿真学报, 2014, 26(5): 1149-1155.