Hardness of Computing the Biclique Partition Number

Alexander R. Block

July 17, 2017
Introduction

- Biclique Partition Number (in short, \( bp \))
  - Minimum number of \textit{bicliques} needed to partition the edges of a graph \( G \), denoted \( \text{bp} (G) \)
  - Note that \( G \) can be any graph
  - \textit{Bicliques} are \textit{complete bipartite graphs}, denoted \( K_{n,m} \)
Introduction (Examples)

\[ K_4 = K_{1,3} + K_{1,2} + K_{1,1} \]

\[ K_{2,2} = K_{2,1} + K_{1,1} \]
Introduction

- Graham and Pollak introduced the biclique partition number in 1972 in the context of network addressing and graph storage problems [GP71, GP72]
- Introduced an extremely prolific research area in Mathematics
- Graham-Pollak Theorem: $bp(K_n) = (n - 1)$. All proofs are algebraic and no purely combinatorial proof is known [GP72, Tve82, Pec84, Vis08, Vis13]
  - Showed $bp(G) \geq \max\{n_+(A(G)), n_-(A(G))\}$ [Witsenhausen, 1980s]
  - Known that $n_-(A(K_n)) = (n - 1)$
  - Can partition $K_n$ into $(n - 1)$ stars
    - A *star* is a biclique of the form $K_{1,i}$ for some positive integer $i$
Applications of Biclique Partition

- Graham and Pollak that a problem on loop switching in networking is equivalent to partitioning a multigraph, yielding their celebrated result \[\text{[GP71, GP72, Tai13]}\]
- Has applications for *perfect hashings* \[\text{[Tai13]}\]

\[
\begin{align*}
\text{INPUT: } & n, r, k \in \mathbb{N} \text{ with } k \leq r \leq n \\
\text{MINIMIZE: } & \text{size of } \mathcal{F} := \{f_i: [n] \to [r]\} \\
\text{CONSTRAINT: } & \forall K \subseteq [n] \text{ with } |K| = k, \exists i \text{ such that } f_i|_K \text{ is injective}
\end{align*}
\]

- Asking for *unique* \(i\) and \(r = k = 2\) asks for \(bp(K_n)\)
- Connections to the nondeterministic state complexity of finite automata, namely used as a lower bound method \[\text{[GH06]}\]
- Play a roll in analysis of HLA reaction matrices used in biology \[\text{[NMWA78]}\]
Applications of Biclique Partition

- A relaxation: $bp_t(G)$, a covering of edges with at most $t$-bicliques
  - Examined by Noga Alon [Alo97]
  - Showed that with $G = K_n$, $bp_t(G)$ is equivalent to finding the max number of boxes in $\mathbb{R}^n$ that are $t$-neighborly
  - Also showed that $bp_t(K_n) \geq \Theta(tn^{1/t})$

- Wyner’s common information $J(R_A, R_B)$:
  - Minimum (amount of) leakage that kills the possibility of key agreement
  - $\min H(L)$ such that $I(R_A, R_B|L) = 0$
  - Bicliques are useless for KA (because 0 mutual information)
  - Roughly corresponds to the biclique partition number

- Close connections to communication complexity and circuit lower bounds [HS12, NW95, Raz92]
Finding bp is Hard

- Suppose we are given a (bipartite) graph $G$
- Does there exist a biclique partition of $G$ of size $k$?
- The problem is NP-Complete for both bipartite and general graphs [Orl77, Cio05]
- We show a proof for bipartite graphs and another for general graphs.
- Proof for bipartite graphs is a reduction from the vertex clique problem

| GIVEN: Graph $G$ with $V(G) = \{v_1, \ldots, v_n\}$ |
| DETERMINE: Fewest number of cliques which include all of $V(G)$ |

- Proof for general graphs is a reduction from the vertex cover problem
Proof of NP-Completeness of \textit{bp} for bipartite graphs

- Suppose we are given a $G$ in the instance of the \textit{vertex clique problem} described before and we want to answer the following question

\begin{center}
\begin{tabular}{ll}
\textbf{GIVEN:} & Bipartite graph $G$ \\
\textbf{DETERMINE:} & Fewest number of bicliques which partition a subset $H \subseteq E(G')$
\end{tabular}
\end{center}

- Construct $G' = (L, R, E')$ with $L = \{x_1, \ldots, x_n\}$, $R = \{y_1, \ldots, y_n\}$ and $E' = \{(x_i, y_i) : \forall i\} \cup \{(x_i, y_j) : (v_i, v_j) \in E(G)\}$

- Let $H' = \{(x_i, y_j) : i = 1, \ldots, n\}$ be the set of edges to be covered.

- Any clique $C$ in $G$ which includes $v_i$ induces a biclique in $G'$ which includes the edge $(x_i, y_i)$.

- If $C'$ is a biclique of $G'$ which includes edges $(x_{j_1}, y_{j_1}), (x_{j_2}, y_{j_2}), \ldots, (x_{j_k}, y_{j_k})$, then by construction it must be the case that $\{v_{j_1}, \ldots, v_{j_k}\}$ is a clique in $G$.

- So the minimum number of cliques that cover all vertices in $G$ is equal to the minimum number of bicliques of $G'$ needed to cover the edges in $H'$. 

8 / 16
Proof of NP-Completeness of \textbf{bp} for general graphs

- Suppose we are given a graph $G$ and need to find a vertex cover of size $k \leq |V(G)|$
- Transform $G$ into $G'$ by replacing every edge with a path of 3 edges
- $G'$ contains no 4-cycles, so only stars are bicliques in $G'$
- This implies that $\text{bp}(G') = \alpha(G')$, where $\alpha(G')$ is the size of the minimal vertex cover of $G'$
- Notice that $\alpha(G') = \alpha(G) + |E|$
- Thus, $\text{bp}(G') = \alpha(G') = \alpha(G) + |E|$
- So $\alpha(G) \leq k$ if and only if $\text{bp}(G') \leq k + |E|$
Approximating \textbf{bp} is Hard

- Since determining \textbf{bp} is NP-Hard, can we approximate?
- Unfortunately, \textbf{bp} is also NP-Hard to approximate
  \cite{Sim90, BMB+08, CHHK14}
- Simon \cite{Sim90} examined reductions which preserved approximability of hard problems
  - Many times, near optimal solution in one problem reduces to a poor solution in another
  - Gives proof that \textbf{bp} is NP-Hard to approximate by a continuous reduction from the vertex clique problem discussed earlier
  - The proof is not very insightful, so it will be skipped in this talk
Chalermsook, Heydrich, Holm, and Karrenbauer [CHHK14] proves an approximation algorithm for \( \text{bp} \) with approximation guarantee of \( O\left(\frac{n_L}{\sqrt{\log(n_L)}}\right) \), where \( |L| = n_L \) and the input graph is bipartite. For our purposes, assume \( |L| = |R| = n \).

The approximation scheme is as follows:

- Choose parameter \( r \) (to be fixed later) and partition \( L \) into \( n/r \) subsets of size \( r \) \((L_1, \ldots, L_{n/r})\).
- For each \( L_i \), run an \( \alpha(r) \)-approximation algorithm to find a biclique cover in each subgraph induced by \( L_i \) (note each \( L_i \) is edge-disjoint).
- Each biclique from each \( L_i \) are put together and form a biclique cover of the whole graph.
  - Note that since \( L_i \) were edge-disjoint, this is also a biclique partition.
Nearly Tight Approximability for $bp$

- This scheme gives approximation guarantee $\frac{n}{r} \alpha(r)$
- Choose the $\alpha(r)$-approximation scheme as follows:
  - Given $L_i$, run a brute force algorithm over all $2^r$ subsets and enumerate all $r$-tuples of each subset
  - Such a defined subset $S$ and its intersection with the set $\{w : v \in S, w$ is a neighbor of $v\}$ induces a biclique
  - Return the smallest tuple of vertex sets which covers all edges (ensure these bicliques are edge-disjoint for $bp$)
  - An optimal solution has at most $r$ bicliques, so this returns an optimal solution (i.e., $\alpha(r) = 1$)
Nearly Tight Approximability for bp

- The running time of this algorithm is $O((2^r)^r)$
- The guarantee of the scheme is $\frac{n}{r} \alpha(r) = \frac{n}{r}$
- Choose $r = \sqrt{\log(n)}$ gives us a guarantee of $O\left(\frac{n}{\sqrt{\log(n)}}\right)$
- $r = \sqrt{\log(n)}$ gives us a polynomial runtime of $O\left(\frac{n}{r} 2^{r^2}\right) = O(n^2)$
- Chalermsook et al. also give an approximation with respect to the number of edges $m$, which has guarantee

$$O\left(\frac{m \log^2 \log m}{\log^3 m}\right)$$
Open Problems

- Does there exist a combinatorial proof for the Graham Pollak Theorem?
  - One exists using the Pigeon Hole Principle, but uses structures with size on the order of $n^n$ [Vis13]
  - Tait [Tai13] claims F.R.K. Chung confirms there exists a “better” combinatorial proof (cited via private communication)

- What is $\text{bp}_2(K_n)$?
  - Best known bounds are
    $\sqrt{n-1} \leq \text{bp}_2(K_n) \leq \lceil \sqrt{n} \rceil + \lfloor \sqrt{n} \rfloor - 2$ [Alo97, HS12]
  - Easy to ask: what is $\text{bp}_t(K_n)$ for constant $t$?
Open Problems

- Are there approximation algorithms with better guarantees?
  - Chalermsook et al. [CHHK14] give better guarantees if $\text{NP} \not\subseteq \text{BPTIME} \left(2^{\text{polylog } n}\right)$ (Bounded Error Probabilistic Time)

- How close is $bp$ to Wyner’s Common Information?
  - How good of an approximation is one to the other?

- How close are $bp$ and the biclique cover number ($bc$)?
  - Known that $bc \leq bp$
  - [Pin14] This relation may be quite loose:
    - $bp(K_n) \geq 2^{bc(K_n)} - 1 - 1$ (note that $bc(K_n) = \lceil \log n \rceil$)
    - $bp(G) \leq \frac{1}{2} \left(3^{bc(G)} - 1\right)$
The biclique partition number is a fertile, rich area of research in mathematics with many connections to other fields.

- Determining $bp$ and $bc$ is an NP-Hard problem
  - Even for bipartite graphs
- $bp$ and $bc$ are NP-Hard to approximate as well
  - Even for bipartite graphs
- Still many open problems in relation to $bp$ and $bc$


[GH06] Hermann Gruber and Markus Holzer.

On the addressing problem for loop switching.

On embedding graphs in squashed cubes.

[HS12] Hao Huang and Benny Sudakov.
A counterexample to the alon-saks-seymour conjecture and related problems.

[NMWA78] Dana S. Nau, George Markowsky, Max A. Woodbury, and D. Bernard Amos.
A mathematical analysis of human leukocyte antigen serology.


[NW95] Noam Nisan and Avi Wigderson.
On rank vs. communication complexity.

[Orl77] James Orlin.
Contentment in graph theory: Covering graphs with cliques.

[Pec84] GW Peck.
A new proof of a theorem of graham and pollak.
*Discrete mathematics, 49(3):327–328, 1984.*

Biclique covers and partitions.

The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear.


On approximate solutions for combinatorial optimization problems.

My favorite application using eigenvalues: Eigenvalues and the graham-pollak theorem.
2013.

[Tve82] Helge Tverberg.
On the decomposition of $k n$ into complete bipartite graphs.

[Vis08] Sundar Vishwanathan.
A polynomial space proof of the graham–pollak theorem.