Hardness of Computing the Biclique Partition Number

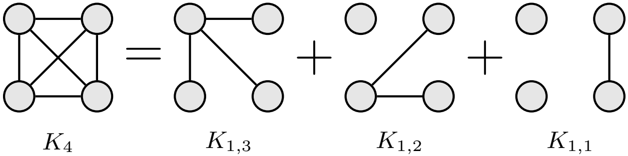
Alexander R. Block

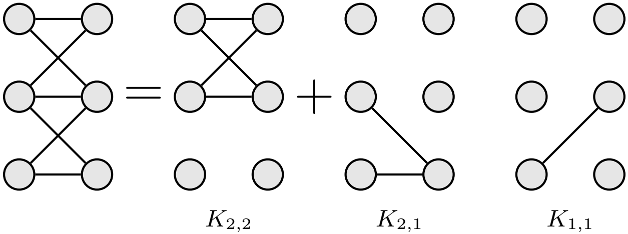
July 17, 2017

Introduction

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

Introduction (Examples)





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: $\mathsf{bp}(K_n) = (n-1)$. All proofs are algebraic and no purely combinatorial proof is known [GP72, Tve82, Pec84, Vis08, Vis13]
  - Showed $\mathsf{bp}(G) \geqslant \max\{n_+(A(G)), n_-(A(G))\}$ [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 [GP71, GP72, Tai13]
- Has applications for *perfect hashings* [Tai13]

<u>INPUT</u>: $n, r, k \in \mathbb{N}$ with $k \leqslant r \leqslant n$ <u>MINIMIZE</u>: size of $\mathcal{F} := \{f_i : [n] \to [r]\}$ <u>CONSTRAINT</u>: $\forall K \subseteq [n]$ with $\lvert K \rvert = k$, $\exists i$ such that $f_i\rvert_K$ is injective

  - Asking for *unique* i and $r = k = 2$ asks for $\mathsf{bp}\left(K_n\right)$

- Connections to the nondetermisnistic state complexity of finite automata, namely used as a lower bound method [GH06]
- Play a roll in analysis of HLA reaction matrices used in biology [NMWA78]

Applications of Biclique Partition

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

- Suppose we are given a G in the instance of the *vertex clique problem* described before and we want to answer the following question

GIVEN: Bipartite graph G
DETERMINE: Fewest number of bicliques which partition a subset $H \subseteq E(G)$

- Construct $G' = (L, R, E')$ with $L = \{x_1, \ldots, x_n\}$, $R = \{y_1, \ldots, y_n\}$ and $E' = \{(x_i, y_i) \colon \forall i\} \cup \{(x_i, y_j) \colon (v_i, v_j) \in E(G)\}$
- Let $H' = \{(x_i, y_j) \colon i = 1, \ldots, n\}$ 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'.

Proof of NP-Completeness of bp for general graphs

- Suppose we are given a graph G and need to find a vertex cover of size $k \leqslant |V(G)|$
- 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 $\mathsf{bp}(G') = \alpha(G')$, where $\alpha(G')$ is the size of the minimal vertex cover of G'
- Notice that $\alpha(G') = \alpha(G) + |E|$
- Thus, $\mathsf{bp}(G') = \alpha(G') = \alpha(G) + |E|$
- So $\alpha(G) \leqslant k$ if and only if $\mathsf{bp}(G') \leqslant k + |E|$

Approximating bp is Hard

- Since determining bp is NP-Hard, can we approximate?
- Unfortunately, bp is also NP-Hard to approximate [Sim90, BMB+08, CHHK14]
- Simon [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 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

Nearly Tight Approximability for bp

- Chalermsook, Heydrich, Holm, and Karrenbauer [CHHK14] proves an approximation algorithm for bp with approximation guarantee of $\mathcal{O}\left(n_L/\sqrt{\log(n_L)}\right)$, 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 from 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 $\{w : v \in S, w \text{ is a neighbor of } v\}$ 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 bicliqes, so this returns an optimal solution (i.e., $\alpha(r) = 1$)

Nearly Tight Approximability for bp

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

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

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

Open Problems

- Are there approximation algorithms with better guarantees?
  - Chalermsook et al. [CHHK14] give better guarantees if $\mathsf{NP} \not\subseteq \mathsf{BPTIME}\left(2^{\mathrm{polylog}\, n}\right)$ (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 $\mathsf{bc} \leqslant \mathsf{bp}$
  - [Pin14] This relation may be quite loose:
    - ★ $\mathsf{bp}\,(K_n) \geqslant 2^{\mathsf{bc}(K_n)-1} - 1$ (note that $\mathsf{bc}\,(K_n) = \lceil \log n \rceil$)
    - ★ $\mathsf{bp}\,(G) \leqslant \frac{1}{2}\left(3^{\mathsf{bc}(G)} - 1\right)$

Conclusions

- 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

[Alo97] Noga Alon.
Neighborly Families of Boxes and Bipartite Coverings, pages 27–31.
Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

[BMB+08] Doina Bein, Linda Morales, Wolfgang W Bein, CO Shields Jr, Z Meng, and Ivan Hal Sudborough.
Clustering and the biclique partition problem.
In *HICSS*, page 475, 2008.

[CHHK14] Parinya Chalermsook, Sandy Heydrich, Eugenia Holm, and Andreas Karrenbauer.
Nearly Tight Approximability Results for Minimum Biclique Cover and Partition, pages 235–246.
Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[Cio05] Sebastian M. Cioabă.
The np-completeness of some edge-partitioning problems.
Master's thesis, Queen's University, 2005.

[GH06] Hermann Gruber and Markus Holzer.

Finding Lower Bounds for Nondeterministic State Complexity Is Hard, pages 363–374.
Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[GP71] Ronald L Graham and Henry O Pollak.
On the addressing problem for loop switching.
Bell System Technical Journal, 50(8):2495–2519, 1971.

[GP72] Ronald L Graham and Henry O Pollak.
On embedding graphs in squashed cubes.
In *Graph theory and applications*, pages 99–110. Springer, 1972.

[HS12] Hao Huang and Benny Sudakov.
A counterexample to the alon-saks-seymour conjecture and related problems.
Combinatorica, 32(2):205–219, 2012.

[NMWA78] Dana S. Nau, George Markowsky, Max A. Woodbury, and D. Bernard Amos.

A mathematical analysis of human leukocyte antigen serology.
Mathematical Biosciences, 40(3):243 – 270, 1978.

[NW95] Noam Nisan and Avi Wigderson.
On rank vs. communication complexity.
Combinatorica, 15(4):557–565, 1995.

[Orl77] James Orlin.
Contentment in graph theory: Covering graphs with cliques.
In *Indagtiones Mathematicae (Proceedings)*, 1977.

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

[Pin14] Trevor Pinto.
Biclique covers and partitions.
In *Electronic Journal of Combinatorics*, 2014.

[Raz92] Alexander A Razborov.

The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear.
Discrete mathematics, 108(1):393–396, 1992.

[Sim90] Hans Ulrich Simon.
On approximate solutions for combinatorial optimization problems.
SIAM Journal on Discrete Mathematics, 1990.

[Tai13] Michael Tait.
My favorite application using eigenvalues: Eigenvalues and the graham-pollak theorem.
2013.

[Tve82] Helge Tverberg.
On the decomposition of kn into complete bipartite graphs.
Journal of Graph Theory, 6(4):493–494, 1982.

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

Journal of Combinatorial Theory, Series A, 115(4):674–676, 2008.

[Vis13] Sundar Vishwanathan.
A counting proof of the graham–pollak theorem.
Discrete Mathematics, 313(6):765–766, 2013.