The method of flag algebras

Zdeněk Dvořák

December 4, 2020

1 The method

Throughout this text, we use the abbreviations (for flags $\mathbf{F_1}, \mathbf{F_2}$ and a graph F) that we introduced in the last lesson.

$$\begin{aligned}\mathbf{F_1} &\equiv p(\mathbf{F_1}; G, \theta)\\ \mathbf{F_1} \circ \mathbf{F_2} &\equiv p(\mathbf{F_1}, \mathbf{F_2}; G, \theta)\\ F &\equiv p(F; G)\end{aligned}$$

Suppose we are trying to obtain an upper bound on the density of a certain induced subgraph D (usually K_2, corresponding to the density of edges) in an n-vertex graph G avoiding subgraphs $F_1, \ldots, F_k$ (we will write $\vec{F} \nsubseteq G$ to indicate this). To do so, we are willing to inspect all graphs with at most m vertices. As the first step, we can use Lemma 1 from the last lecture:

$$D = \sum_{A \in \mathcal{H}_m, \vec{F} \nsubseteq A} p(D; A) \cdot A. \tag{1}$$

Since $\sum_{A \in \mathcal{H}_m, \vec{F} \nsubseteq A} A = 1$, this gives

$$D \leq \max_{A \in \mathcal{H}_m, \vec{F} \nsubseteq A} p(D; A). \tag{2}$$

This bound usually is not very good; let us try to improve it.

Let σ be a type and let $\mathbf{Z_1}, \ldots, \mathbf{Z_m}$ be some flags of type σ with at most $(m + |\sigma|)/2$ vertices (often, one just takes all such flags avoiding $\vec{F}$). Let B be a symmetric positive semidefinite $n \times n$ matrix, and let $\tilde{\mathbf{Z}} = (\mathbf{Z_1}, \ldots, \mathbf{Z_m})$.

Then, using Lemmas 5, 4, and 3 from the last lecture, we have

$$\begin{aligned}
0 &\le E_\theta[\tilde{\mathbf{Z}} B \tilde{\mathbf{Z}}^T] \\
&= \sum_{i,j} B_{i,j} E_\theta[\mathbf{Z_i} \mathbf{Z_j}] \\
&= \sum_{i,j} B_{i,j} E_\theta[\mathbf{Z_i} \circ \mathbf{Z_j}] + O(1/n) \\
&= \sum_{i,j} B_{i,j} E_\theta\Big[\sum_{\mathbf{H} \in \mathcal{H}_{\sigma,m}} p(\mathbf{Z_i}, \mathbf{Z_j}; \mathbf{H}) \cdot \mathbf{H} \Big] + O(1/n) \\
&= \sum_{i,j} B_{i,j} \sum_{\mathbf{H} \in \mathcal{H}_{\sigma,m}} p(\mathbf{Z_i}, \mathbf{Z_j}; \mathbf{H}) E_\theta[\mathbf{H}] + O(1/n) \\
&= \sum_{i,j} B_{i,j} \sum_{\mathbf{H} \in \mathcal{H}_{\sigma,m}} p(\mathbf{Z_i}, \mathbf{Z_j}; \mathbf{H}) E_\theta[p(\mathbf{H}; H, \theta)] \cdot H + O(1/n) \\
&= \sum_{A \in \mathcal{H}_m, \vec{F} \not\subseteq A} \Big(\sum_{i,j} c_{i,j,A} B_{i,j} \Big) \cdot A + O(1/n),
\end{aligned}$$

where

$$c_{i,j,A} = \sum_{\mathbf{H} \in \mathcal{H}_{\sigma,m}, H \cong A} p(\mathbf{Z_i}, \mathbf{Z_j}; \mathbf{H}) E_\theta[p(\mathbf{H}; H, \theta)] \tag{3}$$

are constants independent of G and B.

We can now combine this with (1), obtaining

$$D \le \sum_{A \in \mathcal{H}_m, \vec{F} \not\subseteq A} \Big(p(D; A) + \sum_{i,j} c_{i,j,A} B_{i,j} \Big) \cdot A + O(1/n), \tag{4}$$

and thus

$$D \le \max_{A \in \mathcal{H}_m, \vec{F} \not\subseteq A} \Big(p(D; A) + \sum_{i,j} c_{i,j,A} B_{i,j} \Big) \cdot A + O(1/n). \tag{5}$$

For suitably chosen B, this may be a better bound than the one obtained from (2). Moreover, the best possible bound can be obtained by semidefinite programming, minimizing M such that

$$p(D; A) + \sum_{i,j} c_{i,j,A} B_{i,j} \le M$$

holds for every A.

Some further remarks:

- We may also obtain inequalities as above for several different types and combine them together.

- The resulting bound is for the density of D in G as an induced subgraph. In case we want the density as a subgraph, we can bound the sum $\sum_{D'} p(D'; G)$ over all supergraphs D' of D with $V(D') = V(D)$ in the same way.

- Let M be the maximum from (5), and let c_A be the coefficient at A in 4. The inequality (5) is obtained as follows: We have that $\sum_A A = 1$, and thus

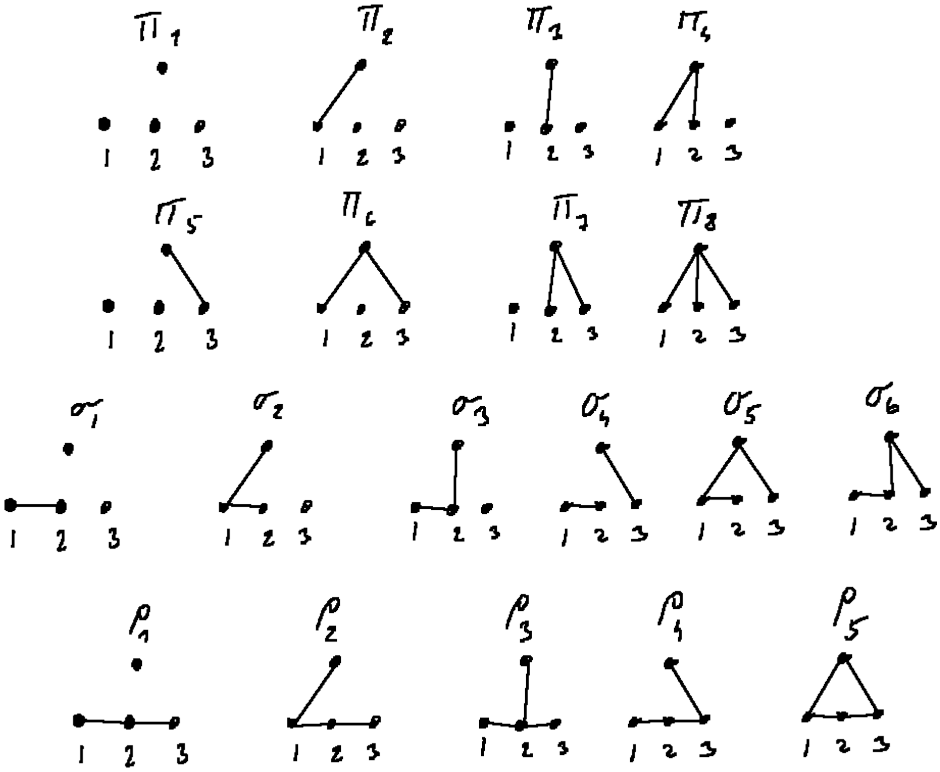
$$\sum_A c_A \cdot A = \Big(\sum_A M \cdot A\Big) - \Big(\sum_A (M - c_A) \cdot A\Big) = M - \Big(\sum_A (M - c_A) \cdot A\Big) \leq M.$$

  If for some A we had $c_A < M$ and $A \neq 0$, then we would actually get a better bound $M - (M - c_A)A < M$; hence, if the bound is (almost) tight for the graph G, we must have $p(A; G)$ very close to 0, and thus A (almost) does not appear in G. Sometimes, this enables us to determine the structure of the extremal graphs.

2 5-cycles in triangle-free graphs

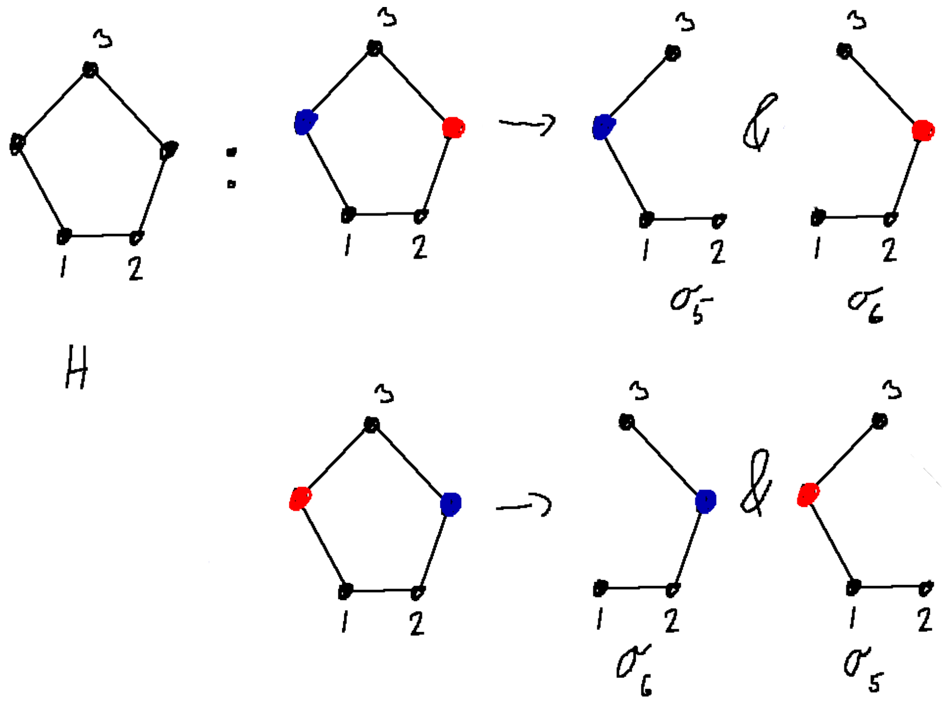
Let us now see an application based on [1]: What is the maximum number of 5-cycles in a triangle-free graph G? Note that every 5-cycle in a triangle-free graph is induced. Hence, we want to find an upper bound on $p(C_5; G)$ under the assumption that $p(K_3; G) = 0$.

We will use $m = 5$; there are 14 triangle-free graphs with 5 vertices, which is certainly small enough that (using computer) we can perform the necessary calculations. We will also use all triangle-free flags with 3 roots and 4 vertices, depicted in the following picture; note there are (up to isomorphism) 3 possible types of these flags.



Let $P, Q, R \succeq 0$ be the symmetric matrices indexed by these flags (their dimensions are 8×8, 6×6, and 5×5, respectively). Let us for example compute the contribution of $A = C_5$ to 4.

- To find the coefficient at $Q_{i,j}$, we need to go over all flags $\mathbf{H}$ whose underlying graph is A (the 5-cycle) and whose type contains the edge only between labels 1 and 2, and add $p(\sigma_{\mathbf{i}}, \sigma_{\mathbf{j}}; \mathbf{H}) E_\theta[p(\mathbf{H}; H, \theta)]$. There is only one such flag, and $E_\theta[p(\mathbf{H}; H, \theta)]$ (the probability that if we add the three distinct roots to the 5-cycle, we will obtain this flag) is $1/6$. There are then two ways how to divide the unlabelled vertices of $\mathbf{H}$ in order to obtain two induced subflags with the same roots and four vertices, giving $p(\sigma_{\mathbf{5}}, \sigma_{\mathbf{6}}; \mathbf{H}) = p(\sigma_{\mathbf{6}}, \sigma_{\mathbf{5}}; \mathbf{H}) = 1/2$. Therefore, the contribution of this type of flags is $\frac{1}{12}(Q_{5,6} + Q_{6,5}) = \frac{1}{6} Q_{5,6}$ since Q is symmetric.



- Similarly, let us compute the coefficient at $R_{i,j}$. There is again only one flag $\mathbf{H}$ whose underlying graph is the 5-cycle and the vertices with labels 1, 2, and 3 in order form a path; and $E_\theta[p(\mathbf{H}; H, \theta)] = 1/6$. The partition of the unlabelled vertices gives $p(\boldsymbol{\rho_2}, \boldsymbol{\rho_4}; \mathbf{H}) = p(\boldsymbol{\rho_4}, \boldsymbol{\rho_2}; \mathbf{H}) = 1/2$. The total contribution is $\frac{1}{6}R_{2,4}$

- The coefficient at $P_{i,j}$ is 0, since there is no flag whose underlying graph is the 5-cycle and the three roots form an independent set.

- Finally, we have $p(C_5; A) = 1$.

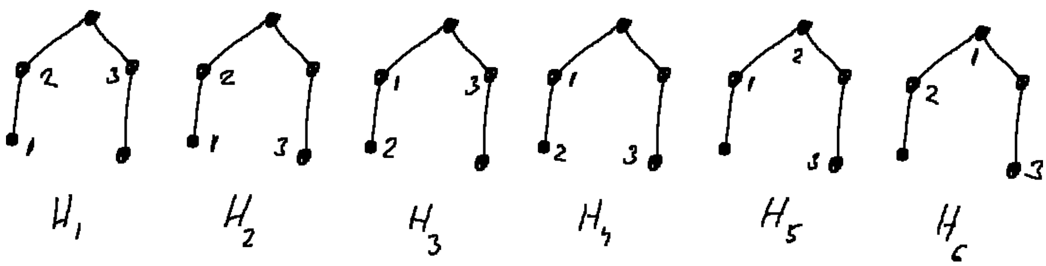
Together, this gives the coefficient $1 + \frac{1}{6}Q_{5,6} + \frac{1}{6}R_{2,4}$ at $A = C_5$.

Let us also do the calculation for $A = P_5$.

- There are three ways how to add labels 1, 2, and 3 to P_5 so that they form an independent set, depending on which label ends up on the middle vertex of the path; let $\mathbf{H_i}$ denote the one where the middle vertex has label i. For each i, we have $E_\theta[p(\mathbf{H_i}; H_i, \theta)] = 1/30$. Partitioning the remaining vertices, we obtain
  - $p(\boldsymbol{\pi_4}, \boldsymbol{\pi_6}; \mathbf{H_1}) = p(\boldsymbol{\pi_6}, \boldsymbol{\pi_4}; \mathbf{H_1}) = 1/2$
  - $p(\boldsymbol{\pi_4}, \boldsymbol{\pi_7}; \mathbf{H_2}) = p(\boldsymbol{\pi_7}, \boldsymbol{\pi_4}; \mathbf{H_2}) = 1/2$
  - $p(\boldsymbol{\pi_6}, \boldsymbol{\pi_7}; \mathbf{H_3}) = p(\boldsymbol{\pi_7}, \boldsymbol{\pi_6}; \mathbf{H_3}) = 1/2$

Together, this gives $\frac{1}{30}(P_{4,6} + P_{4,7} + P_{6,7})$.

- There are six ways how to add the labels to P_5 so that vertex labelled 1 is adjacent to the one labelled 2 and both are non-adjacent to the vertex labelled 3, depicted in the following picture.



The contributions are as follows; for each i, we have $E_\theta[p(\mathbf{H_i}; H_i, \theta)] = 1/30$, and by listing $Q_{a,b}$, we mean $p(\sigma_\mathbf{a}, \sigma_\mathbf{b}; \mathbf{H_i}) = p(\sigma_\mathbf{b}, \sigma_\mathbf{a}; \mathbf{H_i}) = 1/2$:

  - $\mathbf{H_1}$: $Q_{4,6}$
  - $\mathbf{H_2}$: $Q_{3,4}$
  - $\mathbf{H_3}$: $Q_{4,5}$
  - $\mathbf{H_4}$: $Q_{2,4}$
  - $\mathbf{H_5}$: $Q_{2,6}$
  - $\mathbf{H_6}$: $Q_{3,5}$

Together, this gives $\frac{1}{30}(Q_{2,4} + Q_{2,6} + Q_{3,4} + Q_{3,5} + Q_{4,5} + Q_{4,6})$.

- There are three ways how to add the labels to P_5 so that vertices labelled 1, 2, and 3 in order induce a path, depicted in the following picture.

The contributions are as follows; for each i, we have $E_\theta[p(\mathbf{H_i}; H_i, \theta)] = 1/30$, and by listing $R_{a,b}$, we mean $p(\rho_\mathbf{a}, \rho_\mathbf{b}; \mathbf{H_i}) = p(\rho_\mathbf{b}, \rho_\mathbf{a}; \mathbf{H_i}) = 1/2$:

  - $\mathbf{H_1}$: $R_{1,4}$

– $\mathbf{H_2}$: $R_{2,4}$

– $\mathbf{H_3}$: $R_{1,2}$

Together, this gives $\frac{1}{30}(R_{1,2} + R_{1,4} + R_{2,4})$.

- Finally, we have $p(C_5; P_5) = 0$.

Together, this gives the coefficient $\frac{1}{30}(P_{4,6} + P_{4,7} + P_{6,7} + Q_{2,4} + Q_{2,6} + Q_{3,4} + Q_{3,5} + Q_{4,5} + Q_{4,6} + R_{1,2} + R_{1,4} + R_{2,4})$ at $A = P_5$.

We conclude that

$$\begin{aligned} C_5 \le\ & \tfrac{1}{6}(6 + Q_{5,6} + R_{2,4})C_5 \qquad (6)\\ & + \tfrac{1}{30}(P_{4,6} + P_{4,7} + P_{6,7} + Q_{2,4} + Q_{2,6} + Q_{3,4} + Q_{3,5} + Q_{4,5} + Q_{4,6} + R_{1,2} + R_{1,4} + R_{2,4})P_5 \\ & + \ldots + O(1/n), \end{aligned}$$

where the other terms are computed similarly. Hence,

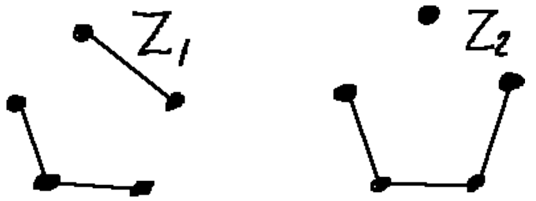
$$\begin{aligned} C_5 \le \max\Big(& \tfrac{1}{6}(6 + Q_{5,6} + R_{2,4}), \qquad (7)\\ & \tfrac{1}{30}(P_{4,6} + P_{4,7} + P_{6,7} + Q_{2,4} + Q_{2,6} + Q_{3,4} + Q_{3,5} + Q_{4,5} + Q_{4,6} + R_{1,2} + R_{1,4} + R_{2,4}),\\ & \ldots\Big) + O(1/n). \end{aligned}$$

We can now use semidefinite programming to obtain the best possible choice of the matrices P, Q, and R. If we want to just obtain some bound, we can directly use the matrices returned by an SDP solver (making sure that they are indeed positive semidefinite, and not slightly off due to rounding errors). In order to obtain an exact result, these matrices need to be rounded to exact rational numbers (again, making sure they still stay positive semidefinite), which may be tricky. The resulting matrices for our problem can be found in [1]. After substituting them to (7), we obtain

$$C_5 \le \max\big(\tfrac{24}{625}, -\tfrac{126}{6250}, \ldots\big) + O(1/n),$$

with all other terms smaller or equal to $\frac{24}{625}$. Therefore, an n-vertex triangle-free graph contains at most $\frac{24}{625}\binom{n}{5} + O(n^4)$ 5-cycles.

Moreover, in the extremal graphs, we must have $p(P_5; G) = o(1)$, as otherwise (6) would give us a better bound. Similarly, the evaluation of the other terms shows that we have coefficients smaller than 24/625 at the following graphs,



implying these induced subgraphs (almost) do not appear in the extremal graphs. This enables us to get a stability result. You can argue (e.g., using the flag algebra method again) that if $p(K_3; G) = 0$ and $p(P_4; G) = o(1)$, then $p(C_5; G) \ll 24/625$; hence, the density of P_4's in any extremal graph is positive, and since $p(Z_2; G) = o(1)$ and $p(P_5; G) = o(1)$, most of these P_4's do not extend to Z_2 or P_5. Hence, for such an induced path $P = v_1v_2v_3v_4$, all but $o(n)$ vertices x of G have $N(x) \cap V(P) \neq \emptyset, \{v_1\}, \{v_4\}$. Since G is triangle-free, we can divide such vertices into those adjacent to v_2 (denoted by A_1), to v_1 and v_3 (denoted by A_2), to v_2 and v_4 (denoted by A_3), to v_3 (denoted by A_4), and to v_1 and v_4 (denoted by A_5). Since G is triangle-free, $A_1 \cup A_3$, $A_2 \cup A_4$, $A_2 \cup A_5$ and $A_3 \cup A_5$ are independent sets. Moreover, a similar density argument (using the fact that $p(Z_2; G) = o(1)$) implies that there can be $\Omega(n^2)$ edges between A_1 and A_4 only for $o(n^4)$ choices of the path P, and thus there exists a choice for P with $o(n^2)$ edges between A_1 and A_4. Therefore, we conclude that every near-extremal graph is close to a blowup of a 5-cycle, in the following sense: There exists a partition A_1, $\ldots$, A_5 of its vertex set such that $\|G[A_i \cup A_{i+2}]\| = o(n^2)$ for $i \in \{1, \ldots, 5\}$, where $A_6 = A_1$ and $A_7 = A_2$.

Finally, let us remark that we can obtain the exact bound (achieved by blowing up each vertex of C_5 to an independent set of size $n/5$, giving a graph with $(n/5)^5$ 5-cycles): Suppose G is a triangle-free graph with n vertices and cn^5 5-cycles. Let G' be the graph obtained from G by blowing up each vertex into an independent set of k vertices (turning edges of G into complete bipartite subgraphs in G'). Clearly, G' is also triangle-free. Moreover, G' has nk vertices and cn^5k^5 5-cycles, and thus $cn^5k^5 \leq (24/625 + o(1))\binom{nk}{5} = (1/5^5 + o(1))n^5k^5$. Hence, $c \leq 1/5^5 + o(1)$ as $k \to \infty$, and thus $c \leq 1/5^5$.

References

[1] Andrzej Grzesik: *On the maximum number of five-cycles in a triangle-freegraph*, Journal of Combinatorial Theory, Series B **102** (2012), 1061-1066.