The non-crossing graph

Nathan Linial *, Michael Saks †, David Statter ‡

December 13, 2004

Abstract

Two sets are non-crossing if they are disjoint or one contains the other. The *non-crossing graph* NC_n is the graph whose vertex set is the set of nonempty subsets of $[n] = \{1, \ldots, n\}$ with an edge between any two non-crossing sets.

Various facts, some new and some already known, concerning the chromatic number, fractional chromatic number, independence number, clique number and clique cover number of this graph are presented. For the chromatic number of this graph we show:

$$n(\lceil \log_2 n \rceil - 1) \geq \chi(NC_n) \geq n(\log_e n - \Theta(1)).$$

1 Introduction

Two sets are non-crossing if they are disjoint or one contains another. The *non-crossing graph* NC_n is the graph whose vertex set is the set of nonempty subsets of $[n] = \{1, \ldots, n\}$, and whose edge set is $\{XY : X, Y \text{ are non-crossing}\}$. Note that the subgraph of NC_n induced on $\{S \subseteq [n] : |S| = k\}$, which we denote here by $NC_n[k]$ is the well known Kneser graph (see, e.g. [6, 3]).

*School of Computer Science and Engineering, Hebrew University, Jerusalem 91904, Israel. E-mail: nati@cs.huji.ac.il. Work supported in part by a grant from the Israel Science Foundation.

†Dept. of Mathematics Rutgers University New Brunswick, NJ. E-mail: saks@math.rutgers.edu.

‡Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel. E-mail: statter@cs.huji.ac.il.

For $1 \leq j \leq k \leq n$ $NC_n[j,k]$ denotes the subgraph of NC_n induced on $\{S \subseteq [n] : j \leq |S| \leq k\}$. For $W \subseteq [n]$, $NC_n[W]$ denotes the subgraph induced on the set of subsets whose size lies in W.

In this note, we collect some facts, some new and some previously known, about various basic parameters associated with these graphs: the chromatic number, fractional chromatic number, independence number, clique number and clique cover number.

A good reason for studying these graphs is that this may be a first step towards understanding the following abstract simplicial complex. A set $\{X_1, \ldots, X_k\}$ of subsets of $[n]$ is a *shattering family* if for each choice of $Y_1, \ldots, Y_k$ with $Y_i \in \{X_i, [n] - X_i\}$ we have $\cap_{i=1}^{k} Y_i \neq \emptyset$. This is a standard definition in machine learning [7] which first emerged out of fundamental work of Vapnik and Chervonenkis [11], and has become of great interest in combinatorics, especially discrete geometry [5]. Shattering is a hereditary property, i.e., any subfamily of a shattering family is shattering and so the set of all shattering families of subsets of $[n]$ is an abstract simplicial complex SF_n with vertex set $2^{[n]} - \{\emptyset\}$. Restricting attention to the 1-skeleton of this complex, i.e., the set of to shattering families of size 2 (*shattering pairs*) yields a graph SG_n on vertex set $2^{[n]} - \{\emptyset\}$. This graph has the property that vertices corresponding to complementary subsets are nonadjacent and have identical neighborhoods. Thus the structure of SG_n is completely captured by restricting the vertex set to subsets of $[n-1]$. The resulting graph is the complement of the graph NC_{n-1} that we are studying here.

2 Independence number

A set of vertices in NC_n is an independent set iff it is an *intersecting antichains*. Namely, a collection of subsets with no one containing another and no two disjoint. These are well-studied in the hypergraph literature. Two fundamental results are:

Lemma 2.1. *(Milner [9]) The largest intersecting antichain of $[n]$ has size $\binom{n}{\lfloor \frac{n}{2} \rfloor + 1}$. Equality holds uniquely for the set of $\lfloor \frac{n}{2} \rfloor + 1$-sets of $[n]$.*

Lemma 2.2. *(see [1]) Let $\mathcal{F} \subseteq 2^{[n]}$ be an intersecting antichain and for $1 \leq k \leq n$, let f_k be the number of subsets of $\mathcal{F}$ of size k. Then:*

$$\sum_{k \leq n/2} \frac{f_k}{\binom{n-1}{k-1}} \leq 1. \tag{1}$$

Equality holds iff $\mathcal{F}$ consists of all k-tuples that contain a given element for some $k \le \lceil \frac{n}{2} \rceil - 1$,

These lemmas imply the following bounds on the independence number of NC_n and its induced subgraphs:

Theorem 2.3. *Let $n \ge k \ge 1$,*

1. $\alpha(NC_n) = \binom{n}{\lfloor \frac{n}{2} \rfloor + 1}$.

2. *If $k > n/2$ then* $\alpha(NC_n[1,k]) = \binom{n}{\lfloor \frac{n}{2} \rfloor + 1}$.

3. *If $k \le n/2$ then* $\alpha(NC_n[1,k]) = \binom{n-1}{k-1}$.

Proof. The first two parts are just Theorem 2.1. The third part, which was noted in [1], is a corollary of Theorem 2.2, i.e., for $k \le n/2$, $f_1 + \ldots + f_k \le \binom{n-1}{k-1}$. □

3 Fractional chromatic number

Recall that a *fractional coloring* of a graph G is a nonnegative function λ on the independent sets such that for each vertex v the sum of $\lambda(I)$ over I containing v is at least 1. The *weight* of λ is defined to be $\sum_I \lambda(I)$. The *fractional chromatic number* $\chi^*(G)$ is the minimum weight of a fractional coloring. A *fractional clique* of G is a nonnegative function f on the vertices, satisfying that the sum of $f(v)$ over any independent set is at most 1. The *weight* of f is $\sum_{v \in V} f(v)$. The *fractional clique number* $\omega^*(G)$ is the maximum weight of any fractional clique. Linear programming duality implies that $\omega^*(G) = \chi^*(G)$.

For positive integer t, let $H(t) = \sum_{i=1}^{t} \frac{1}{i}$. Also $parity(t)$ is 1 if t is odd and 0 otherwise.

Theorem 3.1. *Let $1 \le k \le n$. Then:*

1. *For $1 \le k \le n/2$, $\chi^*(NC_n[1,k]) = nH(k)$.*

2. *For $n/2 < k \le n$, $nH(\lfloor n/2 \rfloor) \le \chi^*(NC_n[1,k]) \le (n+1)H(\lfloor n/2 \rfloor) - H(n-k) + parity(n)$.*

In particular, $nH(\lfloor n/2 \rfloor \le \chi^*(NC_n) \le (n+1)H(\lfloor n/2 \rfloor) + parity(n)$.

Proof. For the first part, suppose $k \le n/2$. To prove the upper bound we combine the well-known fractional colorings of Kneser graphs. Let $T_i^{(j)}$ be the independent set consisting of all the j-tuples that contain the element i. We define the fractional coloring λ of $NC_n[1,k]$ by placing weight $\frac{1}{j}$ on $T_i^{(j)}$ for each i and $1 \le j \le k$ and zeros elsewhere. It is easy to verify that λ is a proper fractional coloring of total weight $nH(k)$.

For the lower bound, Lemma 2.2 implies that the function f on subsets S of $[n]$ of size at most $n/2$ given by $f(S) = 1/\binom{n-1}{|S|-1}$ if $|S| \le n/2$ is a fractional clique. The restriction of this clique to $NC_n[1,k]$ has weight $nH(k)$.

For the second part of the theorem, assume $k > n/2$. The lower bound is the same as the lower bound for $k = \lfloor n/2 \rfloor$. For the upper bound, define $S_i^{(j)}$, for $j \le n/2$ and $i \in [n]$ to be the union of $T_i^{(j)}$ and the set of subsets of size $n-j+1$ that don't contain i. $S_i^{(j)}$ is independent. If we put weight $1/j$ on each $S_i^{(j)}$ for each $1 \le j \le n/2$ and $i \in [n]$ then the total weight on any set of size at most $n/2$ is 1, while the total weight on a set of size $j > (n+1)/2$ is $(n-j)/(n-j+1)$, and in the case that n is odd, there is no weight on sets of size $(n+1)/2$. To get a fractional coloring of $NC_n[1,k]$, it suffices to put weight $1/(n-j+1)$ on the independent set consisting of all j element sets of n, for each $(n+1)/2 < j \le k$ and when n is even, weight 1 on the independent set consisting of all $(n+1)/2$ element sets. The total weight of this coloring is the upper bound stated in the theorem.

□

4 Chromatic Number

For any graph G, the chromatic number $\chi(G)$ is at least $\chi^*(G)$, so Theorem 3.1 implies $\chi(NC_n) \ge nH(\min\{k, n/2\}) = n \ln n + O(n)$. Here we prove:

Theorem 4.1. *Let $1 \le k \le n$ be an integer. Then $\chi(NC_n[1,k]) \le \lceil \log_2(k+1) \rceil n$. In particular, $\chi(NC_n) \le n \lfloor \log_2(n+1) \rceil$.*

The theorem follows easily from the following lemma:

Lemma 4.2. *Let $W \subseteq [n]$ where all elements of W are at least $|W|$. Then $\chi_n(NC_n[W]) \le n$.*

Proof. We describe a coloring of $NC_n[W]$ with color set $[n]$. Let $w_1 < \ldots < w_{|W|}$ denote the elements of W. For $A \subseteq [n]$ with $|A| \in W$, let $i(A)$ be the index such that $|A| = w_{i(A)}$. For $a \in A$, the *rank of a in A*, $r_A(a)$, is defined to be $|A \cap [1, a]|$. Color A by the element $c(A)$ of rank $w_1 + 1 - i(A)$ in A; $w_1 + 1 - i(A) \geq 1$ since $w_1 \geq |W| \geq i(A)$. To see that c is a proper coloring, let $S \subseteq T$ with $|S|, |T| \in W$. Then:

$$r_T(c(T)) = w_1 + 1 - i(T) < w_1 + 1 - i(S) = r_S(c(S)) \leq r_T(c(S)),$$

which implies $c(T) \neq c(S)$. □

As an immediate corollary we get:

Corollary 4.3. *If k and n are positive integers with $2k - 1 \leq n$, then*

$$\chi(NC[k, 2k-1]) \leq n.$$

Proof of Theorem 4.1. Given $1 \leq k \leq n$, let $m = \lceil \log_2(k+1) \rceil$ and partition the set $[1, k]$ into sets $W_1, \ldots, W_m$ with $W_i = [2^{i-1}, 2^i - 1]$ for $i < m$ and $W_m = [2^{m-1}, k]$. By Corollary 4.3 each of the graphs $NC_n[W_i]$ can be colored with n colors so $NC_n[1, k]$ can be colored with at most mn colors.

Theorems 3.1 and 4.1 show that $n \ln n + O(n) \leq \chi(NC_n) \leq n \log_2 n + O(n)$. This gives a constant factor gap between the upper and lower bounds.

Problem 4.4. *What is the asymptotic behavior of $\chi(NC_n)$?*

We believe that the upper bound is the truth. The result in Theorem 4.1 can be improved slightly for some values of k, by choosing a better coloring of the final group W_m of vertices but this only affects the constant in the $O(n)$ term.

One possible way to improve the upper bound more significantly would be to improve Corollary 4.3. This result can be improved somewhat if k is very close to $n/2$, but for k sufficiently less than $n/2$ we conjecture that such an improvement is not possible:

Conjecture 4.5. *For every k and n with $4k \leq n$, $\chi(NC_n[k, 2k-1]) = n$.*

The condition $4k \leq n$ is somewhat arbitrary here; some condition is needed to avoid k being too close to $n/2$.

We can prove two special cases of this conjecture. Say that a coloring of subsets is *elemental* if the color assigned to each subset is an element of the subset. (For example, the coloring in the proof of Lemma 4.2 is elemental.)

Theorem 4.6. *For k, n with $2k-1 \leq n$, every proper elemental coloring of $NC_n[k, 2k-1]$ uses n colors.*

Theorem 4.7. $\chi(NC_n[2,3]) = n$ *for every* $n \geq 6$.

Proof of Theorem 4.6. For $A \subseteq [n]$ we denote by $NC_A[k, 2k-1]$, the induced subgraph of $NC_n[k, 2k-1]$ restricted to subsets of A.

Fix an elemental coloring c of $NC_n[k, 2k-1]$ We will show that for each A of size $2k-1$, the restriction of c to the subgraph $NC_A[k, 2k-1]$ uses all colors in A; it follows immediately that c uses all n colors.

To this end we will prove a stronger claim, namely that for each $i \in [0, k-1]$ and S of size $k+i$, the restriction of c to $NC_S[k, 2k-1]$ uses at least $2i+1$ colors. The basis $i = 0$ is trivial. Suppose S has size $k+i$ with $i \geq 1$. Let $\alpha = c(S)$ and $\beta = c(S - \{\alpha\})$. By induction, the restriction of c to $NC_{S-\{\beta\}}[k, 2k-1]$ uses at least $2i-1$ colors. These colors can't include α or β, so $NC_{S-\{\beta\}}[k, 2k-1]$ uses at least $2i+1$ colors. □

Proof of Theorem 4.7. The proof is by induction on n. For the basis case $n = 6$, let c be an optimal coloring of $NC_6[2,3]$. For $T \subseteq [6]$ of size 3 let $\mu(T)$ denote the induced subgraph whose vertex set consists of the six pairs contained in T or $\overline{T}$. Since T and $\overline{T}$ are adjacent to each other and to each vertex of $\mu(T)$ it suffices to show that for some T, c uses 4 colors on $\mu(T)$ since then c must use at least 6 colors.

Call T *good* if $\mu(T)$ has fewer then 4 colors. If T is good then either all three pairs in T are the same color or all three pairs in $\overline{T}$ are the same color. The color used for these three pairs can't be used for any other pair. Since there are $\frac{1}{2}\binom{6}{3} = 10$ different subgraphs $\mu(T)$, if all T are good then we need at least 10 colors. Thus some T is not good, completing the basis case.

For the induction step, let $n \geq 7$. If we can show that there is a color class C and element i such that every set in C contains i, then we can restrict the vertex set to $[n] - \{i\}$ and finish by induction. We may assume that at most $n-1$ colors are used for pairs. Since $n \geq 7$, $\binom{n}{2} > 3(n-1)$, so there is a color class with at least 4 pairs. It is easy to see that a color class of pairs and triples that includes at least 4 pairs must have an element i that belongs to all of the sets so we are done. □

5 The clique number

A clique in NC_n is a collection of sets such that any two are either disjoint or comparable under inclusion. Such a family is sometimes called *laminar* in the literature. It is well known and easy to show that in a laminar collection $\mathcal{C}$ of sets that does not contain the empty set, the relation $X \longrightarrow Y$ if $Y \subset X$ and there is no $Z \in \mathcal{C}$ with $Y \subset Z \subset X$ is a collection of directed rooted trees whose roots are maximal sets and whose leaves are minimal sets of $\mathcal{C}$. From this it is easy to deduce:

Proposition 5.1. *For any $1 \leq k \leq n$, $\omega(NC_n[1,k]) = 2n - \lceil n/k \rceil$, In particular, $\omega(NC_n) = 2n$*

The case $k = n$ is exercise 11.19 in [8]. The general case is similar.

Proof. For the lower bound, we may assume that the union of all of the sets in $\mathcal{C}$ is $[n]$. The number of nodes in a directed forest is at most $2L - R$ where L is the number of leaves and R is the number of roots. The maximal sets of $\mathcal{C}$ partition $[n]$ so there are at least $\lceil n/k \rceil$ of them and the minimal sets are disjoint so there are at most n of them, so the lower bound follows.

For the upper bound, choose a partition $A_1, \ldots, A_r$ of $[n]$ with $r = \lceil n/k \rceil$ and each A_i at most k. Let $\mathcal{C}_i$ be a maximal chain of sets having largest set A_i. Take $\mathcal{C}$ to be the union of $\mathcal{C}_i$ together with the remaining $n-r$ singletons. $\square$

6 Clique cover number

Recall that $\theta(G)$ is the minimum number of cliques needed to cover the vertices of G. We'll need the following well known fact (which follows easily from the construction of de Bruijn, Tengburgen and Kruyswijk about the existence of a symmetric chain decomposition of $2^{[n]}$ (see, e.g., [4]).

Lemma 6.1.
1. *For $W \subseteq [1, \lfloor n/2 \rfloor]$, it is possible to partition $\{A \subseteq [n] : |A| \in W\}$ into chains (under inclusion) such that the largest set of each chain has size k.*
2. *For $W \subseteq [\lfloor n/2 \rfloor + 1, n]$, it is possible to partition $\{A \subseteq [n] : |A| \in W\}$ into chains (under inclusion) such that the smallest set of each chain has size k.*

3. *For $j, k \in [n]$, with $\binom{n}{j} \leq \binom{n}{k}$ there is a matching from the j-sets to the k-sets such that each matched pair is comparable under inclusion.*

The following proposition reduces the problem of computing $\theta(NC_n[W])$ for arbitrary W to the case that $|W|$ is a singleton or is a two element set with one element $\leq n/2$ and the other $> n/2$.

Proposition 6.2. *Let $W \subseteq [n]$ be nonempty.*

- *If $W \subseteq [1, \lfloor n/2 \rfloor]$ then $\theta(NC_n[W]) = \theta(NC_n[\{w_{max}\}]$, where w_{max} is the largest element of W.*
- *If $W \subseteq [\lfloor n/2 \rfloor + 1, n]$ then $\theta(NC_n[W]) = \theta(NC_n[\{w_{min}\}]$, where w_{min} is the smallest element of W.*
- *Otherwise, let W' be the set consisting of the largest element of $W \cap [1, \lfloor n/2 \rfloor]$ and the smallest element of $W \cap [1 + \lfloor n/2 \rfloor, n]$ Then $\theta(NC_n[W]) = \theta(NC_n[W']$.*

Proof. For the first part, let Π be a clique partition of $NC_n[\{w_{max}\}]$. By Lemma 6.1 we can cover all sets whose size belongs to W by chains whose largest member has size w_{max}. For each clique $C \in \Pi$, let C' be the union of the chains in the cover whose maximal set is in C. C' is a clique and the collection Π' of all such cliques is a cover of $NC_n[W]$.

The proofs of the second and third parts are obtained by the obvious modification of the above proof. □

In light of this proposition, to analyze $\theta(NC_n[W])$ for arbitrary W it suffices to consider the case that W is a singleton or is a two element set with one element $\leq n/2$ and the other $> n/2$.

Theorem 6.3. *Let $1 \leq j \leq n/2 < k \leq n$. Then:*

1. $\theta(NC_n[\{k\}]) = \binom{n}{k}$.
2. *If j divides n then* $\theta(NC_n[\{j\}]) = \binom{n-1}{j-1}$.
3. *If $n/3 < j < n/2$ then* $\theta(NC_n[\{j\}]) = \lceil \frac{1}{2}\binom{n}{j} \rceil$.
4. *If $j + k \leq n$ then* $\theta(NC_n[\{j, k\}]) = \binom{n}{k}$.

Proof. The first part is trivial. The second part is precisely Baranyai's theorem [10]. The third part is the fact that $NC_n[\{k\}]$ has a matching that misses at most one set. This is a simple corollary from Tutte's matching theorem [2]. Details of the proof can be found in Problem 12.16 in [8]. The last part follows from the last part of Lemma 6.1. $\square$

7 Quo vadis

It seems very interesting to consider higher-dimensional analogues of the questions considered here. We recall the general notion of shattering as mentioned in the Introduction, and consider the problem of largest independent set in the d-dimensional case. This translates into the following question: Given n and d determine the smallest m such that every $m \times n$ $0,1$ matrix with distinct rows contains a $d \times 2^d$ block with no repeated columns. The same problem with $2^d \times d$ block with no repeated rows is answered by a classical theorem of Sauer. Indeed using this very same theorem and a very simple construction, it is easy to provide bounds where the upper bound is about quadratic in the lower bound. It seems reasonable that the lower bound $m \geq \Omega(n^{2^d-1})$ is closer to the truth, but this does not seem easy.

References

[1] Béla Bollobás, *Combinatorics*, Cambridge University Press, Cambridge, 1986.

[2] Reinhard Diestel, *Graph theory*, Graduate Texts in Mathematics, vol. 173, Springer-Verlag, New York, 2000.

[3] Chris Godsil and Gordon Royle, *Algebraic graph theory*, Graduate Texts in Mathematics, vol. 207, Springer-Verlag, New York, 2001.

[4] C. Greene and D.J. Kleitman, *Strong versions of Sperner's theorem*, J. Combinatorial Theory (A) **20** (1976), 80–88.

[5] Matoušek Jiří, *Lectures on discrete geometry*, Graduate Texts in Mathematics, vol. 212, Springer-Verlag, New York, 2002.

[6] ______, *Using the Borsuk-Ulam theorem*, Universitext, Springer-Verlag, Berlin, 2003.

[7] Michael J. Kearns and Umesh V. Vazirani, *An introduction to computational learning theory*, MIT Press, Cambridge, MA, 1994.

[8] László Lovász, *Combinatorial problems and exercises*, North-Holland Publishing Co., Amsterdam, 1993.

[9] E.C. Milner, *A combinatorial theorem on systems of sets*, J. London Math. Soc. **43** (1968), 204–206.

[10] J. H. van Lint and R. M. Wilson, *A course in combinatorics*, Cambridge University Press, Cambridge, 2001.

[11] V. N. Vapnik and A. Ja. Červonenkis, *The uniform convergence of frequencies of the appearance of events to their probabilities*, Teor. Verojatnost. i Primenen. **16** (1971), 264–279.