**Graphs, Capacities, Entropies**
lecture notes, Fall 2022

*First lecture* (September 6, 2022)

# 1 Introduction

Information theory deals with the theoretical limits of information transmission. In those cases when a negligible error is allowed on the decoding side, its theorems are of probabilistic nature. If, however, we insist on zero-error decoding then combinatorial problems emerge. Such conisderations gave rise to several interesting combinatorial concepts. A famous example for such a notion is the Shannon capacity of graphs introduced by Claude Shannon [54] in 1956. We start our discussion by introducing this invariant.

## 1.1 Shannon capacity of graphs

Trivial question: How many different $t$-length sequences can be given over the alphabet $\{0, 1\}$? The answer is trivially $2^t$.

Another trivial question: How many different $t$-length sequences can be given over the alphabet $[n] := \{1, 2, \dots, n\}$? The answer is obviously $n^t$.

Still very easy question: How many different (that is, pairwise distinguishable) $t$-length sequences can be given over the alphabet $[4] = \{1, 2, 3, 4\}$ if we can only distinguish between odd and even numbers, that is 1 and 3, as well as 2 and 4 "looks the same" when we want to "decode" our sequences? The answer is again $2^t$: it does not matter which odd and which even numbers we use, the two odd values as well as the two even values can be identified.

Easy question: How many different (that is, pairwise distinguishable) $t$-length sequences can be given over the alphabet $\{h, k, \ell, n, x\}$ if while $x$ can be distinguished from all the other four letters, among those four only the pairs $kn$, $n\ell$ and $h\ell$ can be distinguished, that is $h$ can be confused with $n$ as well as $h$ with $k$ and $k$ with $\ell$? (Think about handwritten letters to feel it more realistic.) Now the answer is $3^t$ since we do not lose anything if we always substitute $h$ with $n$ and $k$ with $\ell$. But $x, n$ and $\ell$ are pairwise distinguishable so all the sequences using only them will be fine.

Very difficult question: How many different (that is, pairwise distinguishable) $t$-length sequences can be given over the alphabet $\{0, 1, 2, 3, 4\}$ if the pair $a, b \in \{0, 1, 2, 3, 4\}$ can be distinguished if and only if $a - b \equiv 1$ or $-1$ modulo 5?

Before explaining why this last question is difficult, let us formulate the problem generally.

We are given a graph $G = (V, E)$, where the vertices, that is the elements of $V$ represent the letters in an alphabet. The meaning of the edges is that the two endpoints can surely be distinguished, while the lack of an edge between two vertices means that the corresponding two vertices can be confused. We are interested in the maximum number of $t$-length sequences that are pairwise distinguishable. Let us denote this number by $M_G(t)$.

**Observation 1**

$$M_G(t) \geq [\omega(G)]^t.$$

*Proof.* Indeed, let $Q \subseteq V$ be a subset of the vertex set that induces a largest clique. Then the $|Q|^t = [\omega(G)]^t$ possible sequences using letters only from $Q$ are all pairwise distinguishable. □

It is obvious that $M_G(t) \leq |V|^t$ so it makes sense to normalize by taking the $t^{\text{th}}$ root and investigate $\sqrt[t]{M_G(t)}$. In fact we are interested in the asymptotics of this value, so consider its lim sup as $t$ goes to infinity. (In fact, the limit always exists.)

Now we are going to define the above asymptotic value in purely graph theoretic terms. This needs that the distinguishability of sequences is expressed in graph theory language. To this end we introduce a graph product.

**Definition 1** *For two graphs $F$ and $G$ their OR-product $F \cdot G$ is defined by*

$$V(F \cdot G) = V(F) \times V(G)$$

*and*

$$E(F \times G) = \{\{(f,g),(f',g')\} : f,f' \in V(F), g,g' \in V(G), \\ ff' \in E(F) \text{ or } gg' \in E(G)\}.$$

*The $t^{\text{th}}$ OR-power if a graph $G$ is meant to be the $t$-fold OR-product of $G$ with itself.*

Note that $G^t$ extends the distinguishability relation from individual letters to $t$-length sequences: the vertices of $G^t$ are exactly $t$-length sequences over the alphabet $V(G)$ and two such sequences are adjacent in $G^t$ if and only if there is at least one position where their corresponding entries are distinguishable, that is, they form an edge of $G$. This means that the maximum number of pairwise distinguishable such sequences is just the clique number of $G^t$.

With the above notion at hand we can now define the Shannon capacity of a graph $G$.

**Definition 2** *The Shannon OR-capacity of a graph $G$ is defined as*

$$C_{\text{OR}}(G) := \limsup_{t\to\infty} \sqrt[t]{\omega(G^t)}.$$

Two remarks are in order. The first is that due to what is called Fekete's Lemma, the above *limsup* is always attained as a *limit*. The second is that choosing to represent the distinguishability relation of our pairs of letters to be represented by the edges of a graph is a quite arbitrary choice: we could have chosen just the opposite and say that edges mean confusability and then we get the complementary graph to describe the exact same situation. Our approach has several advantages but also a disadvantage, namely that traditionally the complementary approach was followed (rather for psychological than for mathematical reasons, I believe). Below we briefly introduce also the complementary notion that involves a different graph product, called AND-product. To make it always clear which language is used the complementary notion will be called Shannon AND-capacity.

**Definition 3** *The AND product of two graphs $G$ and $H$ is given by*

$$V(G \wedge H) = V(G) \times V(H)$$

*and*

$$E(G\wedge H) = \{(g,h)(g',h') : g,g' \in V(G), h,h' \in V(H), (gg' \in E(G) \text{ and } hh' \in E(H)) \\ \text{or } (g = g' \text{ and } hh' \in E(H)) \text{ or } (gg' \in E(G) \text{ and } h = h')\}.$$

*We denote by $G^{\wedge t}$ the $t$-fold AND product of $G$ with itself.*

Note that if the edges of $H$ mean confusability of its endvertices then the edges of $H^{\wedge t}$ can be interpreted as the confusability of the sequences forming their endpoints. Thus the largest number of pairwise distinguishable (=non-confusable) $t$-length sequences can be expressed by the independence number $\alpha(H^{\wedge t})$.

**Definition 4** *The Shannon AND-capacity of a graph $H$ is defined as*

$$C_{\mathrm{AND}}(H) := \limsup_{t\to\infty} \sqrt[t]{\alpha(H^{\wedge t})}.$$

It is easy to verify by the foregoing that

$$C_{\mathrm{AND}}(H) = C_{\mathrm{OR}}(\overline{H}),$$

where $\overline{H}$ is the complementary graph of $H$.

We have already seen that $M_G(t) \geq [\omega(G)]^t$. In graph terms this means $\omega(G^t) \geq [\omega(G)]^t$ that immediately implies that

$$C_{\mathrm{OR}}(G) \geq \omega(G)$$

for every graph $G$. Next we give a general upper bound for $C_{\mathrm{OR}}(G)$. Recall that the chromatic number $\chi(F)$ of a graph $F$ is the minimum number of colors with which the vertices of $F$ can be colored so that adjacent vertices receive different colors.

**Theorem 1** *For every graph $G$ we have*

$$C_{\mathrm{OR}}(G) \leq \chi(G).$$

To prove this statement we will prove the following two lemmas.

**Lemma 1** *Let $\varphi(G)$ be a graph parameter satisfying the following two conditions:*

1. *$\omega(G) \leq \varphi(G)$;*
2. *$\varphi$ is submultiplicative with respect to the OR-product, that is*

$$\varphi(F \cdot G) \leq \varphi(F)\varphi(G)$$

*holds for any finite simple graphs $F$ and $G$.*

*Then*

$$C_{\mathrm{OR}}(G) \leq \varphi(G)$$

*for all finite simple graphs $G$.*

**Lemma 2** *The chromatic number satisfies the two conditions in Lemma 1.*

It is straightforward that the above two lemmas imply Theorem 1

*Proof of Lemma 1.* Let $\varphi$ be a parameter as in the statement. Then we can write

$$C_{\mathrm{OR}}(G) = \lim_{t\to\infty} \sqrt[t]{\omega(G^t)} \leq \lim_{t\to\infty} \sqrt[t]{\varphi(G^t)} \leq \lim_{t\to\infty} \sqrt[t]{[\varphi(G)]^t} = \varphi(G).$$

Here the two inequalities follow, respectively, from the first and the second properties of $\varphi$ given in the statement. □

*Proof of Lemma 2.* The first property is obvious: all vertices of a largest clique must get distinct colors in any proper coloring. To prove the second property consider an optimal coloring $c : V(F) \to [\chi(F)]$ of $G$ and an optimal coloring $h : V(G) \to [\chi(G)]$. We give a proper coloring $r$ of $F \cdot G$ using pairs of colors from these two proper colorings in the folowing way. Let $r : V(F \cdot G) \to [\chi(F)\chi(G)]$ be defined by

$$r : (f, g) \mapsto (c(f), h(g))$$

for all $(f, g) \in V(F \cdot G)$. This is indeed a proper coloring since if $\{(f, g), (f', g')\} \in E(F \cdot G)$ then by definition we either have $ff' \in E(F)$ or $gg' \in E(G)$. In the first case we must have $c(f) \neq c(f')$, in the second case we must have $h(g) \neq h(g')$. In either case $(c(f), h(g)) \neq (c(f'), h(g'))$ thus our coloring is proper. We have used $\chi(F)\chi(G)$ colors implying that $\chi(F \cdot G) \leq \chi(F)\chi(G)$ as stated. $\square$

As already said above, with proving Lemmas 1 and 2 we also proved Theorem 1.

Now we can return to the question we declared to be very difficult above. One can observe that answering it we determine $C_{\mathrm{OR}}(C_5)$, where $C_n$ denotes the cycle on $n$ vertices. In all our questions preceding this one we dealt with graphs $G$ satisfying $\omega(G) = \chi(G)$ and thus by the chain of inequalities

$$\omega(G) \leq C_{\mathrm{OR}}(G) \leq \chi(G)$$

the value of the Shannon OR-capacity simply coincides with the common value of the clique number and the chromatic number. For $C_5$, however this is not the case as $\omega(C_5) = 2 < 3 = \chi(C_5)$. (This in itself does not necessarily mean that determining $C_{\mathrm{OR}}(C_5)$ should be very difficult but it turned out to be so.) In fact, $C_{\mathrm{OR}}(C_5)$ is indeed strictly more than 2. This follows from the observation that the five 2-length sequences

$$00$$

$$12$$

$$24$$

$$31$$

$$43$$

form a clique in $C_5^2$ thus implying

$$\omega(C_5^2) \geq 5.$$

It follows from this that $\omega(C_5^{2k}) \geq 5^k$ for any positive integer $k$ (simply use the 2-length blocks forming a 5-element clique in $C_5^2$) implying that $C_{\mathrm{OR}}(C_5) \geq \sqrt{5}$. Our best upper bound so far is given by the chromatic number of $C_5$ which is 3.

Now we leave the problem of the exact value of $C_{\mathrm{OR}}(C_5)$ open for the time being and turn to discuss an important influence of the Shannon capacity problem on graph theory.

# 2 Perfect graphs

The fact that $\chi(G) = \omega(G)$ implies that $C_{\mathrm{OR}}(G)$ coincides with the common value of these two parameters made the French mathematician Claude Berge elaborate on the question, what can one say about graphs whose chromatic number equals their clique number. (see [5] about the details of how his ideas about perfect graphs developed). He realized that to obtain an interesting structural property one better requires $\chi(G') = \omega(G')$ for all the induced subgraphs of $G$ as well. (Otherwise we could always just add a complete graph on $\chi(G)$ vertices to $G$ without really changing its inner structure yet obtaining equality between the chromatic number and the clique number.)

Thus Berge defined the following class of graphs.

**Definition 5** *A graph $G$ is called perfect if $\chi(G') = \omega(G')$ holds for all of its induced subgraphs.*

It turns out that several interesting classes of graphs are perfect.

*Second lecture* (September 9, 2022)

## 2.1 Classes of perfect graphs

It turns out that several interesting classes of graphs are perfect. Let us see some examples. In the following cases we always talk about such graph classes that are hereditary in the sense that if $G$ belongs to the class and $G'$ is an induced subgraph of $G$ then $G'$ also belongs to the class. Therefore in these cases it will be enough to show that $\chi(G) = \omega(G)$ for any member of the class, that will immediately imply the similar inequality also for the induced subgraphs.

**Proposition 1** *Bipartite graphs are perfect.*

*Proof.* This is quite obvious. If a bipartite graph $G$ has at least one edge then we have $\chi(G) = 2 = \omega(G)$, if it has no edge, then we have $\chi(G) = \omega(G) = 1$. With the remark in the first paragraph of this subsection this completes the proof. □

**Proposition 2** *The complementary graph of any bipartite graph is perfect.*

*Proof.* This follows from Kőnig's theorem stating that we have $\tau(G) = \nu(G)$ for all bipartite graphs. Indeed, let $G$ be a graph as in the statement on $n$ vertices. Let $M$ be a matching of size $\nu(\overline{G})$ in the complementary graph $\overline{G}$ and color the pairs of vertices matched by $M$ the same color while giving a unique color to the rest of the vertices. This way (denoting the number of vertices by $n$) we use

$$\nu(\overline{G}) + n - 2\nu(\overline{G}) = n - \nu(\overline{G}) = n - \tau(\overline{G}) = \alpha(\overline{G}) = \omega(G)$$

colors for a proper coloring, so $\chi(G) \leq \omega(G)$. (The second equality above follows from Kőnig's theorem, the third one from Gallai's identity, the last one is obvious.) Since the reverse inequality always holds, with the remark in the first paragraph of this subsection the proof is completed. □

**Proposition 3** *Line graphs of bipartite graphs are perfect.*

*Proof.* By the remark in the first paragraph of this subsection this statement is equivalent to the equality of the edge-chromatic number $\chi_e(G)$ to the largest degree $\Delta(G)$ for any bipartite graph $G$, that is also a theorem of Kőnig. To prove it we simply have to realize that every regular bipartite graph (possibly containing parallel edges) contains a perfect matching (this is easy by checking Hall's condition) and that any bipartite graph $G$ can be extended by the addition of vertices and edges to a $\Delta$-regular bipartite graph where $\Delta = \Delta(G)$ (by also using parallel edges if necessary). □

**Proposition 4** *The complementary graph of the line graph of any bipartite graph is perfect.*

*Proof.* Note that if $G$ is the complementary graph of the line graph of a bipartite graph $F$, then $\omega(G) = \nu(F)$. Thus by the remark in the first paragraph of this subsection it will be enough to show that we can color $G$ with $\nu(F) = \tau(F)$ colors, where the last equality is again from Kőnig's theorem and the fact that $F$ is bipartite. Consider a set of size $\tau(F)$ of vertices of $F$ whose elements cover all edges in $F$. Observe that the edges of $F$ (that are vertices in $G$) incident to the same vertex of $F$ form an independent set in $G$, thus are allowed to receive the same color. Therefore we can indeed color $G$ with as many colors as many vertices of $F$ are needed to cover all edges in $F$ showing $\chi(G) \leq \tau(F) = \nu(F) = \omega(G)$. As the reverse inequality is also true the proof is completed. □

The perfectness of interval graphs was proven in the basic courses on combinatorics. All interval graphs are chordal, so this is a special case of the fact that chordal graphs are perfect. This statement can be found as Proposition 5.5.2 in Diestel's book [14] together with its proof.

The complementary graphs of interval graphs are special cases of comparability graphs (see their definition below) that are also perfect as we prove next.

**Definition 6** *Let $(P, <)$ be a partially ordered set (a poset). Its comparability graph $G = G_{(P,<)}$ is the graph with $V(G) = P$ and $E(G) = \{uv : u <_P v\}$. A graph $G$ is called a comparability graph if there exists a poset for which $G$ is its comparability graph.*

**Theorem 2** *Comparability graphs are perfect.*

*Proof.* Consider a comparability graph $G$ and orient its edges according to the relation "$<$" in the poset it represents, that is for every edge $uv$ take the orientation $u \to v$ if $v < u$ and $v \to u$ if $u < v$. Let $t$ denote the number of vertices in a longest oriented path of the so obtained oriented graph. Consider such a longest path. Observe that since between any two of its vertices we have an oriented path, any two of these vertices must be connected by the transitivity property of our partial order. (If $u < v$ and $v < w$, then $u < w$, so $uw$ is also an edge.) Therefore $t \leq \omega(G)$. Now we show that $G$ can be colored by $t$ colors. Color all those vertices with color 1 that have outdegree 0 in our oriented version of $G$. Clearly, they form an independent set, so using the same color for all of them is valid. Now delete these vertices and color with color 2 all those vertices that have outdegree 0 in the remaining graph. Delete these, too, and continue the process until all vertices are colored (and thus deleted). This way we get a proper coloring with, say, $m$ colors. We show that $m \leq t$, that is, there exists an oriented path of length $m$. To show this consider an arbitrary vertex colored with color $m$. This vertex must have an outneighbor colored $(m-1)$, otherwise it would have already been colored one step earlier. By the same logic, this outneighbor must have an outneighbor of color $(m-2)$, that one must have an outneighbor of color $(m-3)$, etc. These vertices then give us an oriented path with one vertex in all the $m$ color classes, therefore we indeed have $m \leq t$ as claimed. But then we have $\chi(G) \leq m \leq t \leq \omega(G)$ and since $\omega(G)$ cannot be larger than $\chi(G)$ we proved that $\chi(G) = \omega(G)$. Since all induced subgraphs of a comparability graph are also comparability graphs, this proves that comparability graphs are perfect. □

To mention some imperfect (=not perfect) graphs we note that odd cycles of length at least 5 are not perfect: they have clique number 2 and chromatic number 3. They are actually minimal imperfect: any proper induced subgraph of them is bipartite, therefore perfect. One can also observe that the complementary graphs of these odd cycles are also imperfect: $\omega(\overline{C_{2k+1}}) = k$, but since any color class can have at most two elements (whenever $k > 1$), their chromatic number is at least (in fact, exactly) $\lceil \frac{2k+1}{2} \rceil = k + 1 > k$. Since the complementary graphs of bipartite graphs are perfect, these graphs are also minimally imperfect.

## 2.2 The perfect graph conjectures

The above examples may suggest that perfect graphs go in pairs: whenever a graph is perfect, so is its complement. Berge formulated it as a conjecture, it became known as the Perfect Graph Conjecture. It was proven about ten years later by Lovász.

**Perfect Graph Theorem.** (Lovász 1973) The complementary graph of a perfect graph is perfect.

Berge also formulated a stronger conjecture, the Strong Perfect Graph Conjecture, stating that no other minimal imperfect graph exists, than the ones mentioned above. By now this is also a theorem, but in this case it took several decades until it was proven.

**Strong Perfect Graph Theorem.** (Chudnovsky Robertson, Seymour, Thomas 2006): A graph not containing a chordless odd cycle of length at least 5 or its complement is perfect.

Below we will see a proof of the Perfect Graph Theorem (PGT). In case of the Strong Perfect Graph Theorem (SPGT) it would be hopeless to prove it in class within a reasonable time, the article containing this proof is more than 150 pages long.

## 2.3 Vertex packing and fractional vertex packing polytopes

We will prove the PGT in a form that involves also some other notions we will need later. We introduce them here.

**Definition 7** *The vertex packing polytope $VP(G)$ of a graph $G$ is the convex hull of the characteristic vectors of its independent sets.*

**Definition 8** *The fractional vertex packing polytope $FVP(G)$ of a graph $G$ on $n$ vertices is the set of non-negative vectors in $R^n$ whose entries belonging to the vertices of a clique sum up to at most 1. Formally, denoting the set of cliques by $Q(G)$, this means*

$$FVP(G) = \{\mathbf{b} \in R^{|V(G)|} : \forall i \ b_i \geq 0, \forall B \in Q(G) \sum_{i \in B} b_i \leq 1\}.$$

Since an independent set can intersect a clique in at most one vertex, the characteristic vectors of independent sets satisfy the inequalities in the definition of $FVP(G)$. From this it is easy to see that

$$VP(G) \subseteq FVP(G)$$

for every graph $G$. It turns out that the two polytopes coincide if and only if the graph is perfect. Below we prove this statement and the PGT in one proof. This approach follows the one in the book [43].

## 2.4 Proof of the Perfect Graph Theorem

As indicated above we will prove the following theorem that combines independent results due to Chvátal [9], Fulkerson [19] and Lovász [35].

**Theorem 3** *The following four statements are equivalent for a graph $G$.*
1. *$G$ is perfect;*
2. *$\overline{G}$ is perfect;*
3. *$VP(G) = FVP(G)$;*
4. *$VP(\overline{G}) = FVP(\overline{G})$.*

*Third lecture* (September 13, 2022)

The proof will use an important lemma that was proved by Lovász [35] when he proved the Perfect Graph Theorem. We say that a vertex $v$ in graph $G$ is substituted by an edge if we add a new vertex $v'$ to $G$ adjacent to $v$ and exactly those other vertices that are also adjacent to $v$.

**Lemma 3** (Substitution Lemma) *Substituting a vertex $v$ by an edge in a perfect graph $G$ preserves its perfectness.*

For the *proof* of the above lemma see it as Lemma 5.5.5 in Diestel's book. (Note that the book uses the term "expanding a vertex" for what we called "substituting a vertex by an edge".)

Observe that for any positive integer $s$ applying the Substitution Lemma subsequently $s-1$ times for the same vertex of a perfect graph shows that if the vertex is "substituted by a clique of size $s$" the resulting graph is still perfect. (By substituting a vertex $v$ with a clique $K_s$ we mean that we add $s-1$ new vertices that are all adjacent to each other and the original vertex $v$ as well as to the original neighbors of $v$ and there are no other new edges.) We may also think about deleting a vertex as substituting it with a clique with 0 vertices. Perfectness is also preserved in this case by definition.

*Proof of Theorem 3.* By the symmetry of the statement it is enough to prove that the first property implies the third one and the third property implies the second one. This is what we will do.

1.⇒ 3.: Assume $G$ is perfect and consider an arbitrary $x \in FVP(G)$. We will show that $x \in VP(G)$ holds, too. We may assume that (every coordinate of) $x$ is rational (otherwise we could approximate it with rationals and repeat the proof that way). Let $N$ be a common multiple of the denominators of the coordinates of $x$ and write all coordinates in the form of $\frac{m_i}{N}$. Note that these values are attached to the vertices of $G$, $\frac{m_i}{N}$ belonging to $v_i$. Now substitute each vertex $v_i$ by a clique of size $m_i$ to obtain a new graph $G'$. (As remarked above this can be done by subsequent substitutions with edges, so by the Substitution Lemma $G'$ is also perfect.) Observe that $G'$ has clique number $N$ thus it can be colored by $N$ colors. Consider such an $N$-coloring and consider the independent sets that form the color classes. Now "collapse" these independent sets back to independent sets of $G$: if an independent set $A$ of $G'$ contains a vertex which is one of the copies of $v_i \in V(G)$ obtained during the substitution process, then (and only then) we include $v_i$ in the collapsed version of $A$. This way each $v_i \in V(G)$ will appear in exactly $m_i$ collapsed independent sets. (It may happen that two different color classes $A$ and $B$ of $G'$ collapse to the same independent set. Therefore each collapsed independent set is taken with multiplicity: if it is the collapsed version of $k$ color classes then we consider it $k$ times.) Putting weight $\frac{1}{N}$ on each of the color classes of $G'$, we distribute altogether unit weight since we used $N$ independent sets for coloring $G'$. Now putting $\frac{1}{N}$ weight also on the collapsed version of each of these independent sets (with multiplicity, thus if the same collapsed independent set appears as a result of $k$ color classes of $G'$, then it receives weight $\frac{k}{N}$ altogether) we again distribute exactly $N \cdot \frac{1}{N} = 1$ total weight. Taking these weights as coefficients of a convex combination of the characteristic vectors of the respective independent sets of $G$ we obtain the vector $x$ as a convex combination of characteristic vectors of independent sets, thus as an element of $VP(G)$.

3.⇒ 2.: It is easy to see that property 3. is hereditary, that is, once it is true for graph $G$ then it also holds for all induced subgraphs of $G$. Therefore it is enough to prove that this property implies $\chi(\overline{G}) = \omega(\overline{G})$, the analogous equality will similarly follow for all induced subgraphs.
Assume $VP(G) = FVP(G)$, we prove $\chi(\overline{G}) = \omega(\overline{G})$ by induction, that is we also assume that this already follows if we have fewer vertices than in $G$. Consider the points of $VP(G)$ whose coordinates sum to the independence number $\alpha(G)$. These points are on a hyperplane and they are maximal points of $VP(G)$, so they form a facet of $VP(G)$. By $VP(G) = FVP(G)$ the facets of this polytope are described by the inequalities given in the definition of $FVP(G)$. Since the origin is not on this facet, this facet must be described by one of those inequalities given by a clique $K \in Q(G)$. Since the characteristic vectors of largest, that is $\alpha(G)$-size independent sets satisfy that the sum of their coordinates is $\alpha(G)$, they are all on this facet, thus they all must have an intersection point with $K$. But then using $K$ as a color class in $\overline{G}$ we find that $\overline{G}$ can be colored with just one more colors than $\overline{G} \setminus K$ and that the latter graph has clique number (the independence number of $G \setminus K$) 1 smaller. From here the proof can be completed by the induction hypothesis which implies that $\overline{G} \setminus K$ can be colored by $\omega(\overline{G} \setminus K) = \omega(\overline{G}) - 1$ colors. □

After the proof of PGT András Hajnal suggested that perhaps the following stronger statement (that is still weaker than the Strong Perfect Graphs Conjecture) could be proven. Lovász [36] found a proof of this statement as well which appeared still in the same year as [35].

**Theorem 4** (Lovász [36]) *A graph is perfect if and only if the following inequality holds for all of its induced subgraphs $G'$.*

$$\alpha(G')\omega(G') \geq |V(G')|.$$

Note that the only if part of Theorem 4 is obvious: every graph $F$ must satisfy $\alpha(F)\chi(F) \geq |V(F)|$ (simply because all color classes in a proper coloring are independent sets therefore cannot have more than $\alpha(F)$ vertices), thus if $G$ is perfect then every induced subgraph $G'$ must satisfy $\alpha(G')\omega(G) = \alpha(G')\chi(G') \geq |V(G')|$. So the real content of the theorem is the reverse implication.

*Fourth lecture* (September 16, 2022)

Gasparian [23] found a proof of this statement that is significantly different from that of Lovász in [36]. In particular, unlike Lovász's proof, it does not use the Substitution Lemma. Both proofs are elegant, in fact, the one by Lovász probably gives more "explanation" for why the statement is actually true. Yet, Gasparian's proof is so much different and surprising (partly by its unexpected use of linear algebra) that I chose to present that in class. This proof can be found in Diestel's book [14] as the proof of Theorem 5.5.6 there.

Note that Theorem 4 has the following immediate consequence (that by now, of course, also follows from the SPGT which completely describes all minimally imperfect graphs.

**Corollary 5** *If $G$ is minimally imperfect, that is, an imperfect graph whoch becomes perfect by the deletion of any of its verices, then*

$$|V(G)| = \alpha(G)\omega(G) + 1.$$

# 3 Fractional chromatic and clique number

**Definition:** Denoting the set of independent sets of a graph $G$ by $S(G)$ a *fractional coloring* is a function (a weighting) $w : S(G) \to R_{+,0}$, such that

$$\forall v \in V(G): \quad \sum_{A \ni v, A \in S(G)} w(A) \geq 1.$$

The *fractional chromatic number* $\chi_f(G)$ is the value

$$\inf \sum_{A \in S(G)} w(A)$$

taken under the above conditions. Instead of inf one can write min as the infimum is always attained.

The above can be formulated as a linear program as follows: Let $A$ be a matrix with $n := |V(G)|$ rows and $s := |S(G)|$ columns in which the columns are the characteristic vectors of the independent sets. This means that $A[i,j] = 1$ if vertex $v_i \in A_j$, where $A_j$ denotes the $j^{\text{th}}$ independent set, and $A[i,j] = 0$ otherwise. Then $\chi_f(G) = \min(\boldsymbol{c} \cdot \boldsymbol{x})$, where $\boldsymbol{c} = (1, \ldots, 1)$ is the $s$-dimensional all-1 vector and the minimization is under the constraints

$$A\boldsymbol{x} \geq \boldsymbol{b}$$

for the $n$-dimensional all-1 vector $\boldsymbol{b} = (1, \ldots, 1)^T$ and

$$\boldsymbol{x} \geq \boldsymbol{0}.$$

All the inequalities are meant coordinatewise.

*Fifth lecture* (September 20, 2022)

A linear program as the above has a dual:

$$\boldsymbol{y}A \leq \boldsymbol{c}, \ \boldsymbol{y} \geq \boldsymbol{0},$$

and we seek

$$\max(\boldsymbol{y} \cdot \boldsymbol{b})$$

under these constraint. With the above matrix a vector $\boldsymbol{y}$ satisfying the constraints is a non-negative weighting of the vertices such that the total weight in any independent set is at most 1. Such a weighting is called a *fractional clique* and the maximum possible total weight (that is the maximum of $\boldsymbol{y} \cdot \boldsymbol{b}$) is the *fractional clique number* $\omega_f(G)$. (The name comes from the fact that a fractional clique which has only integer coordinates is necessarily the characteristic vector of a clique.)

By the Duality Theorem of linear programming

$$\chi_f(G) = \omega_f(G).$$

This is a minimax theorem as any *feasible solution* for the first linear program gives an upper bound for the value given by any feasible solution of the dual program:

$$\boldsymbol{y} \cdot \boldsymbol{b} \leq \boldsymbol{y}(A\boldsymbol{x}) = (\boldsymbol{y}A)\boldsymbol{x} \leq \boldsymbol{c} \cdot \boldsymbol{x}.$$

Therefore if we present a fractional coloring and a fractional clique for a graph giving the same value, then they are necessarily optimal.

It should be clear from the foregoing that we always have

$$\omega(G) \leq \omega_f(G) = \chi_f(G) \leq \chi(G),$$

where the equality in the middle is just the already stated consequence of the Duality Theorem. The inequality $\omega_f(G) \geq \omega(G)$ follows from the fact that the characteristic vector of any clique gives a fractional clique (as no clique and independent set can have more than 1 vertex in common). Similarly, the inequality $\chi_f(G) \leq \chi(G)$ follows by observing that any proper coloring gives rise to a fractional coloring by attaching weight 1 to the independent sets that are color classes and weight 0 to the rest. Thus if $\chi(G) = \omega(G)$, in particular, if $G$ is perfect, then $\chi_f(G)$ also equals to the common value of these two parameters.

## 3.1 Fractional chromatic number and zero-error capacity

Now we show that $\chi_f(G)$ satisfies the conditions in Lemma 1 therefore it is also an upper bound on the Shannon OR-capacity of $G$. The condition $\omega(G) \leq \chi_f(G)$ we have already seen, what still needs proof is the submultiplicativity with respect to the OR-product. In fact, more is true: the fractional chromatic number is just multiplicative with respect to this product.

**Lemma 4** *For any two finite simple graphs $F$ and $G$ we have*

$$\chi_f(F \cdot G) = \chi_f(F)\chi_f(G).$$

*Sketch of proof.* Consider optimal fractional colorings $f : S(F) \to R_{+,0}$ and $g : S(G) \to R_{+,0}$. Observing that for $A \in S(F), B \in S(G)$ we have $A \times B \in S(F \cdot G)$ we can define a fractional coloring $h : A \times B \mapsto f(A)g(B)$ and do similarly for optimal fractional cliques $w : V(F) \to R_{+,0}$ and $z : V(G) \to R_{+,0}$ by assigning the vertex $(u, v) \in V(F \cdot G)$ weight $w(u)z(v)$. Calculating the total weight distributed this way we get that both

$$\chi_f(F \cdot G) \leq \chi_f(F)\chi_f(G)$$

and

$$\omega_f(F \cdot G) \geq \omega_f(F)\omega_f(G)$$

hold proving the statement. □

Combining Lemmas 1 and 4 we immediately get

**Theorem 6** *For any graph $G$ we have*

$$C_{\mathrm{OR}}(G) \leq \chi_f(G).$$

In particular, the above gives $C_{\rm OR}(C_5) \leq \frac{5}{2}$. (Recall that our best lower bound so far was $\sqrt{5}$. The upper bound $\frac{5}{2}$ was already known by Shannon. In [54] one can find a completely different argument that also leads to the inequality of Theorem 6 which we present next.

A noisy channel $\mathcal{C}$ is usually described by a stochastic matrix $W_{\mathcal{C}}$ (meaning that every row is a probability distribution) where the rows are indexed by the input characters $x^{(1)}, \ldots, x^{(n)}$, the columns by the output characters $y^{(1)}, \ldots, y^{(k)}$ and entry $W_{\mathcal{C}}[j, i]$ is the probability of receiving $y^{(j)}$ at the output when input letter $x^{(i)}$ is sent.

From this matrix the distinguishability graph $G_{\mathcal{C}}$ of the channel is obtained as

$$V(G_{\mathcal{C}}) = \{x^{(1)}, \ldots, x^{(n)}\} \; E(G_{\mathcal{C}}) = \{x^{(j)}x^{(k)} : \; \forall i \; W_{\mathcal{C}}[j,i] \cdot W_{\mathcal{C}}[k,i] = 0\}.$$

Then the zero-error capacity $C_0(\mathcal{C}) = \log_2 C_{\rm OR}(G_{\mathcal{C}}) = \limsup_{t\to\infty} \frac{1}{t} \log_2 \omega(G^t)$. (Taking the logarithm is traditional in information theory because this way we measure the amount of transmitted information in bits.)

It is customary in information theory to also consider communication with feedback. In case of our channel model it means that after sending a character via the noisy channel $\mathcal{C}$ the sender is informed of which output it resulted in and should decide about the next character to be sent only after this information is received. If $M_{\mathcal{C},f}(t)$ denote the number of messages one can transmit using $t$ transmitted characters with zero probability of error, then the zero-error capacity with feedback for this channel is

$$C_{0,f}(\mathcal{C}) = \limsup_{t\to\infty} \frac{1}{t} \log_2 M_{\mathcal{C},f}(t).$$

It is easy to observe that if $C_0(\mathcal{C}) = 0$ (this happens exactly when all pairs of input letters can be confused, that is $G_{\mathcal{C}}$ has no edges), then we also have $C_{0,f}(\mathcal{C}) = 0$. This is simply because any two sequences of input characters can result in the same output irrespective of whether we know what was received before. If, however, $C_0(\mathcal{C}) > 0$ then we may have

$$C_{0,f}(\mathcal{C}) > C_0(\mathcal{C})$$

as we will soon see.

A primary observation is that if feedback is allowed then even from the zero-error capacity point of view the channel is not fully described by the distinguishability graph $G_{\mathcal{C}}$. Consider for example the following two channels given by the matrices $W_1$ and $W_2$, both with input and output alphabets identical to $\{0, 1, 2, 3\}$. Let us have $W_1(0, 0) = W_2(0, 0) = 1$, i.e., the input character 0 results in a 0 output character with probability 1 in both channels (implying that $W_1(i, 0) = W_2(i, 0) = 0$ for every $i \in \{1, 2, 3\}$). For input characters $i, j$ different from 0 we have

$$W_1(i,j) = \frac{1}{3},$$

that is, in the first channel these inputs result in any of the other nonzero inputs with (the same but this is irrelevant) positive probability and with probability 1 they do not result in output 0. For the second channel we have

$$W_2(1,1) = W_2(2,1) = W_2(2,2) = W_2(3,2) = W_2(3,3) = W_2(1,3) = \frac{1}{2}$$

while all other transition probabilities are zero. That is here all three nonzero characters can become only two of the other three nonzero characters at the output with positive probability. Observe that the distingushability graph is a star with three edges where the center of the star is the vertex representing input letter 0 in both cases. We will see, however that the zero-error capacity with feedback is different for the two channels. Indeed, when we send input letter 1 and get the feedback that the receiver has received 2, then in case of the first channel the receiver knows only that one of the three nonzero characters was sent while in case of the second channel the receiver knows that either 1 or 2 was sent. So the uncertainty of the receiver in the second case is smaller and since we (the sender) know this, we can exploit this when continuing the transmission. It still does not matter what are the actual values of the positive probabilities in our channel matrices but it does matter which output character can be the result of sending a certain input character (which is more information than just saying that two input characters can or cannot result in the same output.) So the channel is now described by a hypergraph: the vertices are the input letters and each hyperedge represents an output letter containing exactly those vertices that represent input letters which may result with positive probability in the output letter that the edge represents. By formula this hypergraph $H_{\mathcal{C}}$ is given as

$$V(H_{\mathcal{C}})) = \{x^{(1)}, \dots, x^{(n)}\},$$
$$E(H_{\mathcal{C}}) = \{U \subseteq V(H_{\mathcal{C}}) : \; \exists y^{(i)} \; W_{\mathcal{C}}[y^{(i)}, x^{(j)}] > 0 \Leftrightarrow x^{(j)} \in U\}.$$

*Sixfth lecture* (September 23, 2022)

Using the above defined hypergraph we can already state the theorem proved by Shannon in [54] expressing the zero-error capacity of a noisy channel with feedback.

**Theorem 7** *Let $\mathcal{C}$ be a noisy channel for which hypergraph $H_{\mathcal{C}}$ describes the possible transitions of input letters to output letters the way described above. Define the value $P_0$ as*

$$P_0 = \min_P \max_{U_j \in E(H_{\mathcal{C}})} \sum_{x^{(i)} \in U_j} P(x^{(i)}),$$

*where the minimization is over all $P$ that are probability distributions on the input alphabet and the inner maximization is over the edges $U_j$ of $H_{\mathcal{C}}$ that belong to the output letters. Then*

$$C_{0,f}(\mathcal{C}) = \begin{cases} \log_2 \frac{1}{P_0} & \textit{if } C_0(\mathcal{C}) > 0 \\ 0 & \textit{if } C_0 = 0. \end{cases}$$

*Sketch of proof.* The $C_0 = 0$ case was already discussed above, so we have to prove only the first line of the above formula.

First we prove that the right hand side is an upper bound. Assume this is not the case, that is, $C_{0,f} > \log_2 \frac{1}{P_0}$. Then there must be some positive integer $t$ (in fact, infinitely many) such that $\frac{M_{\mathcal{C},f}(t)}{M_{\mathcal{C},f}(t-1)} > \frac{1}{P_0}$, that is the maximum number of different messages that can be transmitted without error if feedback is used by transmitting $t-1$ input letters is less than $P_0$ times the maximum number of different messages one can transmit under similar conditions by sending $t$ input characters. Consider such a $t$ and the $M_{\mathcal{C},f}(t)$ different messages. For each of them the first letter to be sent is given by some optimal encoding strategy. The fraction of all possible messages starting with the individual input characters gives a probability distribution on the input alphabet. Whatever this distribution is, the definition of $P_0$ ensures that there must be some output character that can be resulted by the first character of at least $P_0 M_{\mathcal{C},f}(t) > M_{\mathcal{C},f}(t-1)$ of our messages. If this output character is received, then there are more than $M_{\mathcal{C},f}(t)$ messages that the receiver should be able to distinguish just by seeing the result of the remaining $t-1$ transmissions. By the definition of $M_{\mathcal{C},f}(t-1)$ this is impossible. (Even though the sender knows by the feedback exactly which are these remaining messages, their number is too large.) This contradiction proves that $C_{0,f} \leq \log_2 \frac{1}{P_0}$

To prove the reverse inequality consider the probability distribution $(p_1, \ldots, p_n)$ on the input that attains the value $P_0$. When the number $M$ of possible messages is huge, we can attach to almost exactly $p_i M$ of them the input letter $x^{(i)}$ as a first character. Sending the first letter via the channel according to this choice, the definition of $P_0$ ensures that whatever the received character will be, the number of messages that remain consistent with that is not more than $P_0 M$. Then we can repeat this process for these messages and keep doing so until the number of still possible messages goes below some (possibly still quite large) constant. Then those remaining messages can be encoded by a constant length zero-error code whenever there are at least two never confusable characters in the input aplphabet. (This is where we use that $C_0(\mathcal{C}) > 0$.) This completes the proof. □

Some remarks are in order.

1. Notice that for the two channels $W_1$ and $W_2$ in our example above we obtain different values for their zero-error capacity with feedback. Both of them has a pair of always distinguishable pair of input letters so their zero-error capacity is positive, we just have to determine the correspomding $P_0$ values. In case of $W_1$ this value is clearly $1/2$ attained when the probability of input character 0 is $1/2$ and the other $1/2$ probability is distributed arbitrarily on the other three input characters. For $W_2$ we can do better by giving probability $2/5$ to input character 0 and $1/5$ to each of the other three input characters. This way the probability of every output character will be only $2/5 < 1/2$ Thus the zero-error capacity with feedback for the first channel is $\log_2 2 = 1$ while for the second channel it is $\log_2(5/2) > 1$.

2. Our second remark shows that if $G$ is the distinguishability graph of some channel, then its zero-error capacity with feedback is at least $\log_2 \chi_f(G)$. Moreover, there exists a channel with this distinguishability graph whose capacity is exactly $\log_2 \chi_f(G)$. Let $S^*(G)$ denote the set of maximal (that is, non-extendable, not necessarily largest) independent sets of $G$. It should be clear from the definition of $P_0$ in Theorem 7 that if $G_{\mathcal{C}}$ is isomorphic to $G$ for some channel $\mathcal{C}$, then $\mathcal{C}$ has minimal zero-error capacity if $E(H_{\mathcal{C}}) = S^*(G)$. Consider this case. Put the probabilities attaining $P_0$ on the vertices and multiply all of them by $P_0^{-1}$. Observe that the definition of $P_0$ ensures that now the total probability on each independent set is at most 1, that is we obtained a fractional clique. In fact, the definition of $P_0$ also guarantees that this is a maximal fractional clique and its value is $P_0^{-1}$. Thus $\chi_f(G) = P_0^{-1}$ and its logarithm is the value of the zero-error capacity with feedback for our channel. Note that since feedback cannot decrease the capacity (we could simply ignore it to simulate its non-existence), the above argument gives a new proof for $C_0(\mathcal{C}) \leq \chi_f(G_{\mathcal{C}})$.

*Seventh lecture* (September 27, 2022)

# 4 Lovász theta number

Recall that for the Shannon capacity of the pentagon graph $C_5$ our best bounds so far give

$$\sqrt{5} \leq C_{\mathrm{OR}}(C_5) = C_{\mathrm{AND}}(C_5) \leq \frac{5}{2}.$$

The equality in the middle here follows simply by the self-complementary property of $C_5$. These bounds were already proven in Shannon's 1956 paper [54]. The situation has changed only more than two decades later when Lovász published his celebrated paper [39] in which the graph paraemeter in the title of this section was introduced. He showed that this new parameter is also an upper bound on Shannon capacity and its value for $C_5$ happens to be $\sqrt{5}$ proving the sharpness of the lower bound.

At this point we have to warn the reader that the Lovász theta number is originally introduced in such a way that it bounds $C_{\mathrm{AND}}$ from above and this is adopted in most publications, see e.g., [29]. Exceptions also exist though, e.g. the complementary definition is used in [40] and later an equivalent notion was also introduced under the name vector chromatic number (or strict vector chromatic number) by Karger, Motwani and Sudan [27]. (The equivalence is proven e.g. in [42].) The complementary type definition is fitting better for our discussion so we will use that. To distinguish it clearly from the more traditional Lovász theta number, usually denoted by $\vartheta(G)$ we will use the notation $\bar{\vartheta}(G)$ and so we have $\bar{\vartheta}(G) = \vartheta(\bar{G})$. With slight abuse of the terminology we will also refer to this number as the Lovász theta number, but when we want to emphasize that it is $\bar{\vartheta}(G)$ rather than $\vartheta(G)$ then we will use the term *complementary theta number* or *OR-theta number* (as opposed to theta number or AND-theta number).

Key to the definition is the orthonormal representation of a graph $G$. Originally our definition would be an orthonormal representation of the complementary graph. To avoid confusion we will call the representation we need an orthonormal corepresentation.

**Definition 9** *An orthonormal corepresentation of a graph $G$ is an attachment of a $d$-dimensional unit length vector (for some arbitrary but fixed $d$) $\boldsymbol{v}_a$ to each vertex $a \in V(G)$ such that if we have $xy \in E(G)$ for some $x, y \in V(G)$ then we must have $\boldsymbol{v}_x \boldsymbol{v}_y = 0$.*

**Definition 10** *The (complementary) Lovász theta number $\bar{\vartheta}(G)$ is defined as*

$$\bar{\vartheta}(G) = \min_{\{\boldsymbol{v}_a : a \in V(G)\}} \min_{\boldsymbol{c}} \max_{a \in V(G)} \frac{1}{(\boldsymbol{c}\boldsymbol{v}_a)^2},$$

*where the inner minimization is over all $d$-dimensional unit vectors $\boldsymbol{c}$ while the outer minimization is over all orthonormal corepresentations of the graph $G$. The unit vector $\boldsymbol{c}$ attaining the minimum in the inner minimization is called the* handle *of the corresponding orthonormal corepresentation.*

This seemingly not that simple definition gives a surprisingly well-behaving parameter. We will prove that it satisfies the conditions in Lemma 1 therefore it is also an upper bound on $C_{\mathrm{OR}}(G)$.

**Lemma 5** *For any graph $G$ we have*

$$\omega(G) \leq \bar{\vartheta}(G).$$

*Proof.* Let $\{\boldsymbol{v}_a : a \in V(G)\}$ be an optimal orthonormal corepresentation of $G$ with handle $\boldsymbol{c}$ and $Q \subseteq V(G)$ be a set of vertices that induces a largest clique in $G$. Then by the Pythagorean theorem we can write

$$1 = \boldsymbol{c}^2 \geq \sum_{a \in Q} (\boldsymbol{c}\boldsymbol{v}_a)^2 \geq |Q| \min_{a \in V(G)} (\boldsymbol{c}\boldsymbol{v}_a)^2 = \frac{|Q|}{\bar{\vartheta}(G)}.$$

Note that the first inequality follows from the Pythagorean theorem using the fact that $\boldsymbol{c}$ has unit length and the vectors $\{\boldsymbol{v}_a : a \in Q\}$ are pairwise orthogonal. Multiplying by $\bar{\vartheta}(G)$ we get

$$\bar{\vartheta}(G) \geq |Q| = \omega(G)$$

as required. □

To prove submultiplictivity with respect to the OR-product we will use the following identity stating which requires the next definition.

**Definition 11** *For two vectors $\boldsymbol{u} = (u_1, \ldots, u_m) \in R^m$ and $\boldsymbol{v} = (v_1, \ldots, v_n) \in R^n$ we let*

$$\boldsymbol{u} \circ \boldsymbol{v} := (u_1v_1, u_1v_2 \ldots, u_1v_n, u_2v_1, u_2v_2, \ldots, u_mv_n) \in R^{mn}.$$

**Identity 8** *For two vectors $\boldsymbol{x}, \boldsymbol{y} \in R^m$ and two others $\boldsymbol{u}, \boldsymbol{v} \in R^n$ we always have*

$$(\boldsymbol{x} \circ \boldsymbol{u})(\boldsymbol{y} \circ \boldsymbol{v}) = (\boldsymbol{x}\boldsymbol{y})(\boldsymbol{u}\boldsymbol{v}).$$

□

*Proof.* Both sides are equal to $\left(\sum_{i=1}^m x_iy_i\right) \cdot \left(\sum_{r=1}^n u_rv_r\right)$. □

**Lemma 6** *For any two finite simple graphs $F$ and $G$ we have*

$$\bar{\vartheta}(F \cdot G) \leq \bar{\vartheta}(F)\bar{\vartheta}(G).$$

*Proof.* Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in R^d$, $\boldsymbol{c} \in R^d$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in R^k$, $\boldsymbol{h} \in R^k$ be optimal orthonormal corepresentations with corresponding handles for graphs $F$ and $G$, respectively. We claim that in that case the vectors $\boldsymbol{u}_i \circ \boldsymbol{v}_j \in R^{dk}$ give an orthonormal corepresentation of the graph $F \cdot G$. Indeed, if vertices $(x, y)$ and $(z, t)$ form an edge in $F \cdot G$, then we have either $xz \in E(F)$ or $yt \in E(G)$ and thus

$$(\boldsymbol{u}_x \circ \boldsymbol{v}_y)(\boldsymbol{u}_z \circ \boldsymbol{v}_t) = (\boldsymbol{u}_x\boldsymbol{u}_z)(\boldsymbol{v}_y\boldsymbol{v}_t) = 0.$$

The unit length criterion is also satisfied as

$$(\boldsymbol{u}_i \circ \boldsymbol{v}_j)^2 = \boldsymbol{u}_i^2\boldsymbol{v}_j^2 = 1.$$

Similarly, $\boldsymbol{c} \circ \boldsymbol{h}$ is a unit vector in $R^{dk}$ since $(\boldsymbol{c} \circ \boldsymbol{h})^2 = \boldsymbol{c}^2\boldsymbol{h}^2 = 1$ by our identity.

This means that the above corepresentation of $F \cdot G$ with potential handle $\boldsymbol{c} \circ \boldsymbol{h}$ provides an upper bound on $\bar{\vartheta}(F \cdot G)$, thus

$$\bar{\vartheta}(F \cdot G) \leq \max_{i,j} \frac{1}{((\boldsymbol{c} \circ \boldsymbol{h})(\boldsymbol{u}_i \circ \boldsymbol{v}_j))^2} = \max_{i,j} \frac{1}{(\boldsymbol{c}\boldsymbol{u}_i)^2(\boldsymbol{h}\boldsymbol{v}_j)^2} =$$

$$= \max_i \frac{1}{(\boldsymbol{c}\boldsymbol{u}_i)^2} \max_j \frac{1}{(\boldsymbol{h}\boldsymbol{v}_j)^2} = \bar{\vartheta}(F)\bar{\vartheta}(G)$$

as needed. □

**Theorem 9** *For any finite simple graph $G$ we have*

$$C_{\mathrm{OR}}(G) \le \bar{\vartheta}(G).$$

*Proof.* By Lemmas 5 and 6 the statement immediately follows from Lemma 1. □

**Proposition 5**

$$\bar{\vartheta}(C_5) = \sqrt{5}.$$

*Hints for the proof.* By the inequalities

$$\sqrt{5} \le C_{\mathrm{OR}}(C_5) \le \bar{\vartheta}(C_5)$$

it is enough to show an orthonormal corepresentation of $C_5$ $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_5 \in R^d$ with potential handle $\boldsymbol{c} \in R^d$ for some $d$ that gives $\max_i \frac{1}{(\boldsymbol{c}\boldsymbol{u}_i)^2} = \sqrt{5}$. To this end consider an umbrella with five unit length ribs and a unit length handle opened up to the position when ribs that are not next to each other become orthogonal to each other. Note that this gives an orthonormal corepresentation of $C_5$ in $R^3$. Using basic analytic geometry one can calculate that for this representation any of the vectors $\boldsymbol{u}_i$ belonging to a rib and the handle $\boldsymbol{c}$ gives $\frac{1}{(\boldsymbol{c}\boldsymbol{u}_i)^2} = \sqrt{5}$ proving the required inequality $\bar{\vartheta}(C_5) \le \sqrt{5}$. Since the reverse inequality was already known, this implies $\bar{\vartheta}(C_5) = \sqrt{5}$. □

By the foregoing we already have

**Corollary 10** (Lovász [39])

$$C_{\mathrm{OR}}(C_5) = \sqrt{5}.$$

□

*Eighth lecture* (October 4, 2022)

Below we state some more of the nice properties of the Lovász theta number. All these are proven in [39].

We have defined $\bar{\vartheta}(G)$ as a minimum above. It can also be defined as a maximum. First we show that the value to be maximized is always a lower bound.

**Lemma 7** *Let $\{\boldsymbol{u}_a : a \in V(G)\}$ be an orthonormal corepresentation of $G$, $\{\boldsymbol{v}_a : a \in V(G)\}$ be an orthonormal corepresentation of $\overline{G}$ and $\boldsymbol{c}, \boldsymbol{h}$ be any vectors (of appropriate dimension for the following formula to make sense). Then*

$$\sum_{a \in V(G)} (\boldsymbol{c}\boldsymbol{u}_a)^2 (\boldsymbol{h}\boldsymbol{v}_a)^2 \le \boldsymbol{c}^2 \boldsymbol{h}^2.$$

*Proof.* Since any two vertices are adjacent either in $G$ or in $\overline{G}$ we have that whenever $a \ne b$

$$(\boldsymbol{u}_a \circ \boldsymbol{v}_a)(\boldsymbol{u}_b \circ \boldsymbol{v}_b) = (\boldsymbol{u}_a \boldsymbol{u}_b)(\boldsymbol{v}_a \boldsymbol{v}_b) = 0,$$

while

$$(\boldsymbol{u}_a \circ \boldsymbol{v}_a)(\boldsymbol{u}_a \circ \boldsymbol{v}_a) = \boldsymbol{u}_a^2 \boldsymbol{v}_a^2 = 1.$$

This means that the vectors $\boldsymbol{u}_a \circ \boldsymbol{v}_a$ form an othonormal system implying that

$$\boldsymbol{c}^2 \boldsymbol{h}^2 = (\boldsymbol{c} \circ \boldsymbol{h})^2 \ge \sum_{a \in V(G)} ((\boldsymbol{c} \circ \boldsymbol{h})(\boldsymbol{u}_a \circ \boldsymbol{v}_a))^2 = \sum_{a \in V(G)} (\boldsymbol{c}\boldsymbol{u}_a)^2 (\boldsymbol{h}\boldsymbol{v}_a)^2.$$

□

**Theorem 11** *If $\{\boldsymbol{v}_a : a \in V(G)\}$ is an orthonormal corepresentation of $\overline{G}$ and $\boldsymbol{h}$ is a unit vector (of the same dimension) then*

$$\bar{\vartheta}(G) \geq \sum_{a \in V(G)} (\boldsymbol{h}\boldsymbol{v}_a)^2.$$

*Proof.* Let $\{\boldsymbol{u}_a : a \in V(G)\}$ be an optimal orthonormal corepresentation of $G$ with handle $\boldsymbol{c}$ and $\{\boldsymbol{v}_a : a \in V(G)\}$ and $\boldsymbol{h}$ be as in the statement. Then by Lemma 7 we have

$$1 = \boldsymbol{c}^2\boldsymbol{h}^2 \geq \sum_{a \in V(G)} (\boldsymbol{c}\boldsymbol{u}_a)^2(\boldsymbol{h}\boldsymbol{v}_a)^2 \geq$$

$$\min_{a \in V(G)} (\boldsymbol{c}\boldsymbol{u}_a)^2 \sum_{a \in V(G)} (\boldsymbol{h}\boldsymbol{v}_a)^2 = \frac{1}{\bar{\vartheta}(G)} \sum_{a \in V(G)} (\boldsymbol{h}\boldsymbol{v}_a)^2$$

giving the statement. □

It turns out that if we maximize the lower bound in the above inequality then we get equality, that is, $\bar{\vartheta}(G)$ satisfies a minmax theorem. This we state below without proof.

**Theorem 12** (Lovász [39])

$$\bar{\vartheta}(G) = \max_{\boldsymbol{v}_j : j \in V(G), \boldsymbol{h}} \sum_{j=1}^{n} (\boldsymbol{h}\boldsymbol{v}_j)^2,$$

*where the maximization is over all possible orthonormal corepresentations of the complementary graph $\overline{G}$ of $G$ and unit vectors (of the same dimension) $\boldsymbol{h}$.*

Note that, as minmax theorems in general, the above equality gives a so-called "good characterization" of $\bar{\vartheta}(G)$, that is a way to convince someone easily about its value once we figured it out and the representations that attain it. This is because presenting an orthonormal corepresentation of $G$ and an orthonormal corepresentation of $\overline{G}$ that provide the same value as an upper and as a lower bound on $\bar{\vartheta}(G)$, we can already be sure that they are optimal and give the right value of $\bar{\vartheta}(G)$. Theorem 12 can also be used to prove that the inequality in Lemma 6 is actually an equality, that is, $\bar{\vartheta}(F \cdot G) = \bar{\vartheta}(F)\bar{\vartheta}(G)$.

Theorem 12 also has the following consequence.

**Corollary 13** *For any finite simple graph $G$ we have*

$$\bar{\vartheta}(G) \leq \chi_f(G).$$

That is to say that $\bar{\vartheta}(G)$ is a better upper bound of $C_{\mathrm{OR}}(G)$ than the fractional chromatic number for any graph, not only the 5-cycle.

*Proof.* Let $\{\boldsymbol{v}_i \in R^d : i \in V(G)$ be an orthonormal corepresentation of $\overline{G}$ and $\boldsymbol{h} \in R^d$ a unit vector (for some $d$). Furthermore let $A \subseteq V(G)$ be an independent set in $G$, that is a clique in $\overline{G}$. The latter implies that the vectors in $\{\boldsymbol{v}_i : i \in A\}$ are pairwise orthogonal, therefore $1 = \boldsymbol{h}^2 \geq \sum_{i \in A}(\boldsymbol{h}\boldsymbol{v}_i)^2$. This means that putting weight $(\boldsymbol{h}\boldsymbol{v}_i)^2$ on vertex $i$ for every $i \in V(G)$ gives a fractional clique of $G$. Thus

$$\chi_f(G) = \omega_f(G) \geq \max_{\boldsymbol{v}_j : j \in V(G), \boldsymbol{h}} \sum_{j=1}^{n} (\boldsymbol{h}\boldsymbol{v}_j)^2 = \bar{\vartheta}(G).$$

□

Another nice property of $\bar{\vartheta}(G)$ shows that in fact, $\bar{\vartheta}(C_5) = \sqrt{5}$ is just a special case of a more general fact. This follows from the following theorem. To state it we need the following definition.

**Definition 12** *A graph $G$ is called vertex-transitive if for all pairs of vertices $u$ and $v$ one can give an automorphism (an isomorphism to itself) of $G$ that maps $u$ to $v$.*

Intuitively, the above means that all vertices "look the same", that is the graph is highly symmetric. It is easy to see that $C_5$ is vertex-transitive.

**Theorem 14** (Lovász [39]) *If $G$ is a vertex-transitive graph on $n$ vertices then*

$$\bar{\vartheta}(G)\bar{\vartheta}(\overline{G}) = n.$$

**Corollary 15** *If $G$ is a vertex-transitive and self-complementary graph on $n$ vertices then*

$$C_{\mathrm{OR}}(G) = \bar{\vartheta}(G) = \sqrt{n}.$$

*Proof.* Let $G$ be a graph as in the statement. Since it is self-complementary we have $\bar{\vartheta}(G) = \bar{\vartheta}(\overline{G})$. Since it is vertex-transitive Theorem 14 implies that $\bar{\vartheta}^2(G) = \bar{\vartheta}(G)\bar{\vartheta}(\overline{G}) = n$ implying $C_{\mathrm{OR}}(G) \leq \bar{\vartheta}(G) = \sqrt{n}$.

The reverse inequality follows by observing that if $G$ is self-complementary and $f\colon V(G) \to V(G)$ is an isomorphism from $G$ to $\overline{G}$ then the 2-length sequences of vertices $(v, f(v))$ induce a clique of size $n$ in $G^2$. Since that implies $\omega(G^{2r}) \geq n^r$ for every positive integer $r$, we get $C_{\mathrm{OR}}(G) \geq \sqrt{n}$. □

Examples for self-complementary vertex-transitive graphs are the so-called Payley graphs (see `https://en.wikipedia.org/wiki/Paley_graph` for their definition and other details). In particular, the Payley graph on 17 vertices is a graph with no clique or independent set of size larger than 3, thereby proving that the Ramsey number $R(4) \geq 18$ which is actually a sharp lower bound. Corollary 15 shows that its Shannon capacity (both AND and OR as it is self-complementary) equals to $\sqrt{17}$.

# 5 More on the fractional chromatic number

It is natural to ask how much can the fractional chromatic number and the chromatic number differ. The first question is where to look for graphs that may provide a large gap. Fortunately, one can find those graphs which are guaranteed to attain the largest possible gap. Key to this is the following notion.

**Definition 13** *A graph homomorphism from a graph $F$ to a graph $G$ is a mapping $\varphi\colon V(F) \to V(G)$ that preserves edges that is for which*

$$uv \in E(F) \Rightarrow \varphi(u)\varphi(v) \in E(G).$$

*The existence of a homomorphism from $F$ to $G$ is denoted by $F \to G$.*

It is worth noting that a proper coloring of a graph $G$ with $n$ colors is equivalent to a homomorphism of $G$ to the complete graph $K_n$. So an alternative definition of the chromatic number could be

$$\chi(G) = \min\{n : G \to K_n\}.$$

This immediately implies the following.

**Proposition 6**

$$F \to G \Rightarrow \chi(F) \leq \chi(G).$$

*Proof.* By the above observation we know that $G \to K_{\chi(G)}$ and since the composition of homomorphisms is also a homomorphism, we can write $F \to G \to K_{\chi(G)}$, that is, we also have $F \to K_{\chi(G)}$ which is equivalent to a proper coloring of $F$ with $\chi(G)$ colors. □

We will see that just like the chromatic number, the fractional chromatic number can also be defined via graph homomorphisms. To see that we first consider another alternative definition that can be given via so-called $b$-fold colorings.

**Definition 14** *For a positive integer $b$ a $b$-fold coloring of a graph $G$ is an attachment of $b$ distinct colors to each vertex such that adjacent vertices get disjoint sets of colors. The minimum number of colors needed for this is the* $b$-fold chromatic number $\chi_b(G)$.

A $b$-fold coloring is easy to turn to a fractional coloring: just attach weight $\frac{1}{b}$ to every independent set that is a color class in your $b$-fold coloring. (Note that you may use two different colors on exactly the same vertices. Then the two color classes coincide and the corresponding independent set gets the weight $\frac{1}{b}$ twice. Or several times if there are other color classes that are the same.)

It is also easy to see that a fractional coloring with all weights rational can also be turned into a $b$-fold coloring for some appropriate $b$. Since irrationals can arbitrarily well be approximated by rationals, this leads to the fact that

$$\chi_f(G) = \inf_b \frac{\chi_b(G)}{b},$$

where again we can write min in place of inf.

Generalizing that a proper coloring is equivalent to a graph homomorphism to a complete graph, one can observe that a $b$-fold coloring of a graph $G$ using $a$ colors is equivalent to a homomorphism to a special target graph called Kneser graph $\mathrm{KG}(a, b)$.

**Definition 15** *For positive integers $n \geq 2k$ the Kneser graph* $\mathrm{KG}(n, k)$ *is defined by*

$$V(\mathrm{KG}(n, k)) = \binom{[n]}{k}$$

*and*

$$E(\mathrm{KG}(n, k)) = \{AB : A, B \in \binom{[n]}{k}, A \cap B = \emptyset\}.$$

Thus we get that

$$\chi_f(G) = \inf \left\{ \frac{a}{b} : G \to \mathrm{KG}(a, b) \right\},$$

where again, we can write min in place of inf.

Kneser graphs form a very interesting family of graphs. They are not yet that exciting for $n = 2k$ in which case they consist of only several independent edges and for $k = 1$ when we have $\mathrm{KG}(n, 1) = K_n$. So the first interesting case is $\mathrm{KG}(5, 2)$ that turns out to be isomorphic to the famous Petersen graph–a fact that already suggests that this family has some interesting properties.

*Ninth lecture* (October 7, 2022)

The FIRST HOMEWORK WAS HANDED OUT TODAY, the problem sheet is also POSSIBLE TO DOWNLOAD FROM THE WEBSITE.

One of the most famous facts about Kneser graphs is the Lovász-Kneser theorem that gives their chromatic number.

**Theorem 16** (Lovász-Kneser theorem [38])

$$\chi(\mathrm{KG}(n,k)) = n - 2k + 2.$$

The above statement was conjectured by Kneser [28] who observed that a coloring with $n-2k+2$ colors is possible. The truth of this conjecture was proven by Lovász more than two decades later in his celebrated paper [38] that is widely considered to be the starting point of a new branch of combinatorics called topological combinatorics. In fact, a very surprising feature of the proof was that it used what is called the Borsuk-Ulam theorem, a famous result in another branch of mathematics called algebraic topology. Several other combinatorial results were found later using similar tools, the book [44] is a very well-written witness of that.
(The Lovász-Kneser theorem appears with proof in the course *Graphs, hypergraphs and their applications* therefore here we do not prove it. Nevertheless, the interested reader who did not take that course is encouraged to try to find a proper coloring of $\mathrm{KG}(n,k)$ with $n-2k+2$ colors.)

What is said above shows that if we have a graph $G$ with a large gap between its chromatic number and fractional chromatic number, then there must exist a Kneser graph for which this gap is at least as large as for $G$. This is because if $\chi_f(G) = \frac{a}{b}$ then there is some Kneser graph $\mathrm{KG}(n,k)$ for which we have $G \to \mathrm{KG}(n,k)$ and $\frac{a}{b} = \frac{n}{k} \geq \chi_f(\mathrm{KG}(n,k))$ where the latter inequality follows from the fact that $F \to \mathrm{KG}(n,k)$ already implies $\chi_f(F) \leq \frac{n}{k}$ and $\mathrm{KG}(n,k)$ obviously admits a homomorphism to itself. So we have $\chi_f(G) \geq \chi_f(\mathrm{KG}(n,k))$, while $\chi(G) \leq \chi(\mathrm{KG}(n,k))$ follows from Proposition 6.

By the above discussion we can already see how large the gap between the fractional and the ordinary chromatic number can be. We see that the largest possible gap can be attained for Kneser graphs and we also see that this gap is unbounded in a very strong sense: If $\frac{n}{k}$ is set to be any constant that is larger than 2 then $n-2k+2$ already goes to infinity as $n$ and $k$ grow. So for $\mathrm{KG}(n,k)$ with $\frac{n}{k} = 2+\varepsilon$ for some $\varepsilon > 0$ and $n-2k+2$ larger than any prescribed number $N$ we have $\chi_f(\mathrm{KG}(n,k)) \leq 2+\varepsilon$ and $\chi(\mathrm{KG}(n,k) > N$.

Below we show that we actually have $\chi_f(\mathrm{KG}(n,k)) = \frac{n}{k}$. That will also need to accept the truth of the following famous theorem (also appearing in the course about hypergraphs mentioned above).

**Theorem 17** (Erdős–Ko–Rado theorem [16]) *If $n \geq 2k$ and $\mathcal{A}$ is a family of pairwise intersecting $k$-element subsets of an $n$-element set then*

$$|\mathcal{A}| \leq \binom{n-1}{k-1}.$$

*Equality holds if and only if $\mathcal{A}$ consists of all the $k$-element sets containing a fixed element of the $n$-element set.*

Note that the above theorem is equivalent to saying that

$$\alpha(\mathrm{KG}(n,k)) = \binom{n-1}{k-1}.$$

Recall that a graph $G$ is called *vertex-transitive* if for all pairs of vertices $u$ and $v$ one can give an automorphism of $G$ that maps $u$ to $v$. Note that Kneser graphs are obviously vertex-transitive because any of its vertices can be put to another one by an automorphism generated by permuting the elements of the basic $n$-element set in the definition of the Kneser graph at hand.

Before stating a general result about the fractional chromatic number of vertex-transitive graphs we prove the following statement that is true for all graphs.

**Proposition 7** *For any finite simple graph $G$*

$$\chi_f(G) \geq \frac{|V(G)|}{\alpha(G)}$$

*holds.*

*Proof.* Giving weight $\frac{1}{\alpha(G)}$ to every vertex no independent set gets more weight than 1, so this is a fractional clique with total weight $\frac{|V(G)|}{\alpha(G)}$. Thus

$$\chi_f(G) = \omega_f(G) \geq \frac{|V(G)|}{\alpha(G)}.$$

□

**Theorem 18** *If $G$ is vertex-transitive, then*

$$\chi_f(G) = \frac{|V(G)|}{\alpha(G)}.$$

*Proof.* We have already seen in Proposition 7 that

$$\chi_f(G) \geq \frac{|V(G)|}{\alpha(G)}$$

for any graph $G$.

Now we prove that for vertex-transitive graphs the reverse inequality also holds. If $G$ is vertex-transitive, then all vertices are contained in the same number of maximum independent sets. Call this number $t$ and give every maximum independent set (that is those of size $\alpha(G)$) weight $\frac{1}{t}$. By the definition of $t$ this is a fractional colouring: all vertices get total weight $t\frac{1}{t} = 1$. If the number of maximum independent sets is $\ell$ then we distributed altogether $\ell\frac{1}{t} = \frac{\ell}{t}$ total weight, thus this is an upper bound on $\chi_f(G)$. Now we show that this upper bound is equal to $\frac{|V(G)|}{\alpha(G)}$.

To this end we calculate the number of pairs $(v, A)$ where $A$ is an independent set of size $\alpha(G)$ and $v \in A$. We have $\ell$ such $A$ each containing $\alpha(G)$ vertices, so the number of such pairs is $\ell\alpha(G)$. On the other hand, we have $|V(G)|$ vertices and each is contained in $t$ independent sets of size $\alpha(G)$, so the number of such pairs is $|V(G)|t$. Thus

$$\ell\alpha(G) = |V(G)|t,$$

that is we obtained

$$\chi_f(G) \leq \frac{\ell}{t} = \frac{|V(G)|}{\alpha(G)}.$$

By the two inequalities above the statement $\chi_f(G) = \frac{|V(G)|}{\alpha(G)}$ follows. □

**Corollary 19** *For every $n \geq 2k$ we have*

$$\chi_f(\mathrm{KG}(n,k)) = \frac{n}{k}.$$

*Proof.* Since $\mathrm{KG}(n,k)$ is vertex-transitive we have

$$\chi_f(\mathrm{KG}(n,k)) = \frac{|V(\mathrm{KG}(n,k))|}{\alpha(\mathrm{KG}(n,k))} = \frac{\binom{n}{k}}{\binom{n-1}{k-1}} = \frac{n}{k}$$

by Theorem 18 and the Erdős–Ko–Rado theorem. □

The above discussion shows that it is not possible to bound the chromatic number from above in terms of the fractional chromatic number. If, however, we also take into account the value of the independence number, then the situation changes. In particular, we are going to prove the following theorem that will be very useful later.

**Theorem 20**

$$\chi(G) \leq \chi_f(G)(1 + \ln \alpha(G)).$$

We will prove Theorem 20 in a more general form. Let $H$ be a hypergraph. A fractional covering of the edges of $H$ is a function $f\colon V(H) \to R_{+,0}$ such that for every edge $E \in E(H)$ we have $\sum_{v \in E} f(v) \geq 1$. The fractional covering number is the minimum possible value of $\sum_{v \in V(H)} f(v)$ over all fractional coverings $f$. Note that if given a graph $G$ we define $H_G$ to be the hypergraph with $V(H) = S(G)$, i.e., the vertices of $H$ are the independent sets of $G$ and each edge $E = E_x$ of $H$ belongs to a vertex $x \in V(G)$ such that $E_x = \{A \in S(G) : x \in A\}$ then $\tau_f(H_G) = \chi_f(G)$ by definition. If we denote the covering number, the minimum number of vertices covering (or pinning) every edge of $H$ by $\tau(H)$, then we can also see that $\tau(H_G) = \chi(G)$. Observe also that $\alpha(G) = \Delta(H_G)$, the maximum degree of $H_G$. So Theorem 20 will follow from the following more general result.

**Theorem 21** (Lovász [37])

$$\tau(H) \leq \tau_f(H)(1 + \ln \Delta(H)).$$

To prove this theorem observe that just like the fractional chromatic number, $\tau_f(H)$ is also the solution of a linear program. Its dual program defines the fractional matching number $\nu_f(H)$ as follows. A fractional matching of hypergraph $H$ is a weighting $w\colon E(H) \to R_{+,0}$ of the edges of $H$ such that for any $x \in V(H)$ we have $\sum_{E \ni x} w(E) \leq 1$. Then

$$\nu_f(H) = \sup \sum_{E \in E(H)} w(E)$$

where the supremum (which, as usually, can be replaced by a maximum) is taken over all fractional matchings of $H$. Thus by the duality theorem of linear programming we have

$$\tau_f(H) = \nu_f(H).$$

Let $\nu_i(H)$ denote the largest number of (not necessarily distinct) edges in $H$ the union of which covers every vertex at most $i$ times. (Note that $\nu_1(H)$ is the usual matching number of the hypergraph $H$.) Then clearly

$$\nu_f(H) \geq \frac{\nu_i(H)}{i}$$

that can be seen by giving weight $\frac{1}{i}$ to every edge in an $i$-fold matching. Let $\tilde{\nu}_i(H)$ denote the largest number of *distinct* edges in $H$ that cover every vertex at most $i$ times. Obviously $\tilde{\nu}_i(H) \leq \nu_i(H)$ and therefore $\nu_f(H) \geq \frac{\tilde{\nu}_i(H)}{i}$ holds, too. To prove Theorem 21 we need the following lemma.

**Lemma 8**

$$\tau(H) \leq \frac{\tilde{\nu}_1(H)}{1 \cdot 2} + \frac{\tilde{\nu}_2(H)}{2 \cdot 3} + \ldots + \frac{\tilde{\nu}_{d-1}(H)}{(d-1) \cdot d} + \frac{\tilde{\nu}_d(H)}{d},$$

*where* $d = \Delta(H)$ *is the maximum degree of* $H$.

*Proof.* To bound $\tau(H)$ from above we prepare a set of vertices pinning all edges in a greedy way: first we choose a vertex that pins the largest number of edges, that is, it has maximum degree, then a next vertex which has maximum degree in the hypergraph remaining if we remove the already pinned edges, etc. This is continued until all edges are pinned. Let $t_i$ denote the number of steps in which we choose a vertex pinning exactly $i$ edges that were not yet pinned in the earlier steps and let $t = t_d + t_{d-1} + \ldots t_1$ be the total number of vertices selected this way. (Note that some of the $t_i$'s may be zero.) Let $H_i$ denote the hypergraph we have after deleting the edges already pinned in the first $t_d + \ldots + t_{i+1}$ steps. Then we have $\Delta(H_i) \leq i$ (since no more vertices can pin at least $i+1$ edges) and $E(H_i) = it_i + (i-1)t_{i-1} + \ldots + 2t_2 + t_1$ (since exactly that many edges will be pinned in the remaining steps). By $\Delta(H_i) \leq i$ we can also say that $\tilde{\nu}_i(H) \geq |E(H_i)|$. Therefore we can write

$$\tilde{\nu}_i \geq it_i + \ldots 2t_2 + t_1$$

for every $i \in \{1, \ldots, d\}$. Divide these inequalities by $i(i+1)$ for $i = 1, \ldots, d-1$ and with $d$ for $i = d$ to obtain

$$\frac{\tilde{\nu}_i}{i(i+1)} \geq \frac{t_i}{i+1} + \frac{(i-1)t_{i-1}}{i(i+1)} + \ldots + \frac{2t_2}{i(i+1)} + \frac{t_1}{i(i+1)}$$

for $i = 1, 2, \ldots d-1$ and

$$\frac{\tilde{\nu}_d}{d} \geq t_d + \frac{(d-1)t_{d-1}}{d} + \ldots + \frac{2t_2}{d} + \frac{t_1}{d}$$

for $i = d$. Adding all these inequalities the coefficient of $t_i$ on the right hand side becomes

$$i\left(\frac{1}{i(i+1)} + \frac{1}{(i+1)(i+2)} + \ldots + \frac{1}{(d-1)d} + \frac{1}{d}\right)$$

$$= i\left(\frac{1}{i} - \frac{1}{i+1} + \frac{1}{i+1} - \frac{1}{i+2} \cdots \frac{1}{d-1} - \frac{1}{d} + \frac{1}{d}\right) = 1$$

Note that this is so also for $i = d$, so the sum of the right hand sides is just $t_d + t_{d-1} + \ldots + t_1 = t$ while the sum of the left hand sides is just the right hand side in the statement of the lemma. So using $\tau(H) \leq t$ the result follows. $\square$.

*Proof of Theorem 21.* If we substitute the inequality $\frac{\tilde{\nu}_i(H)}{i} \leq \nu_f(H)$ for every $i$ in the statement of Lemma 8 and then use $\tau_f(H) = \nu_f(H)$ then we obtain

$$\tau(H) \leq \tau_f(H)\left(\frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{d} + 1\right).$$

Using the well-known fact that

$$1 + \frac{1}{2} + \ldots + \frac{1}{d} \leq 1 + \ln d$$

we obtain

$$\tau(H) \leq \tau_f(H)(1 + \ln d)$$

which is just the statement of the theorem we had to prove. $\square$

Using the above theorem we will be able to prove the followimg result.

**Theorem 22** (McEliece–Posner theorem [45])

$$\lim_{t\to\infty} \sqrt[t]{\chi(G^t)} = \chi_f(G).$$

*Tenth lecture* (October 11, 2022)

First we prove the McEliece–Posner theorem.

*Proof.* Note that by $\chi_f(F \cdot G) = \chi_f(F)\chi_f(G)$ we have $\chi_f(G^t) = [\chi_f(G)]^t$ and thus

$$\lim_{t\to\infty} \sqrt[t]{\chi(G^t)} \geq \lim_{t\to\infty} \sqrt[t]{\chi_f(G^t)} = \lim_{t\to\infty} \sqrt[t]{[\chi_f(G)]^t} = \chi_f(G).$$

To prove the reverse inequality we use Theorem 20 and the easy fact that $\alpha(G^t) = [\alpha(G)]^t$. By these we can write

$$\lim_{t\to\infty} \sqrt[t]{\chi(G^t)} \leq \lim_{t\to\infty} \sqrt[t]{\chi_f(G^t)(1 + \ln \alpha(G^t))} =$$

$$\lim_{t\to\infty} \sqrt[t]{[\chi_f(G)]^t(1 + \ln[\alpha(G)]^t)} = \chi_f(G) \lim_{t\to\infty} \sqrt[t]{(1 + t \ln \alpha(G))} = \chi_f(G).$$

$\square$

# 6 Witsenhausen rate

Consider the following communication problem that was considered generally by Witsenhausen [56], while the following example follows the idea of Alon and Orlitsky [4] (this example is also quoted in [53]).

Alice and Bob teach in a school attended by 250 pupils. Every afternoon the students form two teams of 100 participants each that will play some game in which two teams are the rivals of each other, e.g., they play “Capture the Flag”. The other 50 just watch the game and the selection of the two teams and the audience is varying from day to day. Bob is the teacher with the students at the beginning of each afternoon when they form the two teams and start playing. But he leaves earlier than the end of the game and he does not know who is the winner. Alice is the gym teacher and later in the day she learns which students were in the winning team but she does not know who were members of the other one. The task is that Alice communicates this knowledge (the winning team) to Bob in a shortest possible message.

Observe that the abstract model of the above situation is that Bob knows an edge of the Kneser graph $\mathrm{KG}(n,k)$ and needs to learn one of the endpoints of this edge (the “winner”). Alice knows this endpoint but she does not know the other end of the edge Bob knows. So, Alice knows a message to transfer that can be any one out of $\binom{250}{100}$ possible messages. But she need not use $\log_2 \binom{250}{100}$ bits for the communication because she knows that Bob has some extra information (called *side-information*), which is an edge containing the vertex she knows. If she knew exactly which edge Bob knows, she could send just 1 bit. She does not know this, but she can still use the fact that Bob knows two vertices that are adjacent, so at least she did not need to send a different message for non-adjacent vertices. Thus the best she can do to communicate the result of an afternoon game is to send the shortest message with which she can encode a set of independent vertices such that the independent sets belonging to all possible messages cover every vertex. This means that these independent sets form a proper coloring of the graph, so the minimum number of different messages needed is just the chromatic number which is $\chi(\mathrm{KG}(250, 100)) = 52$, so the nuber of bits to communicate is $\lceil \log_2 52 \rceil = 6$ each afternoon.

We soon come back to this example but let us first formulate the general problem. A source emits one message of a set $\mathcal{X}$ of $M$ possible messages every day (or: in every time slot) that we have to encode so that it should be possible to decode it without error. Under this condition we want it to be compressed as much as possible, that is, use the shortest possible encoding. We also know that the decoder will have some side-information. The exact side-information is not known to us but we know at least a set $\mathcal{Y}$ of the possible pieces of side-information and we also know which pairs $(x, y)$ of $x \in \mathcal{X}, y \in \mathcal{Y}$ can appear together with positive probability. The question is the minimum length of the message we have to send.

The general model of the situation is this: we can consider a graph $G$ with vertex set $V(G) := \mathcal{X}$ and edge set

$$E(G) := \{\{x, x'\} : \exists y \in \mathcal{Y} \;\; \mathrm{Prob}(x, y)\mathrm{Prob}(x', y) > 0\}.$$

The question is how many bits we should communicate in one message. It is clear that in case we have to communicate only one outcome at a time, then the answer is $\lceil \log \chi(G) \rceil$. (We use the convention that if not stated otherwise, then log always means $\log_2$ in our information theoretic context.) But the question arises, as in most communication problems in information theory: What happens if we are allowed to communicate several outcomes of the source together? Can we gain with that possibility in terms of the length of messages per outcome? And as in many similar situations, the answer is yes.

Now we formulate the graph theory model belonging to the last questions. If we have $t$ outcomes of the source we have to communicate together then a message can be considered as a $t$-length sequence $(x_1, \ldots, x_t)$ formed by the possible individual messages. And we have to send such a message which can distinguish two outcomes in case they may have the same sequence of side-information $(y_1, \ldots, y_t)$ that comes with it at the decoding side. This means exactly that we have to encode a vertex of the AND power graph $G^{\wedge t}$ in such a way that adjacent vertices of this graph should be encoded by different codewords. Then the minimum length of binary codewords we should send is $\lceil \log \chi(G^{\wedge t}) \rceil$ that, as we will see, can be shorter than $t \lceil \log \chi(G) \rceil$. In fact, we are interested in the normalized asymptotic value of this length.

**Definition 16** *The logarithmic* Witsenhausen rate *of a graph $G$ is defined to be*

$$r(G) := \lim_{t \to \infty} \frac{1}{t} \log \chi(G^{\wedge t}).$$

*In combinatorial investigations we usually use the non-logarithmic version*

$$R(G) = \lim_{t \to \infty} \sqrt[t]{\chi(G^{\wedge t})}.$$

Note that the information theoretic content would suggest to write lim inf in the above definitions but it is easy to see (again by Fekete's Lemma as in the case of graph capacities) that the limits exist. Also note the obvious fact, that $r(G) = \log R(G)$.

Let us return now to our example with Alice, Bob and the Capture the Flag game at their school. We have seen that if Alice is obliged to communicate her information day-by-day then during $t$ days she will need $6t$ bits. What happens if she is allowed to tell the results of many days together? In the asymptotics she will need $t \cdot r(\mathrm{KG}(250, 100))$ bits and that is significantly fewer. To see this, it is enough to bear in mind that $G^{\wedge t} \subseteq G^t$ and therefore $\chi(G^{\wedge t}) \leq \chi(G^t)$ which implies

$$r(G) \leq \lim_{t\to\infty} \frac{1}{t} \log \chi(G^t) = \log \lim_{t\to\infty} \sqrt[t]{\chi(G^t)} = \log \chi_f(G)$$

by the McEliece-Posner theorem. Since $\chi_f(\mathrm{KG}(250, 100)) = \frac{250}{100} = 2.5$, Alice needs no more that $t \log 2.5 \approx 1.322t$ bits instead of $6t$ in this case.

## 6.1 General properties of the Witsenhausen rate

As we have already seen above

$$R(G) \leq \chi_f(G)$$

always holds. This already implies that $R(G) \leq \log \chi(G)$, too. We also obviously have $R(G) \geq \omega(G)$ by $\chi(G^{\wedge t}) \geq \omega(G^{\wedge t}) = [\omega(G)]^t$, therefore $\chi(G) = \omega(G)$ (in particular, if the graph is perfect) implies that in that case $R(G)$ coincides with this common value (just like $C_{\mathrm{OR}}(G)$ does in that case.)

The smallest imperfect graph, $C_5$ was also considered by Witsenhausen and he showed that $C_5^{\wedge 2}$ can be properly colored by 5 independent sets that implies $R(C_5) \leq \sqrt{5}$, since $R(G) \leq \sqrt[t]{\chi(G^{\wedge t})}$ holds for every positive integer $t$. (The proof is similar to that of $C_{\mathrm{OR}}(G) \geq \sqrt[t]{\omega(G^t)}$ using $\chi(G^{\wedge t}) \leq [\chi(G)]^t$ in place of $\omega(G^t) \geq [\omega(G)]^t$.) Lovász's result about the Shannon capacity of $C_5$ also determined the exact value of $R(C_5)$ via the folowing more general fact.

**Lemma 9**

$$R(G) \geq \frac{|V(G)|}{C_{\mathrm{OR}}(\overline{G})}.$$

*Proof.* By the general fact $\chi(F) \geq \frac{|V(F)|}{\alpha(F)}$ applied to $G^{\wedge t}$ (and denoting $|V(G)|$ by $n$) we can write

$$R(G) = \lim_{t\to\infty} \sqrt[t]{\chi(G^{\wedge t})} \geq \lim_{t\to\infty} \sqrt[t]{\frac{|V(G^{\wedge t})|}{\alpha(G^{\wedge t})}} =$$

$$\lim_{t\to\infty} \sqrt[t]{\frac{n^t}{\omega(\overline{G}^t)}} = \frac{n}{C_{\mathrm{OR}}(\overline{G})}.$$

□

**Corollary 23**

$$R(C_5) = \sqrt{5}.$$

*Proof.* $R(C_5) \geq \sqrt{5}$ follows from the above theorem and $C_{\mathrm{OR}}(C_5) = \sqrt{5}$. The reverse inequality follows from the observation that $\chi(C_5^{\wedge 2}) \leq 5$ that was already shown by Witsenhausen [56] by providing 5 independent sets in $C_5^{\wedge 2}$ covering all 25 vertices of $V(C_5^{\wedge 2})$. Note that in such a construction all the independent sets should be 5-vertex cliques of the second OR-product of the complementary

$C_5$, so the construction shows strong similarities with Shannon's construction that proved $C_{\mathrm{OR}}(C_5) \geq \sqrt{5}$. It is a little more involved in the sense that it should give 5 disjoint such cliques the union of which is the whole vertex set of $C_5^2$. Labelling the vertices of our $C_5$ by $0, 1, 2, 3, 4$ so that $ij$ is an edge if and only if $|i - j| \equiv 1 \pmod 5$ the construction is as follows (each $A_i$ is a color class).

$$A_1 = \{00, 12, 24, 31, 43\}, \quad A_2 = \{11, 23, 30, 42, 04\}, \quad A_3 = \{22, 34, 41, 03, 10\},$$

$$A_4 = \{33, 40, 02, 14, 21\}, \quad A_5 = \{44, 01, 13, 20, 32\}.$$

□

As a generalization of the Erdős–Ko–Rado theorem Lovász proved in [39] that $\bar{\vartheta}(\overline{\mathrm{KG}(n,k)}) = \binom{n-1}{k-1}$ which gives that all the inequalities in

$$\binom{n-1}{k-1} = \omega(\overline{\mathrm{KG}(n,k)}) \leq C_{\mathrm{OR}}(\overline{\mathrm{KG}(n,k)}) \leq \bar{\vartheta}(\overline{\mathrm{KG}(n,k)}) = \binom{n-1}{k-1}$$

are equalities. This implies by the previous lemma that

$$R(\mathrm{KG}(n,k)) = \frac{\binom{n}{k}}{\binom{n-1}{k-1}} = \frac{n}{k} = \chi_f(\mathrm{KG}(n,k)).$$

In particular, in the story of Alice and Bob, the upper bound we gave on $R(\mathrm{KG}(250, 100))$ was sharp.

*Eleventh lecture* (October 14, 2022)

# 7 Shannon capacity and Ramsey numbers

We know that the chromatic number of a graph can be arbitrarily large even if the graph contains no triangle. One of the best-known constructions proving this is that of Mycielski graphs. Triangle-free Kneser graphs also can have arbitrarily large chromatic number but their fractional chromatic number is bounded, it is always less than 3. Nevertheless a similar result is true for the fractional chromatic number, in fact, the fractional chromatic number of the series of Mycielski graphs is known to tend to infinity by the following cute result that we state below. Recall that the Mycielski construction is an iterative construction that creates a graph $M(G)$ from any graph $G$ with the properties that $\omega(M(G)) = \omega(G)$ and $\chi(M(G)) = \chi(G) + 1$.

**Theorem 24** (Larsen–Propp–Ullman [34]) *For any finite simple graph $G$*

$$\chi_f(M(G)) = \chi_f(G) + \frac{1}{\chi_f(G)}$$

*holds.*

It is also known that there are triangle-free graphs with arbitrarily large value of their $\bar{\vartheta}$-function, see e.g. [2]. It is a very natural question, and here the answer is not known, whether the Shannon OR-capacity of a triangle-free graph can also be arbitrarily large or it is bounded. Below we will show that this is closely related (in fact, equivalent) to a very famous problem of Erdős that at first look seems very different.

Recall that the Ramsey number $R(k_1, \ldots, k_m)$ denotes the smallest integer $N$ for which if we color the edges of $K_N$ with $m$ colors then a monochromatic $K_{k_i}$ must occur in the $i^{\text{th}}$ color for some $i \in \{1, \ldots, m\}$. Ramsey's theorem states that $R(k_1, \ldots, k_m)$ is finite for all possible choices of the parameters.

Our question above is about triangle-free graphs, basically we want to know how the clique number of their OR-powers can grow. First we look at a more general question: how large can the clique number be in the $m$-fold OR-product of graphs $G_1, \ldots, G_m$ if we have some given upper bound for the clique number of all these graphs. This is answered by the following result.

**Theorem 25** (Erdős–McEliece–Taylor [17]) *Let $G_1, \ldots, G_m$ be graphs with $\omega(G_i) \leq k_i$ for every $i \in \{1, \ldots, m\}$. Then*

$$\omega(G_1 \cdot G_2 \cdot \ldots \cdot G_m) \leq R(k_1 + 1, k_2 + 1, \ldots, k_m + 1) - 1.$$

*Moreover, this bound is sharp in the sense that there exist graphs $G_1, \ldots, G_m$ for which $\omega(G_i) \leq k_i$ holds for every $i$ while*

$$\omega(G_1 \cdot G_2 \cdot \ldots \cdot G_m) = R(k_1 + 1, k_2 + 1, \ldots, k_m + 1) - 1.$$

*Proof.* Let $G_1, \ldots, G_m$ satisfy $\forall i\ \omega(G_i) \leq k_i$ and let $Q \subseteq V(G_1) \times \ldots \times V(G_m)$ induce a largest clique in $G_1 \cdot \ldots \cdot G_m$. We define an edge coloring $c$ of the edges of this clique. For an edge between $\boldsymbol{x} = (x_1, \ldots, x_m)$ and $\boldsymbol{y} = (y_1, \ldots, y_m) \in Q$ let $c(\boldsymbol{xy})$ be defined as the minimum $i$ for which $x_i y_i \in E(G_i)$. Note that since $Q$ induces a clique, such an $i$ exists by the definition of the OR-product. Let $H_i$ denote the graph with vertex set $Q$ containing the edges colored with $i$ by the above rule. It is not hard to see that $H_i$ is isomorphic to a "blown up" version of $G_i$ that one obtains by substituting independent sets (possibly of size 0) into the vertices of $G_i$. In particular $\omega(H_i) \leq \omega(G_i)$ (a clique of size $r$ in $H_i$ requires $r$ pairwise adjacent vertices of $G_i$). This means that we could edge-color the clique of the product graph induced by $Q$ with $m$ colors so that the largest monochromatic clique in the $i^{\text{th}}$ color has size at most $\omega(G_i) = k_i$. This proves that $|Q| \leq R(k_1 + 1, k_2 + 1, \ldots, k_m + 1) - 1$.

To prove the second statement consider a complete graph on $N = R(k_1 + 1, k_2 + 1, \ldots, k_m + 1) - 1$ vertices with an $m$-edge-coloring containing no clique of size $k_i + 1$ monochromatic in the $i^{\text{th}}$ color for any $i$. By the definition of $R(k_1 + 1, k_2 + 1, \ldots, k_m + 1)$ such an edge-coloring of $K_N$ exists. Let $G_i$ be defined as the graph spanned by the edges of color $i$ for $i = 1, \ldots, m$. These graphs satisfy the requirement $\omega(G_i) \leq k_i$ for every $i$. In their OR-product the vertices given by the $m$-tuples of $m$ identical vertices give a clique of size $N$, since any two vertices are adjacent in one of the $m$ graphs. □

With a little more consideration we can see that if all $k_i$ values are the same, that is, $k_1 = \ldots = k_m =: k$, then we can choose all graphs being the same, i.e., $G_1 = \ldots = G_m =: G$ in the second part of the statement of Theorem 25. Indeed, in this case taking $G$ as the vertex-disjoint union of the $m$ graphs defined in the proof will do the job.

The last paragraph thus gives that the maximum possible value of $\omega(G^t)$ when $G$ is triangle-free is equal to $R(3; t) := R(3, 3, \ldots, 3)$ where $t$ denotes the number of 3's in the argument of $R(.)$. Thus the largest possible value of $C_{\mathrm{OR}}(G)$ when $G$ is triangle-free is equal to $\lim_{t \to \infty} \sqrt[t]{R(3; t)}$. It is a famous and still open problem of Erdős what the value of this limit is. In particular, he offered \$ 250 for its determination and \$ 100 for deciding whether it is finite or infinite, cf. [47] and see also the lecture by Noga Alon [3] between 46:58 and 48:32.

# 8 Types, typical sequences and asymptotic graph parameters within a given distribution

The following notions and statements are discussed in more detail in the book by Csiszár and Körner [11], most of what is written in this section follows the discussion there.

**Definition 17** *The type of a sequence* $\boldsymbol{x} = (x_1, \ldots, x_t) \in \mathcal{X}^t$ *is the probability distribution* $P$ *over* $\mathcal{X}$ *expressing the relative frequency of the elements of* $\mathcal{X}$ *in* $\boldsymbol{x}$*, i.e., for which*

$$P(a) = \frac{|\{i : x_i = a\}|}{t}$$

*for every* $a \in \mathcal{X}$*.*

**Definition 18** *Given a probability distribution* $P$ *on the finite set* $\mathcal{X}$ *we say that a sequence* $\boldsymbol{x} \in \mathcal{X}^t$ *is* $(P, \varepsilon)$*-typical for some* $\varepsilon > 0$ *if*

$$\forall a \in \mathcal{X} \quad \left| \frac{|\{i : x_i = a\}|}{t} - P(a) \right| < \varepsilon.$$

One of the main properties of $(P, \varepsilon)$-typical sequences is that if the probability of a $t$-length sequence $\boldsymbol{x} = x_1 \ldots x_t$ is defined as

$$P^t(\boldsymbol{x}) = \Pi_{i=1}^t P(x_i)$$

then asymptotically the probability of the set of $(P, \varepsilon)$-typical sequences tends to 1. (The probability of a set of sequences is meant to be the sum of the probabilities of the sequences in the set.)

It will also be important for us that the number of possible types of $t$-length sequences is only polynomial in $t$ as expressed by the following easy lemma.

**Lemma 10** (Type Counting Lemma [11]) *The number of different types of* $t$*-length sequences over an alphabet* $\mathcal{X}$ *of size* $n$ *is less than* $(t + 1)^n$*.*

*Proof.* The number of appearances of a given letter $a \in \mathcal{X}$ in a sequence $\boldsymbol{x} \in \mathcal{X}^t$ is one of the numbers $0, 1, \ldots, t$. So even if we could choose these numbers independently for every $a \in \mathcal{X}$ (that we clearly cannot, as the number of appearances of all letters should add up to $t$) we would have $(t + 1)$ independent choices $n$ times giving the upper bound $(t + 1)^n$. □

Note that the upper bound proved in the previous lemma is indeed only polynomial in $t$. It is exponential in $n$, but in most of our investigations this $n$, the size of our alphabet, is a constant. As a first application of the Type Counting Lemma we show that the entropy of a probability distribution $P$ is just the asymptotic exponent of the number of sequences having the given type $P$.

*Notation:* The set of $t$-length sequences of type $P$ is denoted by $T_P^t$. The set of $(P, \varepsilon)$-typical $t$-length sequences will be denoted by $T_{[P]_\varepsilon}^t$.
The entropy of the probability distribution $P = (p_1, \ldots, p_n)$ is

$$H(P) := \sum_{i=1}^n p_i \log \frac{1}{p_i}.$$

**Lemma 11** *For every type $P$ of sequences in $\mathcal{X}^t$ we have*

$$(t+1)^{-|\mathcal{X}|}2^{tH(P)} \leq |T_P^t| \leq 2^{tH(P)}.$$

*In particular,*

$$\lim_{k\to\infty} \frac{1}{kt} \log |T_P^{kt}| = H(P).$$

(Note that the role of $k$ above is only to avoid the problem that could be caused by $T_P^\ell = \emptyset$ for those lengths $\ell$ on which $P$ cannot be the exact type of a sequence.)

*Proof.* The probability of a sequence $\boldsymbol{x}$ of type $P$ according to the product distribution coming from $P$ is

$$P^t(\boldsymbol{x}) = \Pi_{i=1}^t P(x_i) = \Pi_{a\in\mathcal{X}}[P(a)]^{tP(a)} = 2^{-tH(P)}.$$

Since

$$1 \geq P^t(T_P^t) = |T_P^t|P^t(\boldsymbol{x}) = |T_P^t|2^{-tH(P)}$$

for some $\boldsymbol{x} \in T_P^t$ this already implies

$$|T_P^t| \leq 2^{tH(P)}.$$

To prove the lower bound it is enough to show that if $Q$ is any probability distribution on $\mathcal{X}$ that can be the type of a sequence of length $t$ then $P^t(T_Q^t) \leq P^t(T_P^t)$, therefore $|T_P^t|2^{-tH(P)} \geq \frac{1}{(t+1)^{|\mathcal{X}|}}$ implying the lower bound. This follows by writing

$$\frac{P^t(T_Q^t)}{P^t(T_P^t)} = \frac{\frac{t!}{\Pi_{a\in\mathcal{X}}[tQ(a)]!}\Pi_{a\in\mathcal{X}}[P(a)]^{tQ(a)}}{\frac{t!}{\Pi_{a\in\mathcal{X}}[tP(a)]!}\Pi_{a\in\mathcal{X}}[P(a)]^{tP(a)}} = \Pi_{a\in\mathcal{X}}\frac{[tP(a)]!}{[tQ(a)]!}P(a)^{t(Q(a)-P(a))}$$

from which applying the inequality

$$\frac{n!}{m!} \leq n^{n-m}$$

we can continue by

$$\frac{P^t(T_Q^t)}{P^t(T_P^t)} \leq \Pi_{a\in\mathcal{X}}\frac{[tP(a)]!}{[tQ(a)]!}P(a)^{t(Q(a)-P(a))} \leq$$

$$\Pi_{a\in\mathcal{X}}[tP(a)]^{t(P(a)-Q(a))}P(a)^{t(Q(a)-P(a))} = \Pi_{a\in\mathcal{X}}t^{t(P(a)-Q(a))} = 1.$$

The second statement is an immediate consequence of the first one and the easy fact that $\lim_{t\to\infty} \frac{-|\mathcal{X}|}{t}\log(t+1) = 0$. □

*Thirteenth lecture* (October 21, 2022)

Now we introduce Shannon OR-capacity within a given distribution $P$ that will be denoted by $C_{\mathrm{OR}}(G,P)$. The idea is that when we look at $\omega(G^t)$ in the definition of $C_{\mathrm{OR}}(G)$ we could also restrict attention to $\omega(G^t[T_P^t])$ for some distribution $P$ that is realizable as a type of sequences of length $t$. (Why this will be useful, we will see later.) Since not every distribution is realizable on a given length (or if there are irrational values among the $P(a)$ probabilities then on no finite length) we will actually look at $\omega(G^t[T_{[P]_\varepsilon}^t])$ considering the lim sup as $t$ goes to infinity and then let $\varepsilon$ tend to 0.

**Definition 19** *The Shannon OR-capacity of graph $G$ within distribution $P$ is defined as*

$$C_{\mathrm{OR}}(G,P) = \lim_{\varepsilon\to 0}\limsup_{t\to\infty} \sqrt[t]{\omega(G^t[T^t_{[P]_\varepsilon}])}.$$

An important consequence of the Type Counting Lemma is the following.

**Proposition 8**

$$C_{\mathrm{OR}}(G) = \max_P C_{\mathrm{OR}}(G,P).$$

*Sketch of proof.* Let $\mathcal{P}_t$ denote the set of probability distributions that can be the type of a sequence of length $t$. By the Type Counting Lemma we can write

$$\omega(G^t) \le (t+1)^{|V|} \max_{P\in\mathcal{P}_t} \omega(G^t[T^t_{[P]_0}])$$

Thus

$$C_{\mathrm{OR}}(G) = \lim_{t\to\infty} \sqrt[t]{\omega(G^t)} \le$$

$$\limsup_{t\to\infty} \sqrt[t]{(t+1)^{|V|} \max_{P\in\mathcal{P}_t} \omega(G^t[T^t_{[P]_0}])} = \limsup_{t\to\infty} \sqrt[t]{\max_{P\in\mathcal{P}_t} \omega(G^t[T^t_{[P]_0}])}.$$

Accepting the at least intuitively clear statement that $\max_{P\in\mathcal{P}_t} \omega(G^t[T^t_{[P]_0}])$ can be assumed to be taken by a sequence of distributions tending to a limiting distribution $P$ we obtain the statement of the Proposition. □

Notice that a "within a distribution" version of many other graph parameters can be introduced in a similar manner and an analog of Proposition 8 will also be true in those other cases. For example, we can speak about the Witsenhausen rate within a distribution $P$ that will be

$$R(G,P) := \lim_{\varepsilon\to 0}\limsup_{t\to\infty} \sqrt[t]{\chi(G^{\wedge t}[T^t_{[P]_\varepsilon}])}$$

and we will have

$$R(G) = \max_P R(G,P).$$

The usefulness of these within a distribution parameters lies in the fact that the graphs $G^t[T^t_{[P]_0}]$ are vertex-transitive. This is obviously so as the vertices are $t$-length sequences in $V^t$ that are just permutations of each other. Since any permutation of the coordinates induces an automorphism of $G^t$ and any vertex of $G^t[T^t_{[P]_0}]$ can be mapped to any other vertex of $G^t[T^t_{[P]_0}]$ by a permutattion, this graph is indeed vertex-transitive. But that means that for example

$$\chi_f(G^{\wedge t}[T^t_{[P]_0}]) = \frac{|V(G^{\wedge t}[T^t_{[P]_0}])|}{\alpha(G^{\wedge t}[T^t_{[P]_0}])} = \frac{|V(G^{\wedge t}[T^t_{[P]_0}])|}{\omega(\overline{G}^t[T^t_{[P]_0}])}.$$

Using also that

$$\chi(G^{\wedge t}[T^t_{[P]_0}]) \le \chi_f(G^{\wedge t}[T^t_{[P]_0}])(1+\ln\alpha(G^{\wedge t}[T^t_{[P]_0}])$$

and taking $t^{\text{th}}$ root and limits one gets that

$$R(G,P)\cdot C_{\mathrm{OR}}(\overline{G},P) = 2^{H(P)}$$

holds for every $P$. This immediately implies for example Lemma 9. We get only an inequality there because we cannot be sure that $R(G,P)$ and $C_{\mathrm{OR}}(\overline{G},P)$ will be maximized by the same $P$. But if, for example, the graph is vertex-transitive, implying that the uniform distribution should achieve both maxima, we obtain that we should have equality. This, in particular, implies Corollary 23.

If the above thoughts are applied to the OR-product instead of the AND-product, then we obtain an inequality which—by the McEliece–Posner theorem and the easy fact that $\lim_{t\to\infty} \sqrt[t]{\alpha(G^t)} = \alpha(G)$ simply by the multiplicativity of $\alpha(G)$ under the OR-product—could be written as

$$\chi_f(G,P) \cdot \alpha(G,P) = 2^{H(P)}.$$

In the next section we sill see that what we temporarily denoted here as $\chi_f(G,P)$ is actually an interesting notion.

# 9 Graph Entropy

Graph entropy was introduced by János Körner in [30]. The forthcoming lectures will largely follow the discussion in the survey paper [55]. In those cases we will simply refer to the corresponding part of [55], the details can be read directly from there (see a link on the course website).

## 9.1 Different formulas for graph entropy

The information theoretic motivation for introducing graph entropy is discussed in Section 2 of [55]. This leads to the definition we also repeat here.

**Definition 20**

$$H(G,P) = \lim_{t\to\infty} \min_{U\subseteq V^t, P^t(U)>1-\varepsilon} \frac{1}{t} \log \chi(G^t[U]) \tag{1}$$

*where* $P^t(U) = \sum_{\mathbf{x}\in U} P^t(\mathbf{x})$.

In the light of the remark given in the paragraph following Definition 18 we can see that $H(G,P)$ is just the logarithmic version of what we denoted by $\chi_f(G,P)$ at the end of the previous section. And thus we can write

$$H(G,P) = H(P) - \log \alpha(G,P).$$

The definition of mutual information is given as Definition 2.3 in [55] that we quote here.

**Definition 21** *Let $X$ and $Y$ be two discrete random variables taking their values on some (possibly different) finite sets and consider the random variable formed by the pair $(X,Y)$. The functional*

$$I(X \wedge Y) = H(X) + H(Y) - H((X,Y))$$

*is called the mutual information of the variables $X$ and $Y$.*

The main result of [30] (apart from the very idea of introducing $H(G,P)$) is the theorem expressing the equivalence of the above definition of graph entropy with the following expression that is given in Definition 1.2".

**Theorem 26** (Körner [30])

$$H(G,P) = \min I(X \wedge Y) \tag{2}$$

*where the minimization is over pairs of random variables $(X,Y)$ having the following properties. The variable $X$ is taking its values on the vertices of $G$, while $Y$ on the stable sets of $G$ and their joint distribution is such that $X \in Y$ with probability 1. Furthermore, the marginal distribution of $X$ on $V(G)$ is identical to the given distribution $P$.*

By what is said above, Theorem 26 is proved if one shows that the following equality holds.

$$\log \alpha(G, P) = \max_{P_X = P, X \in Y \in S(G)} H(X, Y) - H(Y), \tag{3}$$

where the maximization is over pairs of random variables $(X, Y)$ having the following properties. The variable $X$ is taking its values on the vertices of $G$, while $Y$ on the independent sets of $G$ and their joint distribution is such that $X \in Y$ with probability 1. Furthermore, the marginal distribution of $X$ on $V(G)$ is identical to the given distribution $P$. (Note that the last conditions are just those already said in Theorem 26.)

*Sketch of proof.* Let $P_t \to P$ be a sequences of probability distributions converging to $P$ such that $P_t$ is realizable as a type of $t$-length sequences. Then an independent set in $G^t[T^t_{[P_t]_0}]$ is given by a set of $t$-length sequences of vertices of $G$ whose type is $P_t$ and at every coordinate all vertices in the sequences belong to the same independent set $A \in S(G)$. Thus these sequences also define a sequence of the independent sets, let $A_i$ denote the one belonging to coordinate $i$. Therefore we can also think about these sequences as sequences of pairs $(x_i, A_i)$ where $A_i$ is the same for all coordinates within an independent set of $G^t[T^t_{[P_t]_0}]$. Consider now all possible $t$-length sequences of pairs $(x_i, A_i)$ with $x_i \in V(G), A_i \in S(G), x_i \in A_i$. Each of these sequences have a type that is a distribution on $V(G) \times S(G)$. We know that for all our sequences the "marginal type" of their $x_i$ part is $P_t$ since we started with sequences that are vertices of $G^t[T^t_{[P_t]_0}]$. Yet the type of the sequences formed by the pairs $(x_i, A_i)$ need not all be the same, since the "marginal type" of the $A_i$ parts may differ. However, in one independent set of $G^t[T^t_{[P_t]_0}]$ this will also be fixed as the $A_i$-sequence must be constant there. What we are trying to do is to choose this fixed joint type so that our independent set is as large as possible. To this end we now also fix a joint type $W_t$ and calculate the size of our independent set assuming that this is the actual joint type for our independent set. Then we will get an expression for this size in terms of the fixed joint type and then maximize over all possible choices. So let us see the calculation.

Once $W_t$ is fixed the asymptotic exponent of the number of sequences of our pairs $(x_i, A_i)$ is $H(W_t)$, that is, their number is $2^{tH(W_t)+o(t)}$. Then the "marginal type" $Q_t$ of the $A_i$ part of our sequences is already given by

$$Q_t(A) = \sum_{x \in V(G)} W_t(x, A)$$

and the number of different sequences we have if considering only the $A_i$ part of our sequences is $2^{tH(Q_t)+o(t)}$. (We may imagine a matrix whose rows are indexed by the $P_t$-type $t$-length sequences of the $x_i$'s, the columns by the $Q_t$-type sequences of the $A_i$'s and every entry is either empty–in case for the corresponding $x_i$-sequence and $A_i$-sequence $x_i \notin A_i$ for some $i$–or contains a unique sequence of $(x_i, A_i)$ pairs, the one determined by the row and the column of this entry. Then the number of non-empty entries will be the same for every column. Similarly, the number of non-empty entries will be the same for every row.) So we have $2^{tH(W_t)+o(t)}$ sequences of pairs $(x_i, A_i)$ and they can be partitioned into $2^{tH(Q_t)+o(t)}$ partition classes according to their subsequences given by their $A_i$ components. (Such a partition class belongs to a column of the matrix mentioned above.) These partition classes have the same size because they can be permuted into each other (as each $A$ appears together with a given $x$ the same number of times that is given by $tW_t(x, A)$ ). Since for an independent set we need to have the $A_i$-part fixed, what we are interested in is the number of sequences in one partition class. Since those have the same size, this is simply the ratio of the number of all sequences and the number of partition classes, i.e.,

$$\frac{2^{tH(W_t)+o(t)}}{2^{tH(Q_t)+o(t)}} = 2^{t(H(W_t)-H(Q_t))+o(t)}.$$

Ignoring the $o(t)$ term that will disappear when taking the accordingly normalized limit, we get that the logarithm of the maximum asymptotic size of an independent set is indeed

$$\log \alpha(G,P) = \max_{P_X=P, X\in Y\in S(G)} H(X,Y) - H(Y)$$

where the maximization is under the conditions said above. Putting this into the already stated equality $H(G,P) = H(P) - \log \alpha(G,P)$ and using that $H(P) = H(P_X)$ was one of our conditions for the above maximization, we get that

$$H(G,P) = H(P) - \log \alpha(G,P) =$$

$$\min_{P_X=P, X\in Y\in S(G)} H(X) + H(Y) - H(X,Y) =$$

$$\min_{P_X=P, X\in Y\in S(G)} I(X \wedge Y)$$

as stated. □

More than two decades after its original definition it turned out that graph entropy can be defined by yet a third at first look completely different formula. Recall that the vertex packing polytope is the convex hull of characteristic vectors of independent sets, cf. Definition 7

**Theorem 27** ([12])

$$H(G,P) = \min_{\boldsymbol{a}\in VP(G)} \sum_{i=1}^{n} p_i \log \frac{1}{a_i}.$$

In the proof we will use that if $(X,Y)$ is a random variable with joint distribution $W$, marginal distributions $P$ for $X$ and $Q$ for $Y$ and conditional distribution $R(y|x)$ for $Y$ given $X = x$ then we can write the mutual information of $X$ and $Y$ as

$$I(X\wedge Y) = \sum_x P(x)\log\frac{1}{P(x)} + \sum_y Q(y)\log\frac{1}{Q(y)} - \sum_{x,y} W(x,y)\log\frac{1}{W(x,y)} =$$

$$\sum_x P(x)\log\frac{1}{P(x)} + \sum_y \left(\sum_x P(x)R(y|x)\right)\log\frac{1}{Q(y)} -$$

$$-\sum_{x,y} P(x)R(y|x)\log\frac{1}{P(x)R(y|x)} =$$

$$\sum_x P(x)\left(\log\frac{1}{P(x)} + \sum_y R(y|x)\log\frac{1}{Q(y)} - \log\frac{1}{P(x)} - \sum_y R(y|x)\log\frac{1}{R(y|x)}\right) =$$

$$= -\sum_x P(x)\sum_y R(y|x)\log\frac{Q(y)}{R(y|x)}.$$

With this in mind the reader can already understand the proof below that is essentially the one presented on pages 4-5 of [55].

*Proof of Theorem 27.* First we prove that

$$\min_{X\in Y\in S(G),P_X=P} I(X\wedge Y)\geq \min_{\boldsymbol{a}\in VP(G)} -\sum_{i=1}^{n} p_i \log a_i.$$

Let the minimum of the left hand side be achieved by a pair $(X,Y)$. Let $Q$ be the marginal distribution of $Y$ for this pair. We know that the marginal distribution for $X$ is $P$ and denote the conditional distribution of $Y$ given $X$ by $R$. By the definition of mutual information and the identities shown above we can write

$$I(X\wedge Y)=-\sum_{i=1}^{n} p_i \sum_{J:i\in J\in S(G)} R(J|i)\log\frac{Q(J)}{R(J|i)}\geq -\sum_{i=1}^{n} p_i \log \sum_{J:i\in J\in S(G)} Q(J)$$

where the inequality follows from Jensen's inequality by the concavity of the log function. Since the vector defined by $a_i=\sum_{J:i\in J\in S(G)} Q(J)$ $(i=1,...,n)$ is in $VP(G)$ we have $I(X\wedge Y)\geq \min_{\boldsymbol{a}\in VP(G)} -\sum_{i=1}^{n} p_i \log a_i$.

To prove the reverse inequality let $\boldsymbol{a}$ be the minimizing vector for the right hand side. Since $\boldsymbol{a}$ is in $VP(G)$ it can be represented as $a_i=\sum_{J:i\in J\in S(G)} Q'(J)$ where the vector of coefficients $Q'(J)$ can be regarded as a probability distribution on $S(G)$. Define the conditional probabilities $R'(J|i)=\frac{Q'(J)}{a_i}$ for $i\in J$ and 0 otherwise. We define another probability distribution on $S(G)$ by $Q^*(J)=\sum_{i=1}^{n} p_i R'(J|i)$. Having in mind the pair $(X,Y)$ with marginal distribution $\mathrm{dist}(X)=P$ and conditional distribution $\mathrm{dist}(Y|X)=R'$ (thereby $\mathrm{dist}(Y)=Q^*$) we can write

$$\min_{X\in Y\in S(G),P_X=P} I(X\wedge Y)\leq -\sum_{i=1}^{n} p_i \sum_{J:i\in J\in S(G)} R'(J|i)\log\frac{Q^*(J)}{R'(J|i)} \tag{4}$$

using again the identities presented just before the proof. (The inequality comes from the fact that the distribution we work with here is not necessarily the one attaining the minimum on the left hand side.) By the concavity of the log function

$$\sum_{J\in S(G)} Q^*(J)\log\frac{Q'(J)}{Q^*(J)}\leq \log\sum_{J\in S(G)} Q'(J)=\log 1=0,$$

hence

$$-\sum_{i,J} p_i R'(J|i)\log Q^*(J)\leq -\sum_{i,J} p_i R'(J|i)\log Q'(J).$$

Thereby we can continue (4) as

$$\min_{X\in Y\in S(G),P_X=P} I(X\wedge Y)\leq -\sum_{i,J} p_i R'(J|i)\log\frac{Q'(J)}{R'(J|i)}=-\sum_{i=1}^{n} p_i\log a_i$$

completing the proof. □

*Fifteenth lecture* (October 28, 2022)

## 9.2 On properties and applications

Some important properties of graph entropy are summarized in Section 3 of [55]. We will particularly need Lemmas 3.1 (monotonicity), 3.2 (sub-additivity), Corollary 3.4 (formula for $H(G, P)$ in terms of the entropies of the components) and Proposition 3.5 (entropy of the complete graph). Note that the latter also follows from Shannon's Source Coding Theorem that we get back when looking at the coding problem that led to the definition of graph entropy specifically for the complete graph. We also remark that although Corollary 3.4 is presented in [55] as a consequence of Lemma 3.3 there that we did not cover on the lecture, Corollary 3.5 of [55] already follows from the mentioned Lemmas 3.1 and 3.2.

The subadditivity of graph entropy has several applications, in the course we illustrate that by a new proof of a result of Krichevskii [33] due to Newman, Ragde, and Wigderson [48, 49], cf. Subsection 5.2, in particular Theorem 5.3 in [55] (on pages 18–20).

*Sixteenth lecture* (November 4, 2022)

## 9.3 About exact additivity

Next we discuss some of the structural results given in Section 6 of [55]. We focus on the content of Subsection 6.2, in particular on Theorems 6.4 and 6.5. To understand the latter we need the notion of convex corners and the definition of their entropy given in Definitions 4.2 and 4,3 (page 12) of [55]. We repeat them here for the reader's cenvenience.

**Definition 22** *A set $\mathcal{A} \in \Re^n_+$ is called a convex corner if it is closed, convex, has a non-empty interior, and satisfies the property that if $0 \le a'_i \le a_i$ for $i = 1, .., n$ then $\mathbf{a} \in \mathcal{A}$ implies $\mathbf{a}' \in \mathcal{A}$.*

**Definition 23** *For a convex corner $\mathcal{A} \subseteq \Re^n_+$ its entropy with respect to a probability distribution $P = (p_1, ..., p_n)$ is defined as*

$$H_{\mathcal{A}}(P) = \min_{\mathbf{a} \in \mathcal{A}} \sum_{i=1}^{n} p_i \log \frac{1}{a_i}. \tag{5}$$

Clearly, $H(G, P) = H_{VP(G)}(P)$ by Theorem 27.

The main question we want to answer is when we will have equality in the subadditivity inequality of graph entropy for a graph and its complement. The following definition gives a name to this property.

**Definition 24** *A graph $G$ is called strongly splitting if for every probability distribution $P$ on its vertex set*

$$H(G, P) + H(\bar{G}, P) = H(P). \tag{6}$$

The main theorem we are going to prove is this (see as Theorem 6.5 in [55]).

**Theorem 28** ([12]) *A graph $G$ is strongly splitting if and only if it is perfect.*

The only if part has the following quick proof.

**Proposition 9** *Let $P_U$ be the uniform distribution on the vertices of a minimal imperfect graph $G$. Then*

$$H(G, P_U) + H(\bar{G}, P_U) > H(P_U).$$

*Proof.* Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be the vectors from $VP(G)$ and $VP(\bar{G})$ achieving the entropy of $G$ and $\bar{G}$, respectively, with respect to $P_U$. Clearly, $\sum_i a_i \leq \alpha(G)$ and $\sum_i b_i \leq \omega(G)$. So we have

$$H(G, P_U) + H(\bar{G}, P_U) = \sum_i \frac{1}{n} \log \frac{1}{a_i} + \sum_i \frac{1}{n} \log \frac{1}{b_i} = \log \frac{1}{(\prod_i a_i)^{\frac{1}{n}} (\prod_i b_i)^{\frac{1}{n}}} \geq$$

$$\log \frac{1}{\frac{\alpha(G)}{n} \frac{\omega(G)}{n}} > \log n,$$

where the first inequality follows from the relation of the arithmetic and geometric mean and the second from Theorem 4. □

Note that Proposition 9 already implies that no imperfect graph is strongly splitting because we can always concentrate a uniform distribution on the vertex set of a minimal imperfect subgraph of an imperfect graph.

Both the above and the reverse implication will easily follow from Theorem 6.4 in [55] that is not proven there, but here we are giving a proof. To state the theorem we need some definitions.

**Definition 25** (Fulkerson [18]) *Let $\mathcal{A} \in \Re^n_+$ be a convex corner. The antiblocker $\mathcal{A}^*$ of $\mathcal{A}$ is defined as*

$$\mathcal{A}^* = \{\boldsymbol{b} \in \Re^n_+ : \boldsymbol{b}^T \cdot \boldsymbol{a} \leq 1 \forall \boldsymbol{a} \in \mathcal{A}\}.$$

Note that we have

$$FVP(G) = [VP(\overline{G})]^*$$

by definition.

**Definition 26** *The pair of convex corners $\mathcal{A}, \mathcal{B} \subseteq \Re^n_+$ is said to form a generating pair if for every probability distribution $P = (p_1, .., p_n)$ there exist $\boldsymbol{a} \in \mathcal{A}$ and $\boldsymbol{b} \in \mathcal{B}$ satisfying $a_i b_i = p_i$ for $(i = 1, .., n)$.*

The result behind the characterization of strongly splitting graphs is the following theorem.

**Theorem 29** ([12]) *For convex corners $\mathcal{A}, \mathcal{B} \in \Re^n_+$ the following two statements are equivalent:*

*(i)* $\mathcal{A}^* \subseteq \mathcal{B}$

*(ii)* $(\mathcal{A}, \mathcal{B})$ *is a generating pair*

We remark that in Theorem 6.4 of [55] a third equivalent statement is also given, but here we restrict ourselves to only that part of the theorem that is needed to prove Theorem 28.

*Proof.* (i) $\Rightarrow$ (ii): Consider an arbitrary probability distribution $P = (p_1, \ldots, p_n)$ and let $\boldsymbol{a}_P$ be the element of $\mathcal{A}$ attaining the minimum in the definition of $H_{\mathcal{A}}(P)$. Then for any $\boldsymbol{a} \in \mathcal{A}$ and $0 \le \lambda \le 1$ we have $\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{a}_P \in \mathcal{A}$ since $\mathcal{A}$ is convex. Thus by the choice of $\boldsymbol{a}_p$ the derivative of $\sum_{i=1}^n p_i \log \frac{1}{\lambda a_i + (1-\lambda) a_{P,i}}$ at $0^+$ is non-negative. The derivative is

$$\frac{d}{d\lambda} \sum_{i=1}^n p_i \log \frac{1}{\lambda a_i + (1-\lambda) a_{P,i}} =$$

$$\sum_{i=1}^n p_i \left( -\frac{1}{(\lambda a_i + (1-\lambda) a_{P,i})^2} (\lambda a_i + (1-\lambda) a_{P,i})(a_i - a_{P,i}) \right) =$$

$$-\sum_{i=1}^n p_i \frac{a_i - a_{P,i}}{\lambda a_i + (1-\lambda) a_{P,i}}$$

and evaluating it at $\lambda = 0$ we obtain

$$\sum_{i=1}^n p_i \frac{a_{P,i} - a_i}{a_{P,i}} \ge 0.$$

From here we can write

$$1 = \sum_{i=1}^n p_i \ge \sum_{i=1}^n a_i \frac{p_i}{a_{P,i}}$$

that means that defining the vector $\boldsymbol{b}_P$ by

$$b_{P,i} := \frac{p_i}{a_{P,i}}$$

we have

$$\boldsymbol{a} \cdot \boldsymbol{b}_P \le 1$$

for our arbitrarily chosen $\boldsymbol{a} \in \mathcal{A}$. In other words this means that $\boldsymbol{b}_P \in \mathcal{A}^*$ and so the distribution $P$ can indeed be obtained as $\boldsymbol{a}_P \circ \boldsymbol{b}_P$ where $\boldsymbol{a}_P \in A$ and $\boldsymbol{b}_P \in A^* \subseteq B$, so $\mathcal{A}$ and $\mathcal{B}$ form a generating pair.

(ii) $\Rightarrow$ (i): Now assume $\mathcal{A}^* \not\subseteq \mathcal{B}$ therefore $\exists \boldsymbol{c} \in \mathcal{A}^* \setminus \mathcal{B}$ and thus also some $\boldsymbol{d} \in \mathcal{B}^*$ such that $\boldsymbol{c} \cdot \boldsymbol{d} > 1$. (The existence of such a $\boldsymbol{d}$ follows from the fact that $(\mathcal{B}^*)^* = \mathcal{B}$.) Define the distribution $P$ by

$$p_i := \frac{c_i d_i}{\boldsymbol{c} \cdot \boldsymbol{d}}.$$

We show that this $P$ cannot be written in the form $\boldsymbol{a} \circ \boldsymbol{b}$ for any $\boldsymbol{a} \in \mathcal{A}, \boldsymbol{b} \in \mathcal{B}$ and therefore $(\mathcal{A}, \mathcal{B})$ is not a generating pair in this case. Assume for contradiction that such a pair $\boldsymbol{a} \in A, \boldsymbol{b} \in \mathcal{B}$ does exist. Then we can write

$$1 \ge \sum_{i=1}^n a_i c_i = \sum_{i=1}^n \frac{p_i}{b_i} c_i \ge \frac{1}{\sum_{i=1}^n p_i \frac{b_i}{c_i}} = \frac{\boldsymbol{c} \cdot \boldsymbol{d}}{d_i b_i} > 1$$

which is a contradiction. Here the first inequality follows from $\boldsymbol{c} \in \mathcal{A}^*$, the second is a consequence of the inequality between the arithmetic and harmonic mean applied to the numbers $\frac{b_i}{c_i}$ and the last inequality is a consequence of $\boldsymbol{b} \in \mathcal{B}, \boldsymbol{d} \in \mathcal{B}^*$ and $\boldsymbol{c} \cdot \boldsymbol{d} > 1$. The contradiction we arrived to proves the statement. □

*Proof of Theorem 28.* The strongly splitting property means that for every probability distribution $P = (p_1, \ldots, p_n)$ we have

$$\sum_{i=1}^{n} p_i \log \frac{1}{p_i} = H(P) = H(G, P) + H(\overline{G}, P) = \sum_{i=1}^{n} p_i \log \frac{1}{a_i b_i}$$

for appropriate $\boldsymbol{a} \in VP(G)$ and $\boldsymbol{b} \in VP(\overline{G})$. If $VP(G)$ and $VP(\overline{G})$ form a generating pair then there are vectors $\boldsymbol{a} \in VP(G)$ and $\boldsymbol{b} \in VP(\overline{G})$ for which $a_i b_i = p_i$ for every $i$ and these vectors should be the minimizing vectors attaining the value of $H(G, P)$ and $H(\overline{G}, P)$, respectively, otherwise the right hand side above would be smaller than the left hand side which we know to be impossible. So $G$ will be strongly splitting if $VP(G)$ and $VP(\overline{G})$ form a generating pair that by Theorem 29 is equivalent to

$$VP(G) \supseteq [VP(\overline{G})]^* = FVP(G)$$

that is equivalent to $G$ being perfect by Theorem 3. So perfect graphs are strongly splitting.

We have already seen a short proof of the reverse implication but it also follows from what is said here if we take into account that $VP(G) \subset FVP(G)$ always holds thus $\boldsymbol{a} \cdot \boldsymbol{b} \leq 1$ is always true for $\boldsymbol{a} \in VP(G), \boldsymbol{b} \in VP(\overline{G})$ and also that for any probability distribution $Q = (q_1, \ldots, q_n)$ we have $\sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log \frac{1}{q_i}$, since Jensen's inequality implies

$$\sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \leq \log \left( \sum_{i=1}^{n} p_i \frac{1}{p_i} \right) = \log 1 = 0$$

with equality if and only if $q_i = p_i$ for all $i$. So if $H(G, P)$ and $H(\overline{G}, P)$ are attained by $\boldsymbol{a} \in VP(G)$ and $\boldsymbol{b} \in VP(\overline{G})$, respectively, then setting $q_i := \frac{a_i b_i}{\boldsymbol{a} \cdot \boldsymbol{b}}$ we have

$$H(G, P) + H(\overline{G}, P) = \sum_{i=1}^{n} p_i \log \frac{1}{a_i b_i} \geq$$

$$\sum_{i=1}^{n} p_i \log \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{a_i b_i} = \sum_{i=1}^{n} p_i \log \frac{1}{q_i} \geq \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = H(P)$$

with eqality if and only if $\boldsymbol{a} \cdot \boldsymbol{b} = 1$ and $q_i = p_i$. So for $G$ being strongly splitting we must have that $VP(G)$ and $VP(\overline{G})$ form a generating pair and this only happens when $G$ is perfect. $\square$.

Note that the above argument shows that $H(G, P) + H(\overline{G}, P) = H(P)$ for any given $P = (p_1, \ldots, p_n)$ if and only if $p_i = a_i b_i$ for some $\boldsymbol{a} \in VP(G)$ and $\boldsymbol{b} \in VP(\overline{G})$. Furthermore, it also shows that such a $P$ for which $\forall i \colon p_i \neq 0$ exists if and only if the vertex set of $G$ can be covered by independent sets $A_1, \ldots, A_m$ and also by cliques $Q_1, \ldots, Q_r$ such that $A_i \cap Q_j \neq \emptyset$ for any $i \in \{1, \ldots, m\}, j \in \{1, \ldots, r\}$, cf. Definitions 6.1, 6.2 and Proposition 6.2 in the survey paper [55] for more details.

*Seventeenth lecture* (November 8, 2022)

We mention without proof that the problem of exact additivity is also solved in the general case when the two graphs involved are not necessarily complementary. An interesting feature of this result is that the complementary case plays a special role in, it seems that it cannot be avoided to solve it separately. This result is the following.

**Theorem 30** ([32]) *For two graphs $F$ and $G$ on the same vertex set $V$ one has*

$$H(F \cup G, P) = H(F, P) + H(G, P)$$

*for every $P$ if and only if the following three conditions are satisfied.*

*(a) $E(F) \cap E(G) = \emptyset$;*

*(b) if $F \cup G$ induces a clique on some $U \subseteq V$ then the graphs induced by $F$ and $G$ on $U$ are perfect;*

*(c) no $P_3$ (path on 3 vertices) of $F \cup G$ has one edge in $F$ and one edge in $G$.*

We remark that the above result implies the following theorem due to Cameron, Edmonds and Lovász [8].

**Theorem 31** ([8]) *If the edges of a complete graph are three-colored (with red, blue, and green, say) in such a way that no three-colored triangle occurs and the graph formed by the red edges, and the graph formed by the blue edges are both perfect then so is the graph formed by the green edges.*

Here is the proof that Theorem 30 implies Theorem 31. Assume that the conditions of Theorem 31 are satisfied and denote the graphs formed by the red, blue and green edges by $R, B$ and $G$ respectively. Note that the perfectness of $R$ and the Perfect Graph Theorem implies $R$ and $G$ both induce perfect graphs on those subsets of their common vertex set $V$ on which their union is complete, therefore perfect the conditions of Theorem 31 imply by Theorem 30 that for every probability distribution on $V$ we have

$$H(R, P) + H(G, P) = H(R \cup G, P).$$

Since $B$ is perfect we also have that

$$H(B, P) + H(\overline{B}, P) = H(B, P) + H(R \cup G, P) = H(P)$$

holds for every probability distribution. The last two equalities imply

$$H(R, P) + H(G, P) + H(B, P) = H(P)$$

for every probability distribution. But this means that we must also have

$$H(G, P) + H(\overline{G}, P) = H(G, P) + H(R \cup B, P) = H(P)$$

for every probability distribution, so $G$ must be perfect. (The last equality follows from knowing that the left hand side is not more than $H(G, P) + H(R, P) + H(B, P) = H(P)$ by sub-additivity, but also cannot be less than $H(P)$, again by sub-additivity.)

## 9.4 Another application: A job scheduling problem

We show another possible application of Theorem 29 to a non-linear optimization problem. It appears in the work of Denardo, Hoffman, Mackenzie, and Pulleyblank [13].

Let $H = (V, E)$ be a hypergraph with all vertices covered by at least one edge and $\mathcal{A}$ be the convex corner defined by the characteristic vectors of the edges of $H$, i.e.,

$$\mathcal{A} = \text{conv}\{\mathbf{1}_A : A \in E\}.$$

Our aim is to find $\boldsymbol{w} \in \mathcal{A}$ minimizing the value of

$$a(\boldsymbol{w}) = \max_{A \in E} \sum_{i \in A} \frac{l_i}{w_i},$$

where the $l_i$'s are given prescribed values.

In the example of [13] the elements of $V$ are the edges of an acyclic directed graph, while $E$ consists of its subsets that form directed paths from a given source to a given sink. The graph describes a project, the edges are the single tasks that should be done for having done the whole work. (Such a graph is often called and taught about as a PERT Chart, where PERT abbreviates "Program Evaluation and Review Technique".) We have a number of workers and know that the $i$th task would last $l_i$ time units if all workers worked on that. The project can obviously be finished in $\sum_{i \in V} l_i$ time units if we let all workers work on each task together until it is completed. This method may have, however, some practical disadvantages. One of those is that each worker has to deal with each single task for a while and another one is that sometimes the workers have to switch from one task to another with no connection between the two. The authors of [13] show that the work can be done in $\sum_{i \in V} l_i$ time units also without these disadvantages. Let $w_i$ mean the proportion of workers working on task $i$. Then task $i$ will be done in $\frac{l_i}{w_i}$ time units and the whole project will be finished in $a(\boldsymbol{w})$ time units. The restriction to $\boldsymbol{w} \in \mathcal{A}$ describes the condition that each worker should work on consecutive tasks.

**Theorem 32** *(Denardo, Hoffman, Mackenzie, Pulleyblank [13])*

$$\min_{\boldsymbol{w} \in \mathcal{A}} a(\boldsymbol{w}) = \sum_{i \in V} l_i.$$

*Furthermore, the above minimum is achieved by the $\boldsymbol{w}$ that achieves the minimum in the definition of $H_{\mathcal{A}}(P)$, where $P$ is the probability distribution on $V$ defined by $p_i = \frac{l_i}{\sum_{i \in V} l_i}$.*

*Proof.* The direction $a(\boldsymbol{w}) \geq \sum_{i \in V} l_i$ for every feasible $\boldsymbol{w}$ is easy. Let $\boldsymbol{w} = \sum_{A \in E} \alpha_A \mathbf{1}_A$ and consider the weighted mean

$$\sum_{A \in E} \alpha_A \sum_{i \in A} \frac{l_i}{w_i} = \sum_{A \in E} \alpha_A \sum_{i \in A} \frac{l_i}{\sum_{A \ni i} \alpha_A} = \sum_{i \in V} l_i,$$

therefore $a(\boldsymbol{w})$ which (not considering the coefficients) is the largest member of the weighted sum on the left hand side cannot be smaller than $\sum_{i \in V} l_i$.

Now consider the $\boldsymbol{w}$ achieving the minimum in the definition of $H_{\mathcal{A}}(P)$. Define the vector $\boldsymbol{b}$ by $b_i = \frac{l_i}{w_i \sum_{i \in V} l_i}$. It follows from Theorem 29 that $\boldsymbol{b}$ is in the antiblocker of $\mathcal{A}$, i.e., for $\boldsymbol{b}$ and every $A \in E$ one has $\sum_{i \in A} b_i \leq 1$ . But this implies that $\sum_{i \in A} \frac{l_i}{w_i} \leq \sum_{i \in V} l_i$ for every $A \in E$. Hence we must have equality for all those $A \in E$ that appear with positive coefficient in the representation of $\boldsymbol{w}$. This also implies that this $\boldsymbol{w}$ must minimize $a(\boldsymbol{w})$, too. □

## 9.5 Kahn and Kim's application to sorting

This application, due to Kahn and Kim [26], is one of the most beautiful ones of graph entropy that provided a deterministic algorithm to a problem which was solved before only in case when randomization is allowed. For the details we refer to Section 7 of [55].

The theorems of Kahn and Kim are stated but not proved in [55]. Here we show the proof of one of them, which is Theorem 7.3 in [55]. It states the following ($G_S$ is the comparability graph of the poset $S$ and $P_U$ is the uniform distribution on its elements).

**Theorem 33** *(Kahn and Kim [26]) For any partial order $S$ and $x, y \in V$ that are incomparable according to $S$ one has*

$$\min\{H(G_{S(x<y)}, P_U), H(G_{S(y<x)}, P_U)\} \leq H(G_S, P_U) + \frac{2}{n}. \tag{7}$$

In the algorithmic context this theorem provides a way to answer to all questions of the type "Is $x < y$?" in such a way that (using the notation introduced in Section 7 of [55]) the need for $\Omega(\log e(S))$ questions is forced.

*Proof.* Let $\boldsymbol{a} \in VP(G_S)$ be the vector attaining $H(G_S, P_U)$. Let us denote $x$ by $x_1$ and $y$ by $x_2$. Let $T, V, W, Z$ be the following four subsets of $S$.

$$T := \{x \in S : x < x_1\},\ V := \{x \in S : x > x_1\},$$

$$W := \{x \in S : x < x_2\},\ Z := \{x \in S : x > x_2\}.$$

Let $K \subset T$ be a clique (thus a chain in $S$) that maximizes the weight

$$w(K) := \sum_{i: x_i \in K} a_i.$$

Choose the chains $L \subseteq V, M \subseteq W, N \subseteq Z$ in a similar manner. Then

$$K \cup \{x_1\} \cup L, \quad \text{and} \quad M \cup \{x_2\} \cup N$$

are also chains, thus

$$w(K) + w(L) + a_1 \leq 1$$

and

$$w(M) + w(N) + a_2 \leq 1.$$

This implies that at least one of the following two inequalities should hold:

$$w(K) + w(N) + \frac{a_1 + a_2}{2} \leq 1$$

or

$$w(L) + w(M) + \frac{a_1 + a_2}{2} \leq 1.$$

We may assume without loss of generality that the first one holds. Assuming this we show that the following vector $\boldsymbol{a}'$ we have

$$\boldsymbol{a}' \in VP(G_{S(x_1<x_2)}),$$

where $G_{S(x_1<x_2)}$ stands for the comparability graph of the poset we obtain from $S$ if we add the relation $x_1 < x_2$ to it along with all its consequences due to transitivity. Let

$$a'_i = \begin{cases} \frac{a_i}{2} & \text{if } i = 1, 2 \\ a_i & \text{otherwise.} \end{cases}$$

By showing $\boldsymbol{a}' \in VP(G_{S(x_1<x_2)})$ we will be done since that implies

$$H(G_{S(x_1<x_2)}, P_U) \leq \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{a'_i} = \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{a_i} + \frac{2}{n} = H(G_S, P_U) + \frac{2}{n}.$$

Since $G_{S(x_1<x_2)}$ is perfect (as it is a comparability graph), we have

$$VP(G_{S(x_1<x_2)}) = FVP(G_{S(x_1<x_2)}),$$

thus for the above it is enough to show that for any clique $Q$ in $G_{S(x_1<x_2)}$ we have

$$\sum_{i:x_i \in Q} a'_1 \leq 1.$$

To this end let $Q$ be a non-extendable clique (clearly, it is enough to consider those) in $G_{S(x_1<x_2)}$. We consider two cases.
First assume that $\{x_1, x_2\} \not\subseteq Q$. Then $Q$ is also a clique in $G_S$ (note that to say this we use that $Q$ is non-extendable, for example with $x_1$ and $x_2$, and therefore it is surely a clique already in $G_S$), therefore

$$w'(Q) := \sum_{i:x_i \in Q} a'_i \leq w(Q) \leq 1.$$

In the second case $\{x_1, x_2\} \subseteq Q$. In that case let

$$K' := \{x \in Q : x <_{s(x_1<x_2)} x_1\}$$

and

$$N' := \{x \in Q : x_2 <_{s(x_1<x_2)} x\}.$$

Then we have $K' \subseteq T$ and $N' \subseteq Z$ and both $K'$ and $N'$ are chains in $S$. This implies by the choice of $K$ and $N$ that

$$w(K') \leq w(K) \text{ and } w(N') \leq w(N).$$

Since there cannot exist any $x \in S$ for which $x_1 <_{S(x_1<x_2)} < x_2$, we must have

$$Q = K' \cup N' \cup \{x_1, x_2\}.$$

But then

$$w'(Q) = w'(K') + w'(N') + \frac{a_1}{2} + \frac{a_2}{2} =$$

$$w(K') + w(N') + \frac{a_1 + a_2}{2} \leq w(K) + w(N) + \frac{a_1 + a_2}{2} \leq 1.$$

Thus $\boldsymbol{a}' \in FVP(G_{S(x_1<x_2)}) = VP(G_{S(x_1<x_2)})$ and the proof is complete. □

*Nineteenth lecture* (November 14, 2022)

# 10 Sperner capacity

We are going to generalize Shannon's graph capacity notion to directed graphs. Formally this generalization will be quite straightforward. The question remains: Why is it interesting? We will see that there are multiple answers to that. One reason is that it leads to interesting mathematical problems: even when the Shannon capacity of an undirected graph is unknown, we might solve the problem for certain orientations nontrivially. And also the other way around: there are undirected graphs for which the value of the Shannon capacity is easy to determine, yet the directed case is nontrivial. (As we will see later, an example for the latter phenomenon will be the case of a cyclically oriented triangle, while examples for the former case are provided by certain orientations of longer odd cycles.) Two other reasons will become clear when we generalize our definitions to graph families. It will turn out that certain natural extremal set theory questions can be formulated in the language of capacities of directed graphs. And we will also see that there is also a natural information theoretic meaning of these capacity values.

## 10.1 Definitions

First we generalize the the OR-product and OR-power to digraphs (=directed graphs).

**Definition 27** *For two digraphs $F$ and $G$ their OR-product $F \cdot G$ is defined by*

$$V(F \cdot G) = V(F) \times V(G)$$

*and*

$$E(F \times G) = \{((f,g),(f',g')) : f,f' \in V(F), g,g' \in V(G), \\ (f,f') \in E(F) \text{ or } (g,g') \in E(G)\}.$$

*The $t^{\text{th}}$ OR-power if a digraph $D$, denoted $D^t$, is meant to be the $t$-fold OR-product of $D$ with itself.*

Note that the only difference between Definitions 27 and **??** is that the orientation of the edges is taken into account. Also note that two vertices of the OR-product of two digraphs can be connected by edges in both directions even if this does not happen in any of the two graphs whose product is taken. If we consider an undirected graph equivalent to the directed graph that contains all of its edges with both of their possible orientations (and no other edge), then the OR product of undirected graphs simply becomes a special case. This kind of phenomenon will be true for all the notions we are going to consider here for digraphs. Having this in mind, the generalization of Shannon capacity is already straightforward.

Notation: A digraph on $n$ vertices containing an edge between any two distinct vertices in both directions will be called a symmetric clique. The number of vertices in a largest symmetric clique in a digraph $D$ will be denoted by $\omega_s(D)$. (We call this the *symmetric clique number* of $D$.

**Definition 28** *The (non-logarithmic) Sperner capacity of a digraph $D$ is defined as*

$$Sp(D) := \limsup_{t\to\infty} \sqrt[t]{\omega_s(D^t)}.$$

We do not distinguish now OR and AND capacities because the literature is much more unified in case of Sperner capacity. Using the notation $Sp$ only emphasizes that the graph after it should be (considered as) a directed graph. Just as in case of $C_{\mathrm{OR}}(G)$ it follows from Fekete's Lemma that the limsup in the definition is actually a limit.

It is straightforward from the above definition that if $D$ is the undirected version of a digraph $\vec{D}$ then we always have

$$Sp(\vec{D}) \leq C_{\mathrm{OR}}(D). \tag{8}$$

The reason is that $\vec{H} \subseteq \vec{D}$ obviously implies $Sp(\vec{H}) \leq Sp(\vec{D})$ and, as we already mentioned, $C_{\mathrm{OR}}(D)$ is just the Sperner capacity of the graph that contains all edges of $D$ in both directions, so $\vec{D}$ is certainly a subgraph of it.

The name *Sperner capacity* comes from the fact that a symmetric clique in the $t^{\mathrm{th}}$ power of the simplest possible directed graph that consists of two vertices (labeled 0 and 1) and a unique directed edge between them is formed by elements of $\{0,1\}^t$ that–if considered as characteristic vectors of a $t$-element set–form a Sperner system: the sets they describe have the property that none of them is the subset of another one.

## 10.2 Bounds

Above we introduced the symmetric clique number of a digraph. Another special type of directed clique will also be important for us. These are called *transitive cliques.*

**Definition 29** *An oriented complete graph $T$ is called a tournament if all pairs of vertices are connected in exactly one of the two possible directions. A tournament is called a* transitive tournament *if its vertices can be labeled by distinct positive integers (by $1, 2, \ldots, n$, say) in such a way that for any two of the labels, $i$ and $j$, the edge between $i$ and $j$ is oriented towards the larger of the two numbers $i$ and $j$. A transitive tournament subdigraph $T$ of a digraph $D$ will be called a transitive clique of $D$ and the number of vertices of a largest transitive clique of $D$ will be its transitive clique number denoted by $\omega_{\mathrm{tr}}(D)$.*

**Proposition 10** *If $T_n$ is a (one could say* the*, as it is unique up to isomorphism) transitive tournament on $n$ vertices then*

$$Sp(T_n) = n.$$

*Proof.* By

$$Sp(T_n) \leq C_{\mathrm{OR}}(K_n) = n$$

we have that $n$ is an upper bound. To prove the reverse inequality assume that $V(T_n) = \{1, \ldots, n\}$ and

$$(i,j) \in E(T_n) \Leftrightarrow i < j.$$

Let $b$ be any positive integer between $t$ and $tn$ and consider the set $Q_b$ of all vertices $x_1x_2\ldots x_t \in V(T_n^t)$ for which $\sum_{i=1}^t x_i = b$. Observe that $Q_b$ induces a symmetric clique in $T_n^t$: indeed, if $y_1\ldots y_t, z_1\ldots z_t$ are two different vertices in

$Q_b$ then there must exist $1 \le r, s \le t$ such that $y_r < z_s$ and $y_s > z_r$, otherwise we could not have $\sum_{i=1}^t y_i = \sum_{i=1}^t z_i$. This implies that any two vertices are connected in both directions in $Q_b$ so it is a symmetric clique.

Since the $nt - t + 1$ sets $Q_b$ belonging to the $nt - t + 1$ possible values of $b$ partition $V(T_n^t)$, the largest one of them has at least $\frac{n^t}{nt-t+1}$ elements. This implies that

$$\omega_s(T_n^t) \ge \frac{n^t}{nt - t + 1}$$

and thus

$$Sp(T_n) \ge \lim_{t\to\infty} \sqrt[t]{\frac{n^t}{nt - t + 1}} = n.$$

Since $n$ was also an upper bound this completes the proof. □

**Corollary 34** *For any digraph $D$ we have*

$$Sp(D) \ge \omega_{\mathrm{tr}}(D).$$

The smallest digraph which is not a transitive tournament is a cyclically oriented triangle that we denote here by $Y_3$. The above bounds give us

$$2 \le Sp(Y_3) \le 3.$$

It was first proven in [7] that the above lower bound is tight. An alternative proof was given in [6] which was then generalized by Alon [1] who proved the following result.

**Theorem 35** *For every directed graph $D$ we have*

$$Sp(D) \le \min\{\Delta_+(D) + 1, \Delta_-(D) + 1\}.$$

*Here $\Delta_+(.)$ and $\Delta_-(.)$ stands for the maximum outdegree and maximum inderee, respectively.*

*Proof.* We define a polynomial of $t$ variables $f_{\boldsymbol{a}}(\boldsymbol{x})$ for each $\boldsymbol{a} \in [V(D)]^t = V(D^t)$ as follows. We assume that $V(D) = \{1, 2, \dots, n\}$.

$$f_{\boldsymbol{a}}(\boldsymbol{x}) := \Pi_{i=1}^t \Pi_{j\in N_+(a_i)}(x_i - j),$$

where $N_+(v)$ is the outneighborhood of vertex $v$.

Let $Q$ be a largest symmetric clique in $D^t$. Observe that if $\boldsymbol{a}, \boldsymbol{b} \in Q$ then

$$f_{\boldsymbol{a}}(\boldsymbol{b}) \ne 0 \;\Leftrightarrow\; \boldsymbol{a} = \boldsymbol{b}.$$

Indeed, if $\boldsymbol{a} \ne \boldsymbol{b}$ then for some $i \in \{1, \dots, t\}$ we must have $b_i \in N_+(a_i)$ resulting in $f_{\boldsymbol{a}}(\boldsymbol{b}) = 0$. On the other hand, if we do not allow loops (and we do not), $f_{\boldsymbol{a}}(\boldsymbol{a}) \ne 0$. This implies that the set of polynomials

$$\{f_{\boldsymbol{a}}(\boldsymbol{x})\}_{\boldsymbol{a}\in Q}$$

are linearly independent, therefore $|Q|$ cannot be larger than the dimension of the linear space they generate. The latter can be bounded from above by the number of monomials of the form $x_1^{j_1} \dots x_t^{j_t}$ that can be present in our polynomials. Since by the definition of $f_{\boldsymbol{a}}(\boldsymbol{x})$ we must have $0 \le j_i \le \Delta_+(D)$ for every $i$ in these monomials, we can have at most $(\Delta_+(D) + 1)^t$ such monomials. Thus we have obtained

$$|Q| \le (\Delta_+(D) + 1)^t.$$

Since $Q$ was chosen to be a largest symmetric clique in $D^t$ this gives

$$Sp(D) = \lim_{t\to\infty} \sqrt[t]{\omega_s(D^t)} \le \lim_{t\to\infty} \sqrt[t]{(\Delta_+(D)+1)^t} = \Delta_+(D)+1.$$

Observing that reversing all edges in a digraph does not change the value of its Sperner capacity we can see that the above also implies

$$Sp(D) \le \Delta_-(D)+1$$

completing the proof. □

Denoting the cyclically oriented $n$-cycle by $Y_n$ the above theorem has the following immediate consequence.

**Corollary 36**

$$Sp(Y_n) = 2.$$

Corollary 36 solves, in particular, the Sperner capacity problem for the cyclically oriented triangle. Since the triangle has only two non-isomorphic orientations and the other one is transitive for which Corollary 34 gives the solution, we know the Sperner capacity for every oriented version of the 3-length cycle. The 5-length cycle is already different: it has four non-isomorphic orientations and the results presented so far give the Sperner capacity only for one of them (the cyclic one). For the rest we know only that 2 is a lower bound by Corollary 34 and $\sqrt{5}$ is an upper bound by (8) and Corollary 10.

*Twentieth lecture* (November 18, 2022)

To get a better upper bound we will introduce first another graph coloring parameter.

## 10.3 Local chromatic number

Imagine a university department where every professor is the instructor of exactly one course and everybody has enough expertise in a few other courses to give a lecture on it if the course instructor needs to be substituted for some reason. The department head makes the schedule of the courses so that no one has a lecture at the same time as another course on which (s)he might potentially substitute the instructor and tries to make it so that if everyone is asked to be available at the time of such lectures on which (s)he is a potential substitute then the total number of timeslots when any of the professors should either lecture or be available is as small as possible. The idea is that it is extremely unlikely that two professors should be substituted at the same time so it is all right if two different courses the same person could lecture on as a substitute are scheduled to the same time. We are interested in the behaviour of the above mentioned minimum value as a parameter of the situation that is of the graph in which two vertices are connected if they represent two courses whose professors could substitute each other. At this point the astute reader may ask why it is obvious that if one professos can substitute another one then this other one can also substitute the first one. It is indeed an important point and we will see that it will worth to look at also the asymmetric situation but let us first assume that there is such a symmetry: if Professor A can substitute Professor B then this is mutual, Professor B can also substitute Professor A.

The investigation of the graph parameter we obtain this way was initiated in a paper by Erdős, Füredi, Hajnal, Komjáth, Rödl and Seress [15] and can be defined formally as follows.

**Definition 30** ([15]) *The* local chromatic number $\chi_{\text{loc}}(G)$ *of a graph* $G$ *is the maximum number of different colors appearing in the closed neighbourhood of any vertex, minimized over all proper colorings of* $G$*. Formally,*

$$\chi_{\text{loc}}(G) := \min_{c} \max_{v \in V(G)} |\{c(u) : u \in \hat{N}(v)\}|,$$

*where* $\hat{N}(v)$*, the closed neighborhood of the vertex* $v \in V(G)$*, is the set of those vertices of* $G$ *that are either adjacent or equal to* $v$ *and* $c$ *runs over all proper colorings of* $G$*.*

It is obvious that if we use only $\chi(G)$ colors for the coloring then no closed neighborhood can contain more than that, so

$$\chi_{\text{loc}}(G) \leq \chi(G)$$

must hold. It is also clear that in any proper coloring with exactly $\chi(G)$ colors each color class must contain a vertex that sees all the other colors in its neighborhood. If this was not so for a color class then every vertex of that color could be recolored to one of the other colors resulting in a proper coloring with fewer than $\chi(G)$ colors, an obvious contradiction. So if there exists a proper coloring that attains a local chromatic number strictly smaller than $\chi(G)$ then it must use strictly more colors than minimally necessary for a proper coloring. At first look one may feel that this is hard to imagine that one can make each closed neighborhood having fewer than $\chi(G)$ colors by "wasting" colors, that is, using more colors than it is necessary for a proper coloring. Yet, surprisingly, this is the case, there are graphs for which the local chromatic number is smaller than the chromatic number.

**Definition 31** *The universal local coloring graph* $U(m, r)$ *with parameters* $m, r$ *(where* $m \geq r$ *and both of them are positive integers) is defined as follows.*

$$V(U(m,r)) = \{(x, A) \colon x \in [m], A \subseteq [m], x \notin A, |A| = r - 1\}$$

$$E(U(m,r)) = \{\{(x, A), (y, B)\} \colon x \in B, y \in A\}.$$

**Proposition 11**

$$\chi_{\text{loc}}(U(m,r)) = r.$$

*Proof.* Let $c \colon V(U(m,r)) \to [m]$ be the coloring for which

$$c \colon (x, A) \mapsto x$$

for every vertex $(x, A) \in V(U(m,r))$. Note that by the definition of $U(m,r)$ this gives $c(\hat{N}(x, A)) = A \cup \{x\}$ which by $|A| = r-1, x \notin A$ implies that $|\hat{N}(x, A)| = r$ for all $(x, A) \in V(U(m,r)$, therefore proving $\chi_{\text{loc}}(U(m,r)) \leq r$. On the other hand, if $T$ is an $r$-element subset of $[m]$ then the $r$ vertices in $\{(x, A) : \{x\} \cup A = T\}$ induce an $r$-vertex clique in $U(m,r)$. Since $\chi_{\text{loc}}(K_r) = r$ for obvious reasons, this shows that $\chi_{\text{loc}}(U(m,r)) \geq r$, so we must have $\chi_{\text{loc}}(U(m,r)) = r$ as stated. □

It is easy to see that $U(m, 2)$ is bipartite for every $m \geq 2$, so these graphs have their local chromatic nuber equal to their chromatic number. However, we are going to prove, that for $r = 3$ the value of $\chi(U(m,r)$ is unbounded. Since, obviously, $U(m,r) \subseteq U(m+1,r)$ this actually means that $\lim_{m\to\infty} \chi(U(m,3)) = \infty$. To prove this we will use the following notion.

**Definition 32** *For a digraph $D$ its line graph $L(D)$ is defined as follows.*

$$V(L(D)) = E(D)$$

*and*

$$E(L(D)) = \{\{(a,b),(c,d)\} : b = c \text{ or } a = d\}.$$

Thus two edges are connected as vertices of the line graph if the head of one of them is identical to the tail of the other one.

It is a well-known exercise (see e.g. as Exercise 9.26a in [41]) that if $\vec{D}$ is a digraph with underlying undirected graph $D$ then we have

$$\chi(L(\vec{D})) \geq \log \chi(D).$$

The proof is as follows: consider an optimal proper coloring of $L(\vec{D})$ with $h := \chi(L(\vec{D}))$ colors as an edge-coloring of $\vec{D}$ and color each vertex $v$ of $D$ with the set of colors that appear on the edges that have $v$ as their tail. This is a proper vertex-coloring of $D$ since if $(u,v) \in E(\vec{D})$ then the color of this edge is an element of the set used to color $u$ and cannot be an element of the set used to color $v$ (otherwise the edge-coloring would not be proper). Since the possible number of subsets of the set of $h$ colors is $2^h$ this gives

$$\chi(D) \leq 2^h$$

which is equivalent to $\chi(L(\vec{D})) = h \geq \log \chi(D)$ that we wanted to prove.

**Theorem 37** *For every $m \geq 3$ we have*

$$\chi(U(m,3)) \geq \log\log m.$$

*Proof.* The vertices of $U(m,3)$ are of the form $(x,A)$ where $A = \{a_1,a_2\}$ is a 2-element set. Consider only the subgraph induced by those vertices for which we have $a_1 < x < a_2$ and observe that this subgraph is isomorphic to $L(\vec{L}(\vec{T}_m))$ where $\vec{L}(\vec{D})$ is the line graph of digrah $\vec{D}$ with its natural orientation: edges of the form $\{(u,v),(v,w)\}$ are oriented as $((u,v),(v,w))$ and $\vec{T}_m$ denotes the transitive tournament on the $m$ vertices $1,2,\ldots,m$. Applying the above mentioned result twice this observation gives $\chi(U(m,3)) \geq \chi(L(\vec{L}(\vec{T}_m))) \geq \log\log m$. □

Note that since $\chi_{\text{loc}}(U(m,3)) = 3$ this means that the gap between the local chromatic number and the chromatic number can be arbitrarily large.

It is a more or less trivial observation that a graph $G$ admits a proper coloring with $m$ colors attaining that its local chromatic number is at most $r$ if and only if a homomorphism from $G$ to $U(m,r)$ exists. Indeed, if the claimed coloring exists then any vertex colored with color $j$ and having the at most $r-1$ colors one can collect in a set $H$ can be mapped to $(j,A)$ for some $A \supseteq H$ to obtain such a homomorphism. Similarly, if such a homomorphism exists then coloring every vertex that are mapped to a vertex $(j,A)$ of $U(m,r)$ by $j$ we get a proper coloring attaining local chromatic number at most $r$. Using this observation we prove the following result.

**Theorem 38** ([31]) *Every finite simple graph $G$ satisfies*

$$\chi_f(G) \leq \chi_{\text{loc}}(G).$$

*Proof.* First we prove that

$$\chi_f(U(m,r)) = r.$$

We have already seen that $\omega(U(m,r)) \geq r$ (since vertices $\{(x, T \setminus \{x\}) : x \in T\}$ form an $r$-clique for any $r$-subset $T$ of $[m]$). This implies $\chi_f(U(m,r)) \geq r$. To prove the reverse inequality we observe that $U(m,r)$ is vertex-transitive since any vertex $(x,A)$ can be moved to any other one $(y,B)$ by an appropriate permutation of the elements of $[m]$. Observe also that the vertices in $\{(x,A)\colon x = \min(A \cup \{x\})\}$ induce an independent set in $U(m,r)$. (This is because if $(x,A)$ and $(y,B)$ are connected, then w.l.o.g. $x < y$ and so $x \in B$ implies that $y$ cannot be the minimal element of $B \cup \{y\}$ as the latter also contains $x$ which is smaller.) The size of this set is exactly $\frac{|V(U(m,r))|}{r}$ and so this is a lower bound on $\alpha(U(m,r))$ providing

$$\chi_f(U(m,r)) = \frac{|V(U(m,r))|}{\alpha(U(m,r))} \leq r$$

proving the statement. □

Since we know $C_{\mathrm{OR}}(G) \leq \chi_f(G)$ Theorem 38 implies that the local chromatic number bounds the Shannon OR-capacity from above. But the proof itself already shows that this upper bound will give nothing new as it is always a worse (or at least not better) upper bound than the fractional chromatic number. The novelty comes in the case when we consider directed graphs. It is quite straightforward to define a directed version of the local chromatic number for digraphs. Indeed, in our introducing "story" there is nothing that should make us assume that if Professor X can substitute Professor Y then this should be mutual and Professor Y should also be able to teach the course of Professor X. This leads to the following definition.

**Definition 33** ([31]) *The directed local chromatic number $\chi_{\mathrm{loc,d}}(D)$ of a digraph $D$ is defined as*

$$\chi_{\mathrm{loc,d}}(D) := \min_c \max_{v \in V(D)} |\{c(u) : u \in \hat{N}_+(v)\}|,$$

*where $\hat{N}_+(v)$, is the closed outneighborhood of the vertex $v$, that is the set of vertices including $v$ and those $u \in V(D)$ for which $(v,u) \in E(D)$ and the minimization is over all proper colorings $c$ of (the underlying undirected graph of) $D$.*

**Theorem 39** ([31]) *For any (loopless) digraph $D$*

$$Sp(D) \leq \chi_{\mathrm{loc,d}}(D).$$

*Twenty first lecture* (November 22, 2022)

*Proof of Theorem 39.* The proof is a straightforward generalization of that of Theorem 35.

Fix a proper coloring $c : V(D) \to [m]$ (with an appropriate choice of $m$) that attains the minimum in the definition of $\chi_{\mathrm{loc,d}}(D)$.

We define a polynomial of $t$ variables $f_{\boldsymbol{a}}(\boldsymbol{x})$ for each $\boldsymbol{a} \in [V(D)]^t = V(D^t)$ as follows.

$$f_{\boldsymbol{a}}(\boldsymbol{x}) := \Pi_{i=1}^t \Pi_{j \in c(N_+(a_i))}(x_i - j),$$

where $c(N_+(v))$ is the set of colors appearing in the outneighborhood of vertex $v$ in the coloring $c$.

Let $Q$ be a largest symmetric clique in $D^t$. Observe that if $\boldsymbol{a}, \boldsymbol{b} \in Q$ then

$$f_{\boldsymbol{a}}(c(b_1), \ldots, c(b_t)) \neq 0 \;\Leftrightarrow\; \boldsymbol{a} = \boldsymbol{b}.$$

Indeed, if $\boldsymbol{a} \neq \boldsymbol{b}$ then for some $i \in \{1, \ldots, t\}$ we must have $b_i \in N_+(a_i)$ and therefore $c(b_i) \in c(N_+(a_i))$ resulting in $f_{\boldsymbol{a}}(c(b_1), \ldots, c(b_t)) = 0$. On the other hand, since we do not allow loops and the coloring $c$ is proper, $f_{\boldsymbol{a}}(c(a_1), \ldots, c(a_t)) \neq 0$. This implies that the set of polynomials

$$\{f_{\boldsymbol{a}}(\boldsymbol{x})\}_{\boldsymbol{a} \in Q}$$

are linearly independent, therefore $|Q|$ cannot be larger than the dimension of the linear space they generate. The latter can be bounded from above by the number of monomials of the form $x_1^{j_1} \ldots x_t^{j_t}$ that can be present in our polynomials. Since by the definition of $f_{\boldsymbol{a}}(\boldsymbol{x})$ we must have $0 \leq j_i \leq \chi_{\mathrm{loc,d}}(D) - 1$ for every $i$ in these monomials, we can have at most $(\chi_{\mathrm{loc,d}}(D))^t$ such monomials. Thus we have obtained

$$|Q| \leq (\chi_{\mathrm{loc,d}}(D))^t.$$

Since $Q$ was chosen to be a largest symmetric clique in $D^t$ this gives

$$Sp(D) = \lim_{t \to \infty} \sqrt[t]{\omega_s(D^t)} \leq \lim_{t \to \infty} \sqrt[t]{(\chi_{\mathrm{loc,d}}(D))^t} = \chi_{\mathrm{loc,d}}(D)$$

completing the proof. □

Note that by the observation mentioned at the end of the proof of Theorem 35, namely that reversing all edges in a digraph does not change the value of its Sperner capacity, we can see that the local chromatic number of the digraph we obtain by reversing all edges of $D$ is also an upper bound on $Sp(D)$.

Let us call the orientation of an odd cycle *alternating* if it has only one vertex that is neither a source nor a sink. Note that up to isomorphism there is only one such orientation for each odd cycle.

**Proposition 12** *Let $\vec{C}_{2k+1}$ be an oriented odd cycle. We have*

$$\chi_{\mathrm{loc,d}}(\vec{C}_{2k+1}) = 2 \quad \Leftrightarrow \quad \text{the orientation of } \vec{C}_{2k+1} \text{ is not alternating.}$$

*If the orientation of $\vec{C}_{2k+1}$ is alternating then $\chi_{\mathrm{loc,d}}(\vec{C}_{2k+1}) = 3$.*

*Proof.* Consider the following auxiliary graph $T := T(\vec{C}_{2k+1})$ on vertex set $V(C_{2k+1})$ in which two vertices $u, v \in V(T)$ are adjacent if and only if $\exists s \in V(T)\colon (s, u), (s, v) \in E(\vec{C}_{2k+1})$. There exists a proper coloring $c$ of $\vec{C}_{2k+1}$ in which no closed outneighborhood contains more than two colors if and only if $c$ can be constant on every connected component of $T$. Such a coloring exists if and only if no two neighboring vertices of $C_{2k+1}$ are in the same component of $T$. Two neighboring vertices of $C_{2k+1}$ can be in the same component of $T$ if and only if the orientation $\vec{C}_{2k+1}$ is such that two neighboring vertices $a, b \in V(\vec{C}_{2k+1})$ are the endpoints of a path $a - u_1 - u_2 - \ldots u_{k-1} - b$ in which any two neighboring vertices are the two outneighbors of a vertex of $\vec{C}_{2k+1}$ that (being a source, that is, outdegree 2 vertex) is not on the path. This happens if and only if $\vec{C}_{2k+1}$ is an alternating odd cycle. This proves that this is the only oriented version of $C_{2k+1}$ for which $\chi_{\mathrm{loc,d}}(\vec{C}_{2k+1}) > 2$. □

Theorem 39 and Proposition 12 have the following immediate consequence.

**Corollary 40** *If $\vec{C}_{2k+1}$ is an oriented odd cycle that is not alternating, then*

$$Sp(\vec{C}_{2k+1}) = 2.$$

□

Since we obviously have $\chi_{\rm loc,d}(\vec{D}) \leq \chi_{\rm loc}(D) \leq \chi(D)$ (where $D$ is the underlying undirected graph of $\vec{D}$), it is also clear from Proposition 12 that for the alternatingly oriented $(2k+1)$-cycle $\vec{C}_{2k+1}^{(\rm alt)}$ we have

$$\chi_{\rm loc,d}(\vec{C}_{2k+1}^{(\rm alt)}) = 3.$$

Thus for their Sperner capacity our best upper bound is still $C_{\rm OR}(C_{2k+1})$, or the best upper bound of the latter in case we want an exact numerical value. For $k = 1$ this value is still 3 and since an alternating 3-cycle is just a transitively oriented triangle, we know from Corollary 34 that this upper bound is sharp. We show that for $k = 2$, that is for the alternating 5-cycle we also have equality.

**Proposition 13** ([20])

$$Sp(\vec{C}_5^{(\rm alt)}) = \sqrt{5}.$$

*Proof.* We have $Sp(\vec{C}_5^{(\rm alt)} \leq \sqrt{5}$ from Corollary 10 by $Sp(\vec{C}_5^{(\rm alt)} \leq C_{\rm OR}(C_5) = \sqrt{5}$, so it is enough to prove that $\sqrt{5}$ is also a lower bound. This can be done by observing that $[\vec{C}_5^{(\rm alt)}]^2$ contains a transitive clique on 5 vertices. Indeed, if the edges of our $C_5$ are $01, 12, 23, 34, 40$ oriented so that 0 and 2 are sources (that is, vertices with outdegree two) and 1 and 3 are sinks (have outdegree zero, so 4 sends an edge to vertex 3 and receives one from vertex 0), then the clique in $C_5^2$ consisting of vertices

$$00, 24, 12, 43, 31$$

is oriented transitively so that in the above ordering every edge is oriented from left to right. So we have $\omega_{\rm tr}([\vec{C}_5^{(\rm alt)}]^2) \geq 5$ and by Corollary 34 this proves the statement. □

It turns out that the appearance of a transitive clique in the square of $\vec{C}_5^{(\rm alt)}$ is the result of a more general phenomenon. This is stated in the following theorem which makes us able to tell the exact Sperner capacity of some more graphs.

**Theorem 41** ([52]) *Let $G$ be a self-complementary graph. Then $G$ and its complementary graph $\overline{G}$ can be oriented so that their oriented versions are still isomorphic as digraphs and the tournament obtained as their union is transitive.*

Before presenting the proof of the above theorem we state its consequences on Sperner capacity. Note that when we speak about an orientation of an undirected graph then we mean that every edge gets one of the two possible orientations, thus the resulting digraph is what is called an oriented graph, one in which no pair of vertices form an edge in both directions.

**Corollary 42** *If $G$ is a self-complementary graph on $n$ vertices, then it has an orientation $\vec{G}$ for which $Sp(\vec{G}) \geq \sqrt{n}$.*

*Proof.* Orient the self-complementary graph $G$ according to Theorem 41 to obtain an oriented version both for $G$ and its complement (that we denote in this proof by $G^c$) whose union is a transitive tournament on the $n$ vertices. Label the vertices with $1, \ldots, n$ to be consistent with the transitive tournament on the vertex set obtained by the union, that is, so that every edge of this transitive tournament points toward its larger endpoint. Note that since the two oriented graph $\vec{G}$ and $\vec{G}^c$ we obtain by the orientation are isomorphic digraphs, their product is isomorphic to $\vec{G}^2$. Notice that $\vec{G} \cdot \vec{G}^c$ contains a transitive tournament on $n$ vertices since the set of sequences $\{(i, i) : i \in [n]\}$ induces a transitive clique.

Thus $\vec{G}^2$ must also contain a transitive clique on $n$ vertices, so $\omega_{\text{tr}}(\vec{G}^2) \geq n$. Since

$$Sp(\vec{G}) = \sqrt{Sp(\vec{G}^2)} \geq \sqrt{\omega_{\text{tr}}(\vec{G}^2)} \geq \sqrt{n},$$

this completes the proof. □

**Corollary 43** ([52]) *If $G$ is a self-complementary vertex-transitive graph on $n$ vertices, then it has an orientation $\vec{G}$ for which*

$$Sp(\vec{G}) = \sqrt{n}.$$

*In particular, then $G$ has an orientation for which*

$$Sp(\vec{G}) = C_{\text{OR}}(G).$$

*Proof.* We know that $Sp(\vec{G}) \leq C_{\text{OR}}(G)$ for all orientations $\vec{G}$ of $G$ and by Corollary 15 we also know that for a self-complementary vertex-transitive graph on $n$ vertices we have $C_{\text{OR}}(G) = \sqrt{n}$. This already proves that $Sp(\vec{G}) \leq \sqrt{n}$ for any orientation of $G$. Corollary 42 implies that for some orientation this upper bound is achieved. □

Before proving Theorem 41 let us summarize a bit the fairly well-understood structure of self-complementary graphs described in the papers [50, 51, 24]. If $G$ and its complementary graph $G^c$ are isomorphic, then there is a permutation $\tau$ of the elements of $V(G) = [n]$ for which $\{\tau(i), \tau(j)\} \in E(G)$ if and only if $\{i, j\} \notin E(G^c)$. Call this $\tau$ the complementing permutation of $G$. We discuss only the case when the cycle decomposition of $\tau$ contains only one cycle as understanding this will only be important for us. Then we can assume without loss of generality that this one-cycle permutation $\tau$ is just $(12\ldots n)$. Imagine thes $n$ numbers in this cyclic order around a cycle. If $\{i-1, i\}$ is an edge of $G$ then $\{i, i+1\}$ is not, $\{i+1, i+2\}$ is an edge again, etc. This already shows that $n$ should be even. In fact, irrespective the number of cycles in the complementing permutations, any self-complementary graph has $n = 4k$ or $n = 4k+1$ vertices, since $|E(K_n)| = \binom{n}{2}$ should be even. So if there is only one cycle in the complementing permutation then $n$ should be a multiple of 4.

The above also shows that if we already know whether $\{i, i+1\}$ is an edge of $G$ or of $G^c$ then it already tells us whether $\{j, j+1\}$ is an edge of $G$ or $G^c$ for every $j$. The situation is similar for any $\{i, i+k\}$ and $\{j, j+k\}$. We will not very directly use this structure but it is likely to help understanding the proof if we have this picture in mind.

*Twenty second lecture* (November 29, 2022)

*Proof of Theorem 41.* (We are following the proof described in [25].) Let the complementing permutation of the self-complementary graph $G$ be $\tau$. If $\tau$ has more than one cycle then put the cycles into some order and orient every edge between vertices belonging to different cycles consistently towards the vertex that is in the cycle coming later in this order. Then if the edges between vertices of the same cycle can be oriented so that the self-complementary graph induced by the vertices in the single cycles satisfy the statement of the theorem then the whole graph $G$ will. So it is enough to consider self-complementary graphs whose complementing permutation contains only one cycle. Assume $G$ is such a graph and the complementing permutation is just the cycle $(12\ldots n)$. We want to find a linear ordering $\sigma$ of these vertices in such a way that if $i$ precedes $j$ in this order and $\{i, j\} \in E(G)$ then $\tau(i)$ also precedes $\tau(j)$ in this order. And similarly, if $i$ precedes $j$ in $\sigma$ while $\{i, j\}$ is a non-edge of $G$ then we want that $\tau^{-1}(i)$ precedes $\tau^{-1}(j)$ in $\sigma$ (as the non-edge $\{i, j\}$ will inherit the orientation of the edge $\{\tau^{-1}(i), \tau^{-1}(j)\}$ of $G$. (We remark that the complementing permutation is not necessarily unique. But we fix one and make our construction work for that arbitrarily chosen one. In that sense we prove a little bit more than stated.)

We may assume without loss of generality that $\{1, 2\} \in E(G)$. The order $\sigma$ will be given so that its first element is 1 and that for any $1 \le k \le n-1$ the first $k$ elements on the cycle formed by $1, 2, \ldots, n$ in this cyclic order, form an arc of consecutive elements. Let $A_k$ be the set of the first $k$ elements in $\sigma$. Then $A_k = \{j+1, \ldots, n, 1, 2, \ldots, k-(n-j)\} = \{j+1, n, 1, \ldots, k+j-n\}$ for some $j$. If we follow this rule and $A_k$ has the form as above, then $\{x\} := A_{k+1} \setminus A_k$ is a single-element set, where there are only two possible values of $x$: $j$ and $k+j-n+1$. We already know that $A_1 = \{1\}$, so $\sigma$ will be completely given if we tell how to choose the $(k+1)^{\text{st}}$ element of $\sigma$ which is the unique element of $A_{k+1} \setminus A_k$ between the above two options. Here is the rule: if the two candidate elements $j$ and $k+1-n+1$ form an edge of $G$, then choose $j$ as the next element, if they are not adjacent in $G$ (that is form an edge of $G^c$) then choose $k+j-n+1$ as the next element.

We only have to show that if we orient the edges according to this order in $G$ and then let the edges of $G^c$ simply inherit the orientation of the edge that is its preimage according to the complementing permutation $\tau$ then all edges of $K_n$ will be oriented consistently with the order $\sigma$, that is the so obtained tournament $\vec{G} \cup \vec{G}^c$ will indeed be transitive.

Consider any two vertices $a$ and $b$. Assume $\{a, b\} \in E(G)$ first. Assume without loss of generality that this edge is oriented from $a$ towards $b$. Then $k := \sigma^{-1}(a) < \sigma^{-1}(b)$, so $a \in A_k$ and $b \notin A_k$. If $b < a$ then $(a+1) \in A_k, (b+1) \notin A_k$ follows, so we are guaranteed that $\sigma^{-1}(a+1) < \sigma^{-1}(b+1)$ as needed. Let $\ell(x)$ denote the minimum $\ell$ for which $x \in A_\ell$. In fact, we have $\ell(x) = \sigma^{-1}(x)$. If $a < b$, then since $\{b+1, a+1\} \notin E(G)$ our construction rule guarantees that $\ell(a+1) < \ell(b+1)$ and thus we again have $\sigma^{-1}(a+1) < \sigma^{-1}(b+1)$.

We still have to check the other case when $\{a, b\} \notin E(G)$. Assume again w.l.o.g. that $k := \sigma^{-1}(a) < \sigma^{-1}(b)$, so $a \in A_k$ and $b \notin A_k$. Now if $a < b$ then we automatically have $\tau^{-1}(a) = (a-1) \in A_k, \tau^{-1}(b) = b-1 \notin A_k$ and thus $\sigma^{-1}(a-1) < \sigma^{-1}(b-1)$ that we need now, or $a = 1$. In the latter case $\tau^{-1}(a) = \tau^{-1}(1) = n$ and since $\{n, b+1\} = \{\tau^{-1}(a), \tau^{-1}(b)\} \in E(G)$ our construction rule guarantees that $\ell(n) < \ell(b-1)$, that is $\sigma^{-1}(n) < \sigma^{-1}(b-1)$ as needed. The situation is similar, when $b < a$: we will have $\ell(a-1) < \ell(b-1)$ and thus $\sigma^{-1}(a-1) < \sigma^{-1}(b-1)$ by $\{a-1, b-1\} \in E(G)$ and the construction rule. (Note that $b = 1$ is not possible by $\sigma^{-1}(a) < \sigma^{-1}(b)$.)

Thus we have proved that all edges of $\vec{G}$ and $\vec{G}^c$ are oriented consistently with the order $\sigma$ while $\vec{G}$ and $\vec{G}^c$ are kept to be isomorphic. This completes the proof. □

# 11 Capacity of graph families

The so-called *compound channel* is a model in information theory for situations when we have a finite set of channels (or a channel with finitely many possible states that makes it behave like that many different channels; this models typically a finite set of physical situations) and we have to transfer information via one of them but we do not know in advance which one. Here "in advance" refers to the time when we need to prepare to the transmission, that is, when we agree on a code to be used with the receiver side.

When negligible error-probability is required then the fact that codes chosen randomly perform well makes the situation relatively easy to handle. The zero-error case, however, that we are going to consider here, is quite different.

From the zero-error point of view our finitely many, say $k$, channels are represented by $k$ distinct graphs on the same vertex set (which is the common input alphabet of the channels) which are the distinguishability graphs of the individual channels. We want to see how much information we can transfer per channel input if zero-error is required whichever of the channels will be used. Thus we need a code which performs well simultaneously to each of the $k$ channels.

The graph theoretic model is thus this. Given graphs $G_1, \ldots, G_k$ with $V(G_1) = \ldots = V(G_k) =: V$. how large can the largest set of $V^t$ be (in particular what is the asymptotic growth rate of such a largest set) if it induces a clique in all the graphs $G_i^t$ for $i = 1, \ldots, k$. So the zero-error capacity of the compound channel described by the graph family $\{G_1, \ldots, G_k\}$ can be defined solely in terms of these graphs and we will call that the Shannon OR-capacity of this graph family.

**Definition 34** ([10]) *Let $\mathcal{G} = \{G_1, \ldots, G_t\}$ be a family of graphs on the common vertex set $V$. Then the (non-logarithmic) Shannon OR-capacity of $\mathcal{G}$ is defined as*

$$C_{\mathrm{OR}}(\mathcal{G}) = \lim_{t\to\infty} \sqrt[t]{\omega(\cap_{i=1}^k G_i^t)}$$

*where $\cap_{i=1}^k G_i^t$ stands for the graph having vertex set $V^t = V(G_1^t) = \ldots = V(G_k^t)$ and edge set $\cap_{i=1}^k E(G_i^t)$.*

It may be hopeless to determine $C_{\mathrm{OR}}(\mathcal{G})$ for a graph family containing individual graphs whose Shannon OR-capacities are themselves unknown. But the problem is highly non-trivial also in such cases when all the $C_{\mathrm{OR}}(G_i)$ values are known, perhaps even trivial. Our main goal is to overcome this other type of difficulty that comes from combining the requirements prescribed by the several different graphs even when their individual capacities are easy to calculate.

We first give an upper bound. Recall Definition 19 in which we defined $C_{\mathrm{OR}}(G, P) = \lim_{\varepsilon\to 0} \limsup_{t\to\infty} \sqrt[t]{\omega(G^t[T^t_{[P]_\varepsilon}])}$. In a similar manner we can define

$$C_{\mathrm{OR}}(\mathcal{G}, P) = \lim_{\varepsilon\to 0} \limsup_{t\to\infty} \sqrt[t]{\omega(\cap_{i=1}^k G^t[T^t_{[P]_\varepsilon}])}$$

and we have

$$C_{\mathrm{OR}}(\mathcal{G}) = \max_P C_{\mathrm{OR}}(\mathcal{G}, P)$$

the same way as we had $C_{\mathrm{OR}}(G) = \max_P C_{\mathrm{OR}}(G, P)$ in Proposition 8. It is immediate from the definition of $C_{\mathrm{OR}}(G, P)$ that the following inequality holds.

$$C_{\mathrm{OR}}(\mathcal{G}, P) \le \min_{G_i\in\mathcal{G}} C_{\mathrm{OR}}(G, P).$$

Thus taking maximum in $P$ on both sides and using the previous equality we get the following upper bound.

**Proposition 14** ([10])

$$C_{\mathrm{OR}}(\mathcal{G}) \le \max_P \min_{G_i\in\mathcal{G}} C_{\mathrm{OR}}(G, P).$$

Somewhat surprisingly, the above bound is actually sharp. This is a very strong result by Gargano, Körner and Vaccaro [21, 22] but before turning to that we make our discussion more general.

The extension of capacities from individual graphs to graph families can also be done for Sperner capacity essentially the same way as for Shannon OR-capacity. Next we are doing this and will see that it has two independent benefits.

*Twenty third lecture* (December 2, 2022)

**Definition 35** *Let $\mathcal{G} = \{G_1, \ldots, G_k\}$ be a family of directed graphs on the common vertex set $V$. Then the (non-logarithmic) Sperner capacity of $\mathcal{G}$ is defined as*

$$Sp(\mathcal{G}) = \lim_{t\to\infty} \sqrt[t]{\omega_s(\cap_{i=1}^k G_i^t)}$$

*where $\cap_{i=1}^k G_i^t$ stands for the digraph having vertex set $V^t = V(G_1^t) = \ldots = V(G_k^t)$ and edge set $\cap_{i=1}^k E(G_i^t)$.*

It is straightforward to see that the statements of the above discussion generalize to the directed case. In particular, we can introduce

$$Sp(G, P) := \lim_{\varepsilon\to 0} \limsup_{t\to\infty} \sqrt[t]{\omega_s(G^t[T^t_{[P]_\varepsilon}])}$$

for any digraph $G$ and also the more general notion of Sperner capacity of a graph family within distribution $P$ as

$$Sp(\mathcal{G}, P) = \lim_{\varepsilon\to 0} \limsup_{t\to\infty} \sqrt[t]{\omega_s(\cap_{i=1}^k G^t[T^t_{[P]_\varepsilon}])}$$

for any family of digraphs $\mathcal{G} = \{G_1, \ldots, G_k\}$. Than we have

$$Sp(G) = \max_P Sp(G, P)$$

for any digraph $G$ and

$$Sp(\mathcal{G}) = \max_P Sp(\mathcal{G}, P)$$

for any family $\mathcal{G}$ of digraphs. Also

$$Sp(\mathcal{G}, P) \le \min_{G_i\in\mathcal{G}} Sp(G_i, P)$$

is valid for Sperner capacities as it was before for Shannon OR-capacity. Thus Proposition 14 also remains valid

**Proposition 15**

$$Sp(\mathcal{G}) \le \max_P \min_{G_i\in\mathcal{G}} Sp(G_i, P).$$

Now let us see those two benefits hinted above. We start with the one that historically was recognized later by Nayak and Rose [46]. Their main observation was that the way we defined the zero-error capacity of the compound channel above has an implicit assumption. Namely, that the receiver is aware of the state of the channel, that is, the receiver knows which graph in the family describes the distinguishabilities in the current state of the channel and uses this knowledge when decoding the received message. Nayak and Rose also observed that if this assumption is dropped then the problem still can be treated as the capacity problem of a graph family, but then we also need to consider directed graphs and the relevant capacity value will be the Sperner capacity of the digraph family so obtained. We discuss the details below.

Let $\mathcal{W} = \{W_1, \dots, W_k\}$ be a collection of stochastic matrices describing a compound channel that we thus identify with $\mathcal{W}$. Thus when this compound channel is in its state $i$ then its behaviour is described by the matrix $W_i$. In particular, the rows of $W_i$ belong to the possible input characters, the columns to the possible output characters and $W_i[s, r] = \mathrm{Prob_i}(r|s)$, that is the probability of receiving character $r$ at the output under the condition that character $s$ was sent and the channel is in its $i^{\text{th}}$ state. (From the zero-error capacity point of view the actual probability values are irrelevant, what matters is only whether a certain probability is positive or equals to zero.) Since we assume that now neither the sender nor the receiver knows the channel state in advance, it is not enough to consider distingushability for the fixed channel matrices $W_i$. It also matters whether we can receive the same output sequence $\boldsymbol{z}$ from a possible input sequence $\boldsymbol{x}$ via channel $W_i$ as from a different input sequence $\boldsymbol{y}$ when it is sent through a possibly different channel $W_j$. If we can, then the receiver may confuse $\boldsymbol{x}$ and $\boldsymbol{y}$ so these two sequences are not distinguishable, they should not appear as two possible codeword of a zero-error code for $\mathcal{W}$.

This situation can be described by introducing a directed graph $G_{i,j}$ for every ordered pair of channels $(W_i, W_j)$. The vertex set of $G_{i,j}$ is the set of input characters (which is the same for all $W_i$'s by assumption) while (denoting the common output alphabet by $\mathcal{Y}$) the edge set is defined as

$$E(G_{i,j}) = \{(a, b) : \forall r \in \mathcal{Y} \;\; W_i(r, a)W_j(r, b) = 0\}.$$

That is, no output letter $r$ can be the result of sending $a$ via channel $W_i$ and sending $b$ via channel $j$. Two codewords $\boldsymbol{x}$ and $\boldsymbol{y}$ will now be distinguishable for the compound channel $\mathcal{W}$ if there is no way to receive the same output $\boldsymbol{z}$ if $\boldsymbol{x}$ is sent through one of the channels $W_i$ and $\boldsymbol{y}$ is sent through a possibly different channel $W_j$ belonging to $\mathcal{W}$. (In the informed decoder case it was enough to distinguish them when they are sent through the same channel.) This will need that for every choice of the pair $(i, j)$ we have a coordinate $h$ for which $x_h$ when sent through channel $W_i$ cannot give the same output as $y_h$ when sent through channel $W_j$. This means exactly that $(x_h, y_h) \in E(G_{i,j})$. Thus $\boldsymbol{x}, \boldsymbol{y} \in V^t$ will be distinguishable for the compound channel $\mathcal{W}$ if and only if both $(\boldsymbol{x}, \boldsymbol{y})$ and $(\boldsymbol{y}, \boldsymbol{x})$ is an edge of $G_i^t$ for every choice of $i$ in $\{1, \dots, k\}$. In other words we need exactly that

$$(\boldsymbol{x}, \boldsymbol{y}), (\boldsymbol{y}, \boldsymbol{x}) \in E(\cap_{i=1}^k G_i^t).$$

Thus the largest possible cardinality of a zero-error code in $V^t$ for $\mathcal{W}$ is exactly $\omega_s(\cap_{i=1}^k G_i^t)$. This already shows that the zero-error capacity $C_0(\mathcal{W})$ of $\mathcal{W}$ is

$$C_0(\mathcal{W}) = Sp(\mathcal{G}(\mathcal{W})),$$

where $\mathcal{G}(\mathcal{W}) = \{G_{i,j} : i, j \in \{1, \dots, k\}\}$ is the family of graphs defined above.

*Remark.* Note that the graphs $G_{i,i}$ above has the property that every edge they contain together with their reversed version, so they can be considered as undirected graphs. And as such, they are exactly the distinguishability graphs of the individual channels $W_i$. ◇

The other benefit we mentioned can probably best be understood via some examples. Let $\mathcal{G} = \{G_1, G_2\}$ where $V = V(G_1) = V(G_2) = \{0, 1, 2\}$ and both $G_1$ and $G_2$ contains only one edge but a different one:

$$E(G_1) = \{(0, 1)\}, \quad E(G_2) = \{(0, 2)\}.$$

Let us try to understand what is the meaning of a symmetric clique in $G_1^t \cap G_2^t$ if we consider the sequences in $V^t$, that is, the vertices of our power graphs as characteristic vectors. These sequences contain a 0, a 1 or a 2 at each position, so they can be considered as characteristic vectors of pairs of sets $(A, B)$ where we have $A, B \subseteq [t]$, $A$ is the subset of coordinates where we have a 1 and $B$ is the subset of coordinates where we have a 2, so we also have $A \cap B = \emptyset$.

For a sequence $\boldsymbol{x} \in V^t$ let $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ denote the sets defined according to the above, that is

$$A(\boldsymbol{x}) = \{i : x_i = 1\} \quad B(\boldsymbol{x}) = \{i : x_i = 2\}.$$

If $\boldsymbol{x}, \boldsymbol{y} \in V^t$ are two sequences for which we have both

$$(\boldsymbol{x}, \boldsymbol{y}) \in E(G_1^t) \cap E(G_2^t) \text{ and } (\boldsymbol{y}, \boldsymbol{x}) \in E(G_1^t) \cap E(G_2^t)$$

that means (it is equivalent to)

$$A(\boldsymbol{x}) \not\subseteq A(\boldsymbol{y}) \cup B(\boldsymbol{y}) \text{ and } B(\boldsymbol{x}) \not\subseteq A(\boldsymbol{y}) \cup B(\boldsymbol{y}),$$

while we also have

$$A(\boldsymbol{y}) \not\subseteq A(\boldsymbol{x}) \cup B(\boldsymbol{x}) \text{ and } B(\boldsymbol{y}) \not\subseteq A(\boldsymbol{x}) \cup B(\boldsymbol{x}).$$

So if a set of sequences in $V^t$ forms a symmetric clique in $G_1^t \cap G_2^t$, then the corresponding system of $(A(\boldsymbol{x}) \cup B(\boldsymbol{x}))$ sets form a Sperner system in which each set on the system has a 2-partition (into the sets $A$ and $B$) such that even the two partition classes are not contained in the other sets. The corresponding $Sp(\mathcal{G})$ value therefore tells us the asymptotic growth of the largest possible such set system as $t$ goes to infinity. This is a very natural extremal set theory question and our formula for $C(\mathcal{G})$ will give a quantitative answer for it since the corresponding $Sp(G_1, P)$ and $Sp(G_2, P)$ values will be possible to calculate exactly.

Here is another example: $\mathcal{G} := \{G_1, G_2\}$ with $V = V(G_1) = V(G_2) = \{0, 1, 2\}$ again, but now we let

$$E(G_1) := \{(0, 1), (0, 2)\} \quad E(G_2) = \{(1, 2)\}.$$

Translating this similarly to extremal set theory as in the previous example we obtain that $Sp(\mathcal{G})$ will measure the asymptotic growth of the size of a largest possible set of pairs

$$\{(A_i, B_i) : A_i, B_i \subseteq [t], A_i \cap B_i = \emptyset\}$$

with the additional property that

$$\forall i \neq j\colon\ A_i \cup B_i \not\subseteq A_j \cup B_j \text{ and } A_i \cap B_j \neq \emptyset$$

and similarly for the role of $i$ and $j$ exchanged.

Our third example has the special feature that it describes an extremal set theoretic problem which was asked independently many years before Sperner capacity was even defined. Let now $k$ be a fixed positive integer and $\mathcal{G} = \{G_1, \ldots, G_{\binom{k}{2}}\}$ with $V = V(G_1) = \ldots = V(G_{\binom{k}{2}}) = [k]$ and the graphs $G_1, \ldots, G_{\binom{k}{2}}$ each contain a unique edge with an arbitrary orientation such that the union of these graphs is a tournament on $V = [k]$. (Thus the family contains all edges of the complete graph $K_n$ with one of its two possible orientations.) When translating to extremal set theory we find that each $\boldsymbol{x} \in V^t$ can be considered as (the characteristic vector of) a $k$-partition $(U_1, \ldots, U_k)$ of $[t]$ and two such $k$-partitions $(A_1, \ldots, A_k)$ and $(B_1, \ldots, B_k)$ are in the required relationship if and only if

$$\forall i \neq j \quad A_i \cap B_j \neq \emptyset.$$

This is almost the same as the following notion that Rényi defined in 1970.

**Definition 36** *Two $k$-partitions of the same set of $t$ elements are called* qualitatively independent *if every partition class of one of them intersects every partition class of the other.*

It should be clear from the definitions that this is almost the same as our conditions above except that for qualitative independence we would also need that $A_i \cap B_i \neq \emptyset$, that is the partition classes with equal indices should behave the same way as those with unequal indices. However, it is easy to observe that this difference is not essential, namely, it does not effect the asymptotic growth rate of the cardinality of the largest such set of partitions. This is because if we once had a subset of $V^t$ forming a symmetric clique in $\cap_{i=1}^{\binom{k}{2}} G_i^t$ then adding just $k$ coordinates at the end of each of the sequences forming such a symmetric clique such that the $r^{\text{th}}$ added coordinate takes the value $r$ for each of our sequences, we get characteristic vectors of qualitatively independent sets without changing the asymptotic growth rate. Rényi asked the maximum number of pairwise qualitatively independent $k$-partitions on a given finite set and so our technique can also handle this problem.

*Twenty fourth lecture* (December 6, 2022)

## 11.1 The Gargano-Körner-Vaccaro theorem

As we have already mentioned the inequality in Proposition 14 and in fact, more generally, the one in Proposition 15 holds with equality.

**Theorem 44** (Gargano-Körner-Vaccaro [21, 22])

$$Sp(\mathcal{G}) = \max_P \min_{G_i \in \mathcal{G}} Sp(G_i, P).$$

Before telling anything about the proof, let us see how this theorem can be used to obtain a quantitative answer for those extremal set theoretic questions that served as our examples. This needs to see how the $Sp(G, P)$ values can be calculated in some simple situations.

*Rule 1: Transitive tournaments.* Let $T_n$ be the transitive tournament on $n$ vertices and $P$ be a probability distribution on its vertex set. Earlier we have seen in the proof of Proposition **??** that if we consider the set $Q_b$ of all vertices $x_1 x_2 \dots x_t \in V(T_n^t)$ for which $\sum_{i=1}^t x_i = b$ for some constant number $b$ then those induce a symmetric clique in $T_n^t$. This condition is clearly satisfied by sequences that all have the same type, in particular, by all sequences having some type $P_t$ on length $t$ such that $\lim_{t\to\infty} P_t = P$. We have also seen in Lemma **??** that $\lim_{k\to\infty} \frac{1}{kt} \log |T_P^{kt}| = H(P)$ if $P$ is a possible type of sequences on length $t$. Applying this on the types $P_t$ and using the continuity of the entropy function we obtain that

$$Sp(T_n, P) = 2^{H(P)}.$$

*Rule 2: Twin vertices.* We call two vertices $a, b$ of a digraph $D$ twins if

$$\forall z \in V(D)\text{:}\ \ (a, z) \in E(D) \Leftrightarrow (b, z) \in E(D) \text{ and } (z, a) \in E(D) \Leftrightarrow (z, b) \in E(D).$$

Let $a, b$ be twin vertices in a digraph $D$, $P$ a probability distribution on $V(D)$, $D_{ab}$ be the digraph obtained from $D$ by identifying the two twin vertices $a$ and

$b$ to a vertex $v_{ab}$ and $P_{ab}$ be the probability distribution on $V(D_{ab})$ defined by $P_{ab}(v_{ab}) = P(a) + P(b)$ and $P_{ab}(u) = P(u)$ if $u \neq a, b$. Then

$$Sp(D, P) = Sp(D_{ab}, P_{ab}).$$

This statement is more or less obvious as substituting every appearance of $a$ and $b$ in sequences giving a symmetric clique in $D^t$ with $v_{ab}$ will clearly give a symmetric clique in $D^t_{ab}$ and if $a$ appeared $tP(a)$ times and $b$ appeared $tP(b)$ times then $V_{ab}$ will appear $t(P(a) + P(B))$ times giving $Sp(D_{ab}, P_{ab}) \geq Sp(D, P)$, while the reverse inequality is equally easy to obtain. Indeed, in a symmetric clique of $D^t_{ab}$ where $v_{ab}$ appears $t(P(a) + P(b))$ times in every sequence we can substitute $tP(a)$ of those appearances by $a$ and $tP(b)$ of those by $b$ to get a symmetric clique of the same size in $D^t$ proving $Sp(D, P) \geq Sp(D_{ab}, P_{ab})$.

*Rule 3: Isolated points.* Let $D$ be a digraph that consists of an isolated point $w$ and the digraph $D_0$ on the vertices in $V(D) \setminus \{w\}$. Let $P$ be a probability distribution on $V(D)$ and $P(D_0) = \sum_{v \in V(D_0)} P(v)$. Denote by $P_0$ the normalized distribution on $V(D_0)$ defined by $P_0(v) = \frac{P(v)}{P(D_0)}$. Then

$$Sp(D, P) = [Sp(D_0, P_0)]^{P(D_0)},$$

or in the logarithmic form (which is perhaps easier to remember)

$$\log Sp(D, P) = P(D_0) \log Sp(D_0, P_0).$$

To prove the validity of this rule we show separately the inequalities

$$\log Sp(D, P) \geq P(D_0) \log Sp(D_0, P_0)$$

and

$$\log Sp(D, P) \leq P(D_0) \log Sp(D_0, P_0).$$

The first one is easy: consider a largest symmetric clique in $D^t_0$ consisting of sequences of type (close to) $P_0$ and add $(1 - P(D_0))t = tP(w)$ $w's$ at the end of each of these sequences. This will clearly give a symmetric clique of the same size and the right type.

To prove the reverse inequality consider a largest symmetric clique consisting of sequences of the same type (close to) $P$ in $D^t$. If $t$ is large enough then its size is about $Sp(D, P)^t$. Each of these sequences contains (about) $tP(w)$ copies of the vertex $w$. For each sequence substitute the subsequence of $tP(w)$ $w$'s by the (more or less) $[Sp(D_0, P_0)]^{tP(w)}$ distinct sequences of a largest symmetric clique (consisting of sequences) of type $P_0$ to obtain $[Sp(D_0, P_0)]^{tP(w)}$ new sequences from each original sequence. This way we get a symmetric clique of size (about)

$$2^{t(\log Sp(D,P) + (1 - P(w)) \log Sp(D_0, P_0))}$$

proving that

$$t \log Sp(D, P) + tP(w) \log Sp(D_0, P_0) \leq t \log Sp(D_0, P_0),$$

that is,

$$\log Sp(D, P) \leq (1 - P(w)) \log Sp(D_0, P_0) = P(D_0) \log Sp(D_0, P_0).$$

Applying our rules to the three examples we have shown above we get the following results.

Example 1: Recall that we gave there $\mathcal{G} = \{G_1, G_2\}$ where $V = V(G_1) = V(G_2) = \{0, 1, 2\}$ and both $G_1$ and $G_2$ contains only one edge but a different one:

$$E(G_1) = \{(0, 1)\}, \quad E(G_2) = \{(0, 2)\}.$$

It is clear by symmetry that for the distribution $P$ achieving the maximum in Theorem 44 we must have $P(1) = P(2) =: p$. Then by Rules 1 nd 3 we have $Sp(G_1, P) = Sp(G_2, P) = (1-p)h(\frac{p}{1-p})$, so

$$\log Sp(\mathcal{G}) = \max_p (1-p)h(\frac{p}{1-p}).$$

Observing that $\max_p (1-p)h(\frac{p}{1-p})$ is also the asymptotic exponent of the number of all sequences in $\{0,1\}^t$ that has the property of not containing two consecutive 1's and knowing that the number of such sequences satisfies the recursion of the Fibonacci series (in fact, their number is given by the Fibonacci numbers), we obtain that

$$Sp(\mathcal{G}) = \frac{1+\sqrt{5}}{2}.$$

Example 2: In our second example we have $\mathcal{G} := \{G_1, G_2\}$ with $V = V(G_1) = V(G_2) = \{0,1,2\}$ again, but now we let

$$E(G_1) := \{(0,1),(0,2)\} \quad E(G_2) = \{(1,2)\}.$$

Symmetry gives again that the optimal distribution $P$ should have $P(1) = P(2)$, call this common value now $q/2$. Then by rules 1 and 2 we have $Sp(G_1, P) = h(q)$ while Rules 1,2 and 3 give $Sp(G_2, P) = q$. This means that here we have

$$\log Sp(\mathcal{G}) = \max_q \min\{q, h(q)\}$$

and this maximum is achieved by the unique $q$ for which we have $q = h(q)$. (This is because the function $f(x) = x$ is monotone increasing, while $h(x)$ is monotone decreasing for $x > 1/2$.)

Example 3: Recall that this is essentially Rényi's qualitative independence problem where the family $\mathcal{G}$ consisted all single edge subgraphs of a(n arbitrary) tournament on $k$ vertices. The calculations are simplest here: symmetry tells us that all vertices should have the same probability $\frac{1}{k}$ resulting in

$$\log Sp(\mathcal{G}) = \frac{2}{k}$$

by Rules 1 and 3.

*Twenty fifth lecture* (December 9, 2022)

Now we give an example of a family $\mathcal{G}$ of very simple (undirected) graphs for which it is already nontrivial to show the result the Gargano-Körner-Vaccaro (from now on GKV) theorem gives.

Let $\mathcal{G} = \{G_1, G_2\}$, where $V(G_1) = V(G_2) = V = \{00, 01, 10, 11\}$ and

$$E(G_1) = \{\{00,10\},\{01,11\}\}$$

$$E(G_2) = \{\{00,01\},\{10,11\}\}.$$

The formula in the GKV theorem is easy to evaluate and it gives $C_{\mathrm{OR}}(\mathcal{G}) = 2$ (the maximizing distribution is the uniform one as it easily follows from symmetry). Can we prove this directly without using the GKV theorem? Note that the upper bound is the easy one, so the problem is to give a construction that achieves this. We show such a construction that is simple but not completely trivial and the main idea of which is similar to that we will see in the general construction proving the theorem.

Consider all $t$-length sequences in $\{0,1\}^t$ that starts with a 0, there are $2^{t-1}$ such sequences. Take them as the sequence of first coordinates in $V^t$. That is we are constructing a set of sequences $\boldsymbol{z} \in V^t$ where $\boldsymbol{z} = z_1 z_2 \dots z_t$, $z_i = x_i y_i \in \{00, 01, 10, 11\}$, and so far we took all those sequences for which $x_1 = 0$. Now let $y_i = x_{i-1}$ for each $i \geq 2$. Finally look at all our $2^{t-1}$ sequences and partition them into two sets according to whether $y_t = 0$ or $y_t = 1$. We claim that both of these sets of sequences satisfy the requirements and since at least one of them has size at least $2^{t-2}$ this shows that

$$C(\mathcal{G}) \geq \lim_{t\to\infty} \sqrt[t]{2^{t-2}} = 2$$

that we need. To show that our set of sequences (any one of the two sets defined) forms a clique in both $G_1^t$ and $G_2^t$ consider two sequences in the set $\boldsymbol{z}$ and $\boldsymbol{z}'$, where $\forall i : z_i = x_i y_i, z_i' = x_i' y_i'$. Let $j$ be the first coordinate where $x_j \neq x_j'$. By the construction $j \geq 2$ and thus $y_j = x_{j-1} = x_{j-1}' = y_j'$, therefore $\{z_j, z_j'\} \in E(G_1)$ proving $\{\boldsymbol{z}, \boldsymbol{z}'\} \in E(G_1^t)$. Proving $\{\boldsymbol{z}, \boldsymbol{z}'\} \in E(G_2^t)$ is similar just now we look at the last coordinate $k$ for which $y_k \neq y_k'$. By the construction we have $k < t$, so we have $x_k = y_{k+1} = y_{k+1}' = x_k'$ and therefore $\{z_k, z_k'\} \in E(G_2)$ which implies $\{\boldsymbol{z}, \boldsymbol{z}'\} \in E(G_2^t)$ that we needed.

The general proof of the GKV theorem is based on the following.

**Key Lemma.** If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two families of graphs with all their members having the same vertex set $V$ and $P$ is a probability distribution on $V$ then

$$Sp(\mathcal{G}_1 \cup \mathcal{G}_2, P) = \min\{Sp(\mathcal{G}_1, P), Sp(\mathcal{G}_2, P)\}.$$

From this the theorem follows immediately by induction on the number of elements of the family $\mathcal{G}$ for which we want to prove the statement. (For $\mathcal{G} = \{G_1, \dots, G_k\}$ we assume that the statement is already proven for $\mathcal{G}_1 := \{G_1, \dots, G_{k-1}\}$, set $\mathcal{G}_2 := \{G_k\}$ and apply the Key Lemma. The base case for $k = 1$ is trivial.)

Here we sketch the proof of the statement of the Key Lemma[1].

Assume w.l.o.g. that $\Sigma_1 := Sp(\mathcal{G}_1, P) \leq \Sigma_2 := Sp(\mathcal{G}_2, P))$. We fix a large enough $t$. We know that there exists a symmetric clique in $\cap_{G\in\mathcal{G}_1} G^t$ consisting of approximately $\Sigma_1^t$ sequences of some type $P_t$ very close to $P$ and there exists a symmetric clique in $\cap_{G\in\mathcal{G}_2} G^t$ consisting of approximately $\Sigma_2^t$ sequences of some type $P_t'$ close to $P$.

SIMPLIFYING ASSUMPTION: $P_t = P_t'$ (close types can be made equal by adding not too many coordinates).

Let $G_1$ be the subgraph of $\cap_{G\in\mathcal{G}_1} G^t$ induced by the sequences of type $P_t$. Notice that $G_1$ is vertex-transitive as a permutation of the indices can bring any sequence of type $P_t$ to any other. Let $\bar{G}_1$ be the undirected complement of $G_1$, i.e., we connect two sequences if at least one of the two directed edges connecting them is missing in $G_1$. We have $\alpha(\bar{G}_1) = \omega_s(G_1) \approx \Sigma_1^t$. As $\bar{G}_1$ is also vertex-transitive we have $\chi_f(\bar{G}_1) = |V(G_1)|/\alpha(\bar{G}_1) \approx |V(G_1)|/\Sigma_1^t$. By Lovász's theorem we have $\chi(\bar{G}_1) \leq \chi_f(\bar{G}_1)(\ln\alpha(\bar{G}_1) + 1) \approx |V(G_1)|/\Sigma_1^t$ (logarithmic terms can safely be ignored). A proper coloring of $\bar{G}_1$ by $\chi(\bar{G}_1)$ colors partitions $G_1$ into symmetric cliques. The average size of a part is approximately $\Sigma_1^t$.

SIMPLIFYING ASSUMPTION: We can choose a common divisor of $A$ of the sizes of all parts with $A$ not much smaller than $\Sigma_1^t$. (This may need to throw away a few elements but that will not change the exponential growth rate.) We refine the partition into parts $D_1, \dots, D_r$ of size exactly $A$.

[1]This sketch of proof is written with the kind help of Gábor Tardos.

We partition the subgraph $G_2$ of $\cap_{G\in\mathcal{G}_2} G^t$ spanned by the vertices of type $P_t$ the same way. We obtain a partition to symmetric cliques of average size approximately $\Sigma_2^t$.

SIMPLIFYING ASSUMPTION: We assume that $A$ divides these sizes, too. We refine the partition till all parts are of size exactly $A$ and denote the parts by $F_1, \ldots, F_r$.

Let $m$ be large and consider the sequences in $(V(G_1))^m$. These are concatenations of $m$ length $t$-sequences of type $P_t$, so altogether they are length $mt$ sequences of type $P_t$. Consider the set $S$ of all the sequences $(x_1, \ldots, x_m)$, where $x_1 \in D_1$ and $x_{i+1} \in D_j$ whenever $x_i \in F_j$ $(1 \le i < m,\ 1 \le j \le r)$. Clearly, we have $A$ choices for $x_1$, and given $x_1$, we have $A$ choices for $x_2$, etc. Thus $|S| = A^m$.

Claim: The sequences in $S$ form a symmetric clique in $G^{mt}$ for $G \in \mathcal{G}_1$. Indeed, let $(x_1, \ldots, x_m)$ and $(y_1, \ldots, y_m)$ be two distinct elements of $S$ and let $i$ be the smallest index for which $x_i \ne y_i$. Then $x_i$ and $y_i$ fall into the same part $D_j$ (the one with $x_{i-1} \in F_j$ or $D_1$ if $i = 1$), so we have the necessary relation somewhere in that segment.

The same claim does not necessarily hold for $G \in \mathcal{G}_2$. To enforce it we partition $S$ according to which part in the $F_j$-partition contains the last segment and consider the largest part $S'$. Clearly $|S'| \ge |S|/r = A^m/r = \Sigma_1^{tm}/r$. If $m$ is large enough the effect of dividing by $r$ is small and we get rates close $\Sigma_1$.

Claim: The sequences in $S'$ form a symmetric clique in $G^{mt}$ for $G \in \mathcal{G}_2$. The proof is identical, but this time we look for the last difference $x_i \ne y_i$ between sequences $(x_1, \ldots, x_m)$ and $(y_1, \ldots, y_m)$ of $S'$. The vertices $x_i$ and $y_i$ are now in the same part of the $F_j$-partition and we are done.

The two claims and our estimate on $S'$ proves the statement and the Gargano-Körner-Vaccaro Theorem (modulo the simplifying assumptions). □

# References


[1] N. Alon, On the capacity of digraphs, *European J. Combin.*, 19 (1998), 1–5.

[2] N. Alon, *Lovász, vectors, graphs and codes*, In: Bárány, I., Katona G., Sali A. (eds) Building Bridges II. Bolyai Society Mathematical Studies, vol 28 (2020). Springer, Berlin, Heidelberg.

[3] Noga Alon, List coloring, lecture at
`https://www.youtube.com/watch?v=IAgHA4TTxy0`

[4] N. Alon and A. Orlitsky, Repeated communication and Ramsey graphs, *IEEE Trans. Inform. Theory*, **41** (1995), 1276–1289.

[5] Claude Berge, Motivations and history of some of my conjectures, *Discrete Mathematics* Volumes 165166 (1997), Pages 61-70.

[6] A. Blokhuis, On the Sperner capacity of the cyclic triangle, J. Algebraic Combin., **2** (1993), 123-124.

[7] A. R. Calderbank, P. Frankl, R. L. Graham, W. Li, and L. Shepp, The Sperner capacity of the cyclic triangle for linear and non-linear codes, *J. Algebraic Combin.*, **2** (1993), 31–48.

[8] K. Cameron, J. Edmonds, L. Lovász, A note on perfect graphs, *Periodica Math. Hungar.* **17** (1986), 173–175.

[9] V. Chvátal, On certain polytopes associated with graphs, *J. Combin. Theory ser. B*, **18** (1975), 138–154.

[10] G. Cohen, J. Körner, G. Simonyi, Zero-error capacities and very different sequences, in: Sequences. Combinatorics, Security and Transmission, papers presented at the Advanced International Workshop on Sequences, Positano, Italy, June 1988, Springer, New York, 1990, R. M. Capocelli (ed.), 144-155.

[11] Imre Csiszár, János Körner, *Information theory. Coding theorems for discrete memoryless systems. Second edition*, Cambridge University Press, Cambridge, 2011.

[12] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, Entropy splitting for antiblocking corners and perfect graphs, *Combinatorica*, **10** (1990), 27–40.

[13] E. V. Denardo, A. J. Hoffman, T. Mackenzie, and W. R. Pulleyblank: A non-linear allocation problem, *IBM J. Res.*, **38** (1994), 301–306.

[14] Reinhard Diestel, *Graph Theory, Fifth edition*, Graduate Texts in Mathematics, 173. Springer, Berlin, 2017.

[15] P. Erdős, Z. Füredi, A. Hajnal, P. Komjáth, V. Rödl, Á. Seress, Coloring graphs with locally few colors, *Discrete Math.*, **59** (1986), 21–34.

[16] P. Erdős, Chao Ko, R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser.*, 12 (1961), 313–318.

[17] P. Erdős, R. J. McEliece, H. Taylor, Ramsey bounds for graph products, *Pacific J. Math.*, **37** (1971), 45–46.

[18] D. R. Fulkerson, Blocking and antiblocking pairs of polyhedra, *Math. Programming*, **1** (1971), 168–194.

[19] D. R. Fulkerson, On the perfect graph theorem, in: *Mathematical Programming* (T. C. Hu and S. M. Robinson eds.), Academic Press, New York, 1973, 69–77.

[20] A. Galluccio, L. Gargano, J. Körner, G. Simonyi, Different capacities of a digraph, *Graphs Combin.*, **10** (1994), 105–121.

[21] L. Gargano, J. Körner, U. Vaccaro, Sperner capacities, *Graphs Combin.*, **9** (1993), 31–46

[22] L. Gargano, J. Körner, U. Vaccaro, Capacities: from information theory to extremal set theory, *J. Combin. Theory Ser. A*, **68** (1994), 296–316.

[23] G. S. Gasparian, Minimal imperfect graphs: a simple approach, *Combinatorice*, **16** (2) (1996), 209–212.

[24] R. A. Gibbs, Self-complementary graphs, *J. Combin. Theory Ser. B*, **16** (1974), 106–123.

[25] A. Gyárfás, Transitive tournaments and self-complementary graphs, *J. Graph Theory*, **38** (2001), 111–112.

[26] J. Kahn and J. H. Kim, Entropy and sorting, *Proc. 24th Annual ACM Symposium on the Theory of Computing*, 178–187.

[27] Karger, R. Motwani and M. Sudan: Approximate graph coloring by semidefinite programming, *Proc. 35th Ann. IEEE Symp. on Found. Comp. Science* (1994), 213.

[28] M. Kneser, Aufgabe 300, Jahresber. Deutsch., Math. Verein. 58 (1955) 27.

[29] D. Knuth, The Sandwhich Theorem, *Electronic J. Combin*, 1 (1994), #A1.

[30] J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs. *Transactions of the 6th Prague conference on Information Theory, etc., 1971*, Academia, Prague, (1973), 411–425.

[31] J. Körner, C. Pilotto, G. Simonyi, Local chromatic number and Sperner capacity, *J. Combin Theory Ser. B*, **95** (2005), 101-117.

[32] J. Körner, G. Simonyi, and Zs. Tuza, Perfect couples of graphs, *Combinatorica*, **12** (1992), 179–192.

[33] R. E. Krichevskii, Complexity of contact circuits realizing a function of logical algebra, *Sov. Phys. Dokl.* **8** (1964), 770–772.

[34] M. Larsen, J. Propp, D. Ullman, The fractional chromatic number of Mycielski's graphs, *J. Graph Theory*, **19** (1995), no. 3, 411–416.

[35] L. Lovász, Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972), 253–267.

[36] L. Lovász, A characterization of perfect graphs, *J. Combin. Theory ser. B*, **13** (1972), 95–98.

[37] L. Lovász, On the ratio of optimal integral and fractional covers, *Discrete Math.* 13, Issue 4, 1975, 383–390.

[38] L. Lovász, Kneser's conjecture, chromatic number, and homotopy, *J. Combin. Theory Ser. A*, **25** (1978), no. 3, 319–324.

[39] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **25** (1979), 1–7.

[40] L. Lovász, *An Algorithmic Theory of Numbers, Graphs, and Convexity*, CBMS-NSF Regional Conference Series in Applied Mathematics 50, SIAM, Philadelphia, Pennsylvania 1986.

[41] L. Lovász, *Combinatorial Problems and Exercises, 2nd Edition*, Akadémiai Kiadó - North Holland, 1993.

[42] László Lovász, *Graphs and Geometry*, AMS, 2019.

[43] L. Lovász, M.D. Plummer: Matching Theory, Akadémiai Kiadó - North Holland, Budapest, 1986 (reprinted by AMS Chelsea Publishing, 2009).

[44] J. Matoušek, *Using the Borsuk-Ulam theorem.* Lectures on topological methods in combinatorics and geometry. Written in cooperation with A. Björner and G. M. Ziegler, Universitext, Springer-Verlag, Berlin, 2003.

[45] R. J. McEliece, E. C. Posner, Hide and seek, data storage, and entropy, *Ann. Math. Statist.*, **42** (1971), 1706–1716.

[46] J. Nayak and K. Rose, Graph capacities and zero-error transmission over compound channels, *IEEE Trans. Inform. Theory*, **51** (2005), 4374–4378.

[47] J. Nešetřil, M. Rosenfeld: I. Schur, C. E. Shannon and Ramsey numbers, a short story; Combinatorics, graph theory, algorithms and applications, *Discrete Math.*, **229** (2001), no. 1-3, 185–195.

[48] Ilan Newman and Avi Wigderson, Lower Bounds on Formula Size of Boolean Functions Using Hypergraph Entropy, SIAM Journal on Discrete Mathematics, Vol. 8 (1995), Issue 4.

[49] I. Newman, P. Ragde, and A. Wigderson, Perfect hashing, graph entropy and circuit complexity, *Proceedings of the fifth annual conference on structure in Complexity Theory*, 1990, 91–100.

[50] G. Ringel, Selbstkomplementäre Graphen, *Arch. Math.*, **14** (1963), 354-358.

[51] H. Sachs, Über selbstkomplementäre Graphen, *Publ. Math. Debrecen*, **9** (1962), 270–288.

[52] A. Sali, G. Simonyi, Orientations of self-complementary graphs and the relation of Sperner and Shannon capacities, *European J. Combin.*, **20** (1999), 93–99.

[53] Edward R. Scheinerman, Daniel H. Ullman, *Fractional Graph Theory*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley and Sons, Chichester, 1997.

[54] C. E. Shannon, The zero error capacity of a noisy channel, *IRE Trans. Inform. Theory*, 2 (1956), 8–19.

[55] G. Simonyi, Graph entropy: a survey, in *Combinatorial Optimization*, DIMACS Series on Discrete Math. and Computer Science Vol. 20 (W. Cook, L. Lovász, P. D. Seymour eds.), AMS, 1995, 399–441.

[56] H. S. Witsenhausen, The zero-error side-information problem and chromatic numbers, *IEEE Trans. Inform. Theory*, **22** (1976).