A hypergraph or set system is a pair \((V, E)\) where \(E\) is a family of subsets of \(V\) called edges. We typically identify a hypergraph with its edge set. A matching in a hypergraph is a set of disjoint edges. The theory of matchings in graphs is well-developed, but unfortunately for \(r\)-uniform hypergraphs there is no necessary and sufficient condition for a perfect matching. It is not hard to show that every \(r\)-regular graph has a matching covering all but \(O(n/r)\) vertices (in fact using the Tutte-Berge formula, a tight result on the number of uncovered vertices can be obtained). We consider approaches which do not appeal to matching theory.

1 Large matchings in regular graphs

The permanent of an \(n\) by \(n\) matrix \(A\) is defined to be the signless determinant

\[
\text{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i, \sigma(i)}.
\]

If \(A\) is the adjacency matrix of a graph, then \(\text{per}(A)\) is interpreted combinatorially as \(\sum_{F \subseteq G} 2^c(F)\) where \(F\) is a subgraph of \(G\) consisting of vertex-disjoint cycles and edges, and \(c(F)\) is the number of components of \(F\). If \(A\) is the incidence matrix of a bipartite graph, then \(\text{per}(A)\) exactly counts the number of perfect matchings of the graph. Arguably the most famous result on permanents is the Brégman-Falikman-Egorichev solution to the Minc (upper bound) and van der Waerden (lower bound) conjectures.

**Theorem 1.** Let \(A\) be an \(n\) by \(n\) zero-one matrix with row sums \(r_1, r_2, \ldots, r_n \geq 1\). Then

\[
\text{per}(A) \leq \prod_{i=1}^{n} r_i!^{1/r_i}.
\]

On the other hand, if \(r_1 = r_2 = \cdots = r_n = r\), then

\[
\text{per}(A) \geq \frac{r^n n!}{n^n}.
\]

A short proof of the upper bound was given by Schrijver, with a generalization to arbitrary non-negative matrices. Using this theorem, we give the following theorem of Alon (2009):
**Theorem 2** (Alon). If $G$ is an $n$-vertex $d$-regular graph, then there exists a subgraph $F$ of $G$ consisting of a vertex-disjoint union of cycles of length at least $g$, where $4g^2e^g + 1 \geq d$, such that $|V(F)| \geq n - n/g$. In particular, $G$ has a matching covering $n - O(n/\sqrt{\log d})$ vertices.

**Proof.** Let $s = n/g$. We use the combinatorial interpretation of the permanent, $\text{per}(A) = \sum_{F \subseteq G} 2^{\text{wt}(F)}$. Suppose for every $F$ in this sum, there are components $C_1, C_2, \ldots, C_t$ such that $|V(C_i)| < g$ and $s \leq \bigcup_{i=1}^t |V(C_i)| < s + g$. Note $2s \leq t(s-1)$. Then $H = G - \bigcup_{i=1}^t V(C_i)$ is a graph whose adjacency matrix $B$ satisfies

$$\text{per}(B) \leq d^{(n-s)/d} \leq d^{n-s}.$$  

Pick a representative $v_i \in V(C_i)$. There are at most $\binom{n}{t}$ ways to pick these vertices, and then at most $d^{s+g-t}$ ways to grow the components $C_1, C_2, \ldots, C_t$. We conclude

$$\text{per}(A) \leq \binom{n}{t} 2^t d^{s+g-t} \cdot \text{per}(B) \leq \binom{n}{t} 2^t d^{n+g-t}.$$ 

On the other hand, $\text{per}(A) \geq (d/e)^n$ using Stirling’s formula and the last theorem. Therefore

$$\binom{n}{t} 2^t d^{n+g-t} \geq \left(\frac{d}{e}\right)^n.$$ 

This implies $td/2e < e^{n/t}d^{n+t}/n$. Since $t \geq s/g \geq n/g^2$, we get $d < 4g^2e^{g+1}$ if $d$ is large enough, a contradiction. $\square$

It would be interesting to know if there is an analog of this theorem for $r$-uniform hypergraphs.

### 2 Large matchings in regular hypergraphs

The approach we use for large matchings in regular hypergraphs is probabilistic. Let $H$ be an $n$-vertex $d$-regular $r$-uniform hypergraph. In the case $r = 2$, we saw that a $d$-regular graph has a matching covering $n - O(n/d)$ vertices as $d \to \infty$. Unfortunately, this does not carry over to $r$-uniform hypergraphs when $r \geq 3$, as the following construction shows.

**Construction.** Let $r \geq 3$ and let $P(S)$ be a projective plane of order $r - 2$, which can be viewed as an $(r - 1)$-regular $(r - 1)$-uniform intersecting on a set $S$ of $r^2 - 3r + 3$ vertices. Then create an $r$-uniform $d$-regular hypergraph $H$ as follows: let $d = t(r-1) = s(r^2-3r+3)$, and let $V(H) = T \cup S_1 \cup \cdots \cup S_s$ where $|T| = t$ and $|S_i| = r^2 - 3r + 3$, and

$$H = \bigcup_{i=1}^s \{x \cup e : x \in T, e \in P(S_i)\}.$$ 

If $M$ is a maximum matching in $H$, then $M = \{i \cup e_i : e_i \in P(S_i) : 1 \leq i \leq s\}$ and this implies $|M| = s = n(r-1)/r(r^2-3r+3)$. Note this is roughly $n/r^2$. 

2
It is possible in an $d$-regular $r$-uniform $n$-vertex hypergraph to get a matching of size at least about $n/r^2$. Suppose we greedily select edges and delete all edges intersecting selected edges. This process deletes at most $(d-1)r + 1$ edges at a time, and since we have $|H| = dn/r$, we conclude that it is possible to pick $dn/((d-1)r + 1)r \approx n/r^2$ disjoint edges.

It turns out the appropriate condition to place on the hypergraph is that no pair of vertices is contained in many edges. The following result was proved by Rödl (1985), and the proof method is known as Rödl’s semirandom method.

**Theorem 3 (Rödl).** Let $r \geq 2$. For all $\epsilon > 0$ and $d \geq 1$, there exists $\delta > 0$ such that if $H$ is a $d$-regular $r$-uniform $n$-vertex hypergraph such that each pair of vertices is contained in at most $\delta d$ edges, then $H$ has a matching of size at least $n/r - \epsilon n$.

The original proof is probabilistic (a randomized greedy algorithm). The idea is that one selects edges independently with probability $p = \epsilon/d$, and then remove from the hypergraph all vertices covered by these edges. We then obtain a new hypergraph in which all but a few vertices have degree roughly $d' = e^{-\epsilon}d$, and repeat the argument, selecting edges with probability $\epsilon/d'$. The condition on pairs of vertices in the theorem is used to control the variance of the degrees of the vertices, and hence the number of vertices at each stage which do not have the right degree. In the end, one expects to select at most $\epsilon n/r \cdot (1 + e^{-\epsilon} + \ldots)$ edges covering almost all the vertices of the hypergraph. As $\epsilon \to 0$, we have selected about $(1 + \epsilon)n/r$ edges to cover the hypergraph, and from this we can easily extract a matching of size $n/r - \epsilon n$. The technical details, which are not trivial, are omitted, since we prove a stronger theorem for *uncrowded hypergraphs*: these are uniform hypergraphs containing no short cycles: i.e. not containing any of the configurations below:

![Hypergraph Configurations](image)

**Theorem 4.** Let $r \geq 2$ and let $H$ be an $n$-vertex $d$-regular uncrowded hypergraph. If $d$ is large enough, then $H$ has a matching of size at least $n/r - 2n/d^{1/(r-1)}$.

**Proof.** We first sketch the proof. Pick a set $E_1$ of edges of $H_0 = H$ independently with probability $p = 1/d$. Let $E'_1$ consist of all edges of $E_1$ intersecting no other edge of $E_1$. Then delete all vertices covered by $E'_1$ to get a hypergraph $H_1$. Each vertex of $H_1$ is said to survive. The survival probability is $$ (1 - p(1 - p)^{(r-1)(d-1)})^d n \approx \exp(-1/\exp(r-1))n := \omega n. $$

The latter comes from the fact that $H$ is uncrowded, so the event that edge $e$ on a surviving vertex is in $H_1$ is independent of other edges on $v$ being in $H_1$, and each edge on $v$ survives
The quantities are as follows: starting with $H$. It is convenient to let $d_1(v)$ be the degree of $v \in V(H)$ in the hypergraph $H_1$. Then the expected value of $d_1(v)$ is approximately $\omega^{-1}d$. By concentration of measure inequalities, there is strong reason to believe that the random variables $|E_1|, |V(H_1)|$ and $d_1(v)$ track near their expectations. For the purpose of this sketch, assume this is the case, so that $|V(H_1)| \approx \omega n$ and $d_1(v) \approx \omega^{-1}d = d_1$ for all $v \in V(H_1)$. Then we repeat the argument: we pick a set $E_2$ of edges of $H_1$ with probability $1/d_1$, let $E_2'$ consist of picked edges intersecting no others, and delete all vertices covered by $E_2'$. In general, at stage $t$, we suppose we have a hypergraph $H_t$ with $|V(H_t)| \approx \omega t n$ and the degrees $d_t(v) \approx d_t = \omega^{(r-1)} t d$ for $v \in V(H_t)$, and we pick a set $E_{t+1}$ of edges of $H_t$ with probability $p = 1/d_t$. If we could proceed until $d_t \leq 1$, the process has at least $t$ stages where $\omega^{(r-1)} t d \leq 1$. In particular, $t \geq \lfloor (\log d)/(r-1) \log(1/\omega) \rfloor$ stages. The number of uncovered vertices is then roughly $\omega t n \approx n/d^{1/(r-1)}$, and so we would have a matching of size at least $n/r - n/d^{1/(r-1)}$.

To make this sketch proof rigorous, we have to deal with (1) the concentration of the random variables (2) the growth of the error terms introduced by deviation terms in the concentration inequalities and (3) ensuring that all the random variables exhibit the desired behavior simulaneously. The latter will be achieved via the Lovász Local Lemma, and the former from the Hoeffding-Azuma concentration inequality.

**Lovász Local Lemma.** Let $A_1, A_2, \ldots, A_n$ be events in a probability space and suppose that for each $A_i$ there is a set $D_i \subset [n] \setminus i$ of size at most $\triangle$ such that $A_i$ is mutually independent with any set of events $A_j$ where $j \in D_i \setminus [n]$. If $\mathbb{P}(A_i) \leq 1/4\triangle$, then

$$
\mathbb{P}\left( \bigcap_{i=1}^{n} \overline{A_i} \right) \geq (1 - 1/\triangle)^n.
$$

**Hoeffding-Azuma Inequality.** Let $(X_i)_{i=1}^{m}$ be a sequence of independent random variables and let $f(X_1, X_2, \ldots, X_n)$ be a real valued $c$-Lipschitz function. Then for any $\lambda > 0$,

$$
\mathbb{P}(|f - \mathbb{E}(f)| > \lambda) \leq 2 \exp(-2\lambda^2/c^2m).
$$

We now proceed to the technical details. Throughout the process, $\omega = \exp(-1/\exp(r-1))$. The process is defined in the sketch above. If $d$ is sufficient large, we establish the existence of quantities $d_i, n_i, t$ and a real $\eta_i > 0$ such that with positive probability, for all $i \leq t$,

$$
||V(H_i)| - n_i| \leq \eta_i n_i \quad \text{and} \quad |d_i(v) - d_i| \leq \eta_i d_i. \quad (1)
$$

The quantities are as follows:

$$
n_i = \omega^i n \quad d_i = \omega^{(r-1)} d \quad t = \frac{\log d}{6(r-1) \log(1/\omega)} \quad \eta_i = 4^i d_i^{1/3}.
$$

It is convenient to let $d_i^+ = (1 + \eta_i) d_i$ and $d_i^- = (1 - \eta_i) d_i$. We proceed by induction on $i \geq 0$, starting with $H_0 = H$, $n_0 = n$, and $d_0 = d$. This choice of parameters ensures for $i \leq t$ that

$$
d_i \geq d_i \geq \omega^{(r-1)} d \geq d^{1/2} \quad \text{and} \quad \eta_i \leq d_i^{-1/3} \leq d^{-1/6}. \quad (2)
$$
Expected values. Suppose we have an $H_i$ with the above statistics, from which we select edges independently with probability $p = \epsilon/d_i$. The survival probability is in the range

\[
[(1 - p(1 - p)^{d_i^{-}})^{d_i^{+}}, (1 - p(1 - p)^{d_i^{+}})^{d_i^{-}}].
\]

This is easily contained in $[(1 - 2\eta_i)\omega, (1 + 2\eta_i)\omega]$ for large enough $d$ due to (2), and this is why $n_{t+1} = \omega n_i$. Similarly, the expected value of $d_{i+1}(v)$ for $v \in V(H_{i+1})$ is in the range $[(1 - 2\eta_i)d_{i+1}, (1 + 2\eta_i)d_{i+1}]$ for large enough $d$.

Concentration. The random variable $|V(H_{i+1})|$ is an $r$-Lipschitz function of $|H_i| \leq 2d_i n_i/r$ independent random variables. Using the Hoeffding-Azuma Inequality with Concentration. The random variable $−(1 − p(1 - p)^{d_i^{-}})^{d_i^{+}}\omega n_i$, we have the following for large enough $d$:

\[
\mathbb{P}(\mathbb{E}(|V(H_i)| − n_{i+1}) > \lambda) \leq 2 \exp(-2\lambda^2/(r^2|H_i|)) \leq \exp(-n_i/d_i^2)
\]

provided $d$ is large enough. Since $H$ is uncrowded, $d_{i+1}(v)$ is a sum of $d_i(v)$ independent indicator random variables: distinct edges $e$ and $f$ on $v$ are independently deleted according as they contain a vertex covered by the picked edges of $H_i$. By the Chernoff Bound (or Hoeffding-Azuma), if $d$ is large enough using (2):

\[
\mathbb{P}(|d_{i+1}(v) - d_{i+1}| > 2\eta_i d_{i+1}) \leq (4rd_i)^{-5}.
\]

The local lemma. We consider the events $B$ and $C_v$ whose probabilities we determined in (3) and (4). The event $C_v$ is mutually independent with any set of events $C_w : w \in W$ if for all $w \in W$, there are no edges $e_1, e_2, \ldots, e_5 \in H_i$ with $v \in e_1, w \in e_5$ and $e_i \cap e_{i+1} \neq \emptyset$ for $1 \leq i \leq 4$. A dependency graph for the events $C_v$ therefore has degree at most $\Delta = (2rd_i)^5$. By (4), the local lemma applies to the events $C_v : v \in V(H_i)$ and so with $\overline{C} = \bigcap_{v \in V(H_i)} \overline{C}_v$, we have

\[
P(\overline{C}) \geq (1 - 1/\Delta)^{|V(H_i)|} > 1 - \mathbb{P}(B)
\]

for large enough $d$, using (3). So with positive probability neither of $B$ and $C$ occur. This establishes (1) for $i \leq t$.

Process termination. A calculation shows $t \geq (\log d)/(6(r - 1) \log(1/\omega))$. Then the number of uncovered vertices is

\[
|V(H_t)| \leq (1 + \eta_t)n_t \leq 2n_t \leq 2\omega^{-t}n \leq 2n/d^{1/(r-1)}.
\]

This completes the proof. □

The proof was extend to the case of linear hypergraphs – hypergraphs with $|e \cap f| \leq 1$ for distinct edges $e$ and $f$ – by Alon, Kim and Spencer (1999):

**Theorem 5** (Alon-Kim-Spencer). Let $r \geq 4$ and let $H$ be an $n$-vertex $d$-regular linear hypergraph. Then $H$ has a matching of size at least $n/r - O(n/d^{1/(r-1)})$. 

5