Lecture 1
Tools

1 Preliminaries

1.1 Notation and terminology

Our graph theoretic notation and terminology are fairly standard. In particular, for a graph $G = (V, E)$ and disjoint vertex subsets $U, W \subset V$, we denote by $N_G(U)$ the external neighborhood of U in G: $N_G(U) = \{v \in V - U : v \text{ has a neighbor in } U\}$. The number of eddes of G spanned by U is denoted by $e_G(U)$, the number of edges of G between U and W is $e_G(U, W)$. When the graph G is clear from the context, we may omit G in the subscript in the above notation.

Path lengths are measured in edges.

When dealing with graphs on n vertices, we will customarily use N to denote the number of pairs of vertices in such graphs: $N = \binom{n}{2}$.

1.2 Asymptotic estimates

We will use the following, fairly standard and easily proven, estimates of binomial coefficients. Let $1 \leq x \leq k \leq n$ integers. Then

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k , \tag{1}$$

$$\frac{\binom{n-x}{k-x}}{\binom{n}{k}} \leq \left(\frac{k}{n}\right)^x , \tag{2}$$

$$\frac{\binom{n-x}{k}}{\binom{n}{k}} \leq e^{-\frac{kx}{n}} . \tag{3}$$

1.3 Chebyshev and Chernoff

Chebyshev's Inequality helps to show concentration of a random variable X around its expectation, based of the first two moments of X. It reads as follows: let X be a random variable with expectation μ and variance σ^2. Then for any $a > 0$,

$$Pr[|X - \mu| \geq a\sigma] \leq \frac{1}{a^2} .$$

The following are very standard bounds on the lower and the upper tails of the Binomial distribution due to Chernoff: If $X \sim Bin(n,p)$, then

- $\Pr\left(X < (1-a)np\right) < \exp\left(-\frac{a^2np}{2}\right)$ for every $a > 0$.
- $\Pr\left(X > (1+a)np\right) < \exp\left(-\frac{a^2np}{3}\right)$ for every $0 < a < 1$.

Another, trivial yet useful, bound is as follows: Let $X \sim Bin(n,p)$ and $k \in \mathbb{N}$. Then

$$\Pr(X \geq k) \leq \left(\frac{enp}{k}\right)^k.$$

Indeed, $\Pr(X \geq k) \leq \binom{n}{k}p^k \leq \left(\frac{enp}{k}\right)^k$.

1.4 Random graphs, asymptotic notation

As usually, $G(n,p)$ denotes the probability space of graphs with vertex set $\{1,\ldots,n\} = [n]$, where each pair of distinct elements of $[n]$ is an edge of $G \sim G(n,p)$ with probability p, independently of other pairs. For $0 \leq m \leq N$, $G(n,m)$ denotes the probability space of all graphs with vertex set $[n]$ and exactly m edges, where all such graphs are equiprobable: $Pr[G] = \binom{N}{m}^{-1}$. One can expect that the probability spaces $G(n,p)$ and $G(n,m)$ have many similar features, when the corresponding parameters are appropriately tuned: $m = Np$; accurate quantitative statements are available, see [2], [3]. This similarity frequently allows to prove a desired property for one of the probability spaces, and the then to transfer it to the other one.

We will also address the model $D(n,p)$ of directed random graphs, defined as follows: the vertex set is $[n]$, and each of the $n(n-1) = 2N$ ordered pairs $1 \leq i \neq j \leq n$ is a directed edge of $D(n,p)$ with probability p, independently from other pairs.

We say that an event $\mathcal{E}_n$ occurs with high probability, or **whp** for brevity, in the probability space $G(n,p)$ if $\lim_{n\to\infty} Pr[G \sim G(n,p) \in \mathcal{E}_n] = 1$. (Formally, one should rather talk about a sequence of events $\{\mathcal{E}_n\}_n$ and a sequence of probability spaces $\{G(n,p)\}_n$.) This notion is defined in other (sequences of) probability spaces in an obvious way.

Let $k \geq 2$ be an integer, and assume that $0 \leq p, p_1, \ldots, p_k \leq 1$ satisfy $(1-p) = \prod_{i=1}^k (1-p_i)$. Then the random graphs $G \sim G(n,p)$ and $G' = \bigcup_{i=1}^k G(n,p_i)$ have the exact same distribution. Indeed, it is obvious that each pair of vertices $1 \leq i < j \leq n$ is an edge in both graphs G, G' independently of other pairs. In G, this edge does not appear with probability $1-p$, and in order for it not appear in G', it should not appear in any of the random graphs $G(n,p_i)$ – which happens with probability $\prod_{i=1}^k (1-p_i) = 1-p$, the same probability as in G. This very useful trick is called *multiple exposure* as it allows to generate (to expose) a random graph $G \sim G(n,p)$ in stages, by generating $G(n,p_i)$ sequentially and taking their union. In case when the last probability p_k is much smaller than the rest, it is also called *sprinkling* – a typical scenario in this case is to expose the bulk of the random graph $G \sim G(n,p)$ first by generating the graphs $G(n,p_i)$, $i = 1,\ldots,k-1$, to come close to a target graph property P, and then to add few random edges from the last random graph $G(n,p_k)$ (to sprinkle these few edges) to finish off the job.

2 Depth First Search and its applications for finding long paths

The *Depth First Search* is a well known graph exploration algorithm, usually applied to discover connected components of an input graph. As it turns out, this algorithm is particularly suitable for finding long paths in graphs, and using it in the context of random graphs can really make wonders. We will see some of them later in this course.

Recall that the DFS (Depth First Search) is a graph search algorithm that visits all vertices of a (directed or undirected) graph. The algorithm receives as an input a graph $G = (V, E)$; It is also assumed that some order σ on the vertices of G is given, and the algorithm prioritizes vertices according to σ. The algorithm maintains three sets of vertices, letting S be the set of vertices whose exploration is complete, T be the set of unvisited vertices, and $U = V \setminus (S \cup T)$, where the vertices of U are kept in a stack (the last in, first out data structure). It initiliazes with $S = U = \emptyset$ and $T = V$, and runs till $U \cup T = \emptyset$. At each round of the algorithm, if the set U is non-empty, the algorithm queries T for neighbors of the last vertex v that has been added to U, scanning T according to σ. If v has a neighbor u in T, the algorithm deletes u from T and inserts it into U. If v does not have a neighbor in T, then v is popped out of U and is moved to S. If U is empty, the algorithm chooses the first vertex of T according to σ, deletes it from T and pushes it into U. In order to complete the exploration of the graph, whenever the sets U and T have both become empty (at this stage the connected component structure of G has already been revealed), we make the algorithm query all remaining pairs of vertices in $S = V$, not queried before. Figure 1 provides an illustration of applying the DFS algorithm.

Observe that the DFS algorithm starts revealing a connected component C of G at the moment the first vertex of C gets into (empty beforehand) U and completes discovering all of C when U becomes empty again. We call a period of time between two consecutive emptyings of U an *epoch*, each epoch corresponds to one connected component of G. During the execution of the DFS algorithm as depicted in Figure 1, there are two components in the graph G, and respectively there are two epochs – the first is Steps 1–6, and the second is Steps 7–16.

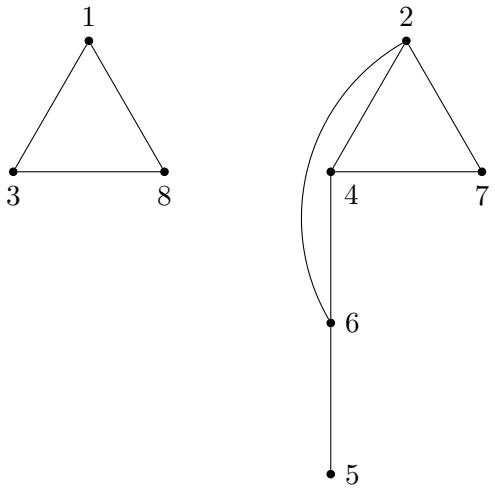
The following properties of the DFS algorithm are immediate to verify:

(P1) at each round of the algorithm one vertex moves, either from T to U, or from U to S;

(P2) at any stage of the algorithm, it has been revealed already that the graph G has no edges between the current set S and the current set T;

(P3) the set U always spans a path (indeed, when a vertex u is added to U, it happens because u is a neighbor of the last vertex v in U; thus, u augments the path spanned by U, of which v is the last vertex).

We now exploit the properties of the DFS algorithms to derive the existence of long paths in expanding graphs.



Step	S	U	T
0	$\emptyset$	$\emptyset$	$\{1,\ldots,8\}$
1	$\emptyset$	1	$\{2,\ldots,8\}$
2	$\emptyset$	1, 3	$\{2,4,\ldots,8\}$
3	$\emptyset$	1, 3, 8	$\{2,4,\ldots,7\}$
4	$\{8\}$	1, 3	$\{2,4,\ldots,7\}$
5	$\{3,8\}$	1	$\{2,4,\ldots,7\}$
6	$\{1,3,8\}$	$\emptyset$	$\{2,4,\ldots,7\}$
7	$\{1,3,8\}$	2	$\{4,5,6,7\}$
8	$\{1,3,8\}$	2, 4	$\{5,6,7\}$
9	$\{1,3,8\}$	2, 4, 6	$\{5,7\}$
10	$\{1,3,8\}$	2, 4, 6, 5	$\{7\}$
11	$\{1,3,5,8\}$	2, 4, 6	$\{7\}$
12	$\{1,3,5,6,8\}$	2, 4	$\{7\}$
13	$\{1,3,5,6,8\}$	2, 4, 7	$\emptyset$
14	$\{1,3,5,6,7,8\}$	2, 4	$\emptyset$
15	$\{1,3,4,5,6,7,8\}$	2	$\emptyset$
16	$\{1,\ldots,8\}$	$\emptyset$	$\emptyset$

Figure 1: Graph G with vertices labeled is on the left, and the protocol of applying the DFS algorithm to G is on the right. Observe that at any point of the algorithm execution the set U spans a path in G.

Proposition 2.1 *Let k, l be positive integers. Assume that $G = (V, E)$ is a graph on more than k vertices, in which every vertex subset S of size $|S| = k$ satisfies: $|N_G(S)| \geq l$. Then G contains a path of length $l - 1$.*

Proof Run the DFS algorithm on G, with σ being an arbitrary ordering of V. Look at the moment during the algorithm execution when the size of the set S of already processed vertices becomes exactly equal to k (there is such a moment as the vertices of G move into S one by one, till eventually all of them land there). By Property **(P2)** above, the current set S has no neighbors in the current set T, and thus $N(S) \subseteq U$, implying $|U| \geq l$. The set U always spans a path in G, by Property **(P3)**. Hence G contains a path of length $l - 1$. □

Proposition 2.2 *[1] Let $k < n$ be positive integers. Assume that $G = (V, E)$ is a graph on n vertices, containing an edge between any two disjoint subsets $S, T \subset V$ of size $|S| = |T| = k$. Then G contains a path of length $n - 2k + 1$ and a cycle of length at least $n - 4k + 4$.*

Proof Run the DFS algorithm on G, with σ being an arbitrary ordering of V. Consider the moment during the algorithm execution when $|S| = |T|$ – there is such a moment, by Property **(P1)**. Since G has no edges between the current S and the current T by Property **(P2)**, it follows by the proposition

assumption that both sets S and T are of size at most $k-1$. This leaves us with the set U whose size satisfies: $|U| \geq n - 2k + 2$. Since U always spans a path by **(P3)**, we obtain a path P of desired length. To argue about a cycle, take the first and the last k vertices of P. By the proposition assumption there is an edge between these two sets, this edge obviously closes a cycle with P, whose length is at least $n - 4k + 4$, as required. □

As we have mentioned already, the DFS algorithm is well suited to handle directed graphs too. Similar results to those stated above can be obtained for the directed case. Here is an analog of Proposition 2.2 for directed graphs; the proof is the same, mutatis mutandis.

Proposition 2.3 *[1] Let $k < n$ be positive integers. Let $G = (V, E)$ be a directed graph on n vertices, such that for any ordered pair of disjoint subsets subsets $S, T \subset V$ of size $|S| = |T| = k$, G has a directed edge from S to T. Then G contains a directed path of length $n - 2k + 1$ and a directed cycle of length at least $n - 4k + 4$.*

3 Pósa's Lemma and boosters

In this section we present yet another technique for showing the existence of long paths in graphs. This technique, introduced by Pósa in 1976 [4] in his research on Hamiltonicity of random graphs, is applicable not only for arguing about long paths, but also for various Hamiltonicity questions. And indeed, we will see its application in this context later.

In quite informal terms, Pósa's Lemma guarantees that expanding graphs not only have long paths, but also provide a very convenient backbone for augmenting a graph to a Hamiltonian one by adding new (random) edges. The fact that expanders are good for getting long paths is already not new to us – Propositions 2.1 and 2.2 are just about this. Pósa's Lemma however quantifies things somewhat differently and as a result yields further benefits.

We start by defining formally the notion of an expander.

Definition 3.1 *For a positive integer k and a positive real α, a graph $G = (V, E)$ is a (k, α)-expander if $|N_G(U)| \geq \alpha|U|$ for every subset $U \subset V$ of at most k vertices.*

By the way of example, Proposition 2.1 can now be rephrased (in a somewhat weaker form – we now require the expansion of all sets of size *up to* k) as follows: if G is a (k, α)-expander, then G has a path of length at least $\alpha k - 1$. For technical reasons, Pósa's Lemma uses the particular case $\alpha = 2$.

The idea behind Pósa's approach is fairly simple and natural – one can start with a (long) path P, and then using extra edges, perform a sequence of simple deformations (rotations) till it will be possible to close the deformed path to a cycle, or to extend it by appending a vertex outside $V(P)$; then this can be repeated if necessary to create an even longer path, or to close it to a Hamilton cycle. The approach is thus called naturally the *rotation-extension* technique. Formally, let $P = x_0x_1 \dots x_h$ be a path in a graph $G = (V, E)$, starting at a vertex x_0. Suppose G contains an edge (x_i, x_h) for

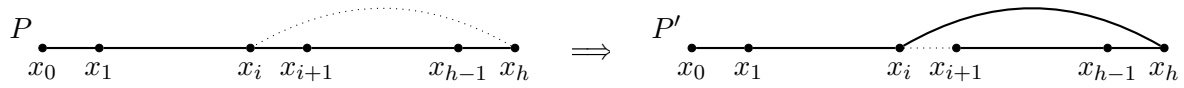


Figure 2: Elementary rotation is applied to a path P to get a new path P' with the same vertex set.

some $0 \leq i < h-1$. Then a new path P' can be obtained by rotating the path P at x_i, i.e. by adding the edge (x_i, x_h) and erasing (x_i, x_{i+1}). This operation is called an *elementary rotation* and it depicted in Figure 2. Note that the obtained path P' has the same length h and starts at x_0. We can therefore apply an elementary rotation to the newly obtained path P', resulting in a path P'' of length h, and so on. If after a number of rotations an endpoint x of the obtained path Q is connected by an edge to a vertex y outside Q, then Q can be extended by adding the edge (x, y).

The power of the rotation-extension technique of Pósa hinges on the following lemma.

Lemma 3.2 *[4] Let G be a graph, P be a longest path in G and $\mathcal{P}$ be the set of all paths obtainable from P by a sequence of elementary rotations. Denote by R the set of ends of paths in $\mathcal{P}$, and by R^- and R^+ the sets of vertices immediately preceding and following the vertices of R along P, respectively. Then $N_G(R) \subset R^- \cup R^+$.*

Proof Let $x \in R$ and $y \in V(G) \setminus (R \cup R^- \cup R^+)$, and consider a path $Q \in \mathcal{P}$ ending at x. If $y \in V(G) \setminus V(P)$, then $(x, y) \notin E(G)$, as otherwise the path Q can be extended by adding y, thus contradicting our assumption that P is a longest path. Suppose now that $y \in V(P) \setminus (R \cup R^- \cup R^+)$. Then y has the same neighbors in every path in $\mathcal{P}$, because an elementary rotation that removed one of its neighbors along P would, at the same time, put either this neighbor or y itself in R (in the former case $y \in R^- \cup R^+$). Then if x and y are adjacent, an elementary rotation applied to Q produces a path in $\mathcal{P}$ whose endpoint is a neighbor of y along P, a contradiction. Therefore in both cases x and y are non-adjacent. □

The following immediate consequence of Lemma 3.2 is frequently applied in Hamiltonicity problems in random graphs.

Corollary 3.3 *Let h, k be positive integers. Let $G = (V, E)$ be a graph such that its longest path has length h, but it contains no cycle of length $h+1$. Suppose furthermore that G is a $(k, 2)$-expander. Then there are at least $\frac{(k+1)^2}{2}$ non-edges in G such that if any of them is turned into an edge, then the new graph contains an $(h+1)$-cycle.*

Proof Let $P = x_0 x_1 \dots x_h$ be a longest path in G and let R, R^-, R^+ be as in Lemma 3.2. Notice that $|R^-| \leq |R|$ and $|R^+| \leq |R| - 1$, since $x_h \in R$ has no following vertex on P and thus does not contribute an element to R^+.

According to Lemma 3.2,

$$|N_G(R)| \leq |R^- \cup R^+| \leq 2|R| - 1 \ ,$$

and it follows that $|R| > k$. (Here the choice $\alpha = 2$ in the definition of a (k, α)-expander plays a crucial role.) Moreover, (x_0, v) is not an edge for any $v \in R$ (there is no $(h+1)$-cycle in the graph), whereas adding any edge (x_0, v) for $v \in R$ creates an $(h+1)$-cycle.

Fix a subset $\{y_1, \ldots, y_{k+1}\} \subset R$. For every $y_i \in R$, there is a path P_i ending at y_i, that can be obtained from P by a sequence of elementary rotations. Now fix y_i as the starting point of P_i and let Y_i be the set of endpoints of all paths obtained from P_i by a sequence of elementary rotations. As before, $|Y_i| \geq k+1$, no edge joins y_i to Y_i, and adding any such edge creates a cycle of length $h+1$. Altogether we have found $(k+1)^2$ pairs (y_i, x_{ij}) for $x_{ij} \in Y_i$. As every non-edge is counted at most twice, the conclusion of the lemma follows. □

The reason we are after a cycle of length $h+1$ in the above argument is that if $h+1 = n$, then a Hamilton cycle is created. Otherwise, if the graph is connected, then there will be an edge e connecting a newly created cycle C of length $h+1$ with a vertex outside C. Then opening C and appending e in an obvious way creates a longer path in G.

In order to utilize quantitatively the above argument, we introduce the notion of boosters.

Definition 3.4 *Given a graph G, a non-edge $e = (u, v)$ of G is called a* booster *if adding e to G creates a graph G', which is Hamiltonian or whose maximum path is longer than that of G.*

Note that technically every non-edge of a Hamiltonian graph G is a booster by definition.

Boosters advance a graph towards Hamiltonicity when added; adding sequentially n boosters clearly brings any graph on n vertices to Hamiltonicity.

We thus conclude from the previous discussion:

Corollary 3.5 *Let G be a connected non-Hamiltonian k-expander. Then G has at least $\frac{(k+1)^2}{2}$ boosters.*

References

[1] I. Ben-Eliezer, M. Krivelevich and B. Sudakov, *The size Ramsey number of a directed path*, Journal of Combinatorial Theory Series B 102 (2012), 743–755.

[2] B. Bollobás, **Random graphs**, 2nd ed., Cambridge University Press, Cambridge, 2001.

[3] S. Janson, T.Łuczak and A. Ruciński, **Random Graphs**, Wiley, New York, 2000.

[4] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Mathematics 14 (1976), 359–364.