Cycle lengths in sparse random graphs

MICHAEL KRIVELEVICH (TEL AVIV U.)

(JOINT WORK WITH YAHAV ALON (TAU) & EYAL LUBETZKY (NYU))

Setting

- The **binomial random graph** $G(n,p)$: $V(G) = [n]$; each edge (i,j) is in $E(G)$ with prob. $p = p(n)$, independently of all other edges.

Definition: *Let G be an n-vertex graph. The set $\mathcal{L}(G)$ is the set of all integers $\ell \in [3,n]$ such that G contains a cycle of length ℓ.*

? What can we say about $\mathcal{L}(G)$ when $G \sim G(n,p)$?

Simple observations, famous results…

- If $k = \Theta(1)$, X_k = r.v. counting k-cycles in $G(n,p)$, then $\mathbb{E}[X_k] = \frac{(n)_k}{2k} \cdot p^k = \Theta(n^k p^k)$, and
$$Var[X_k] = \Theta\left(n^k p^k + \sum_{i=1}^{k-1} n^{2k-i-1} p^{2k-i}\right).$$
So if $np \to \infty$, we have: $Var[X_k] = o(\mathbb{E}[X_k]^2)$, and by the second moment method:
$$\Pr[X_k = 0] \leq \frac{Var[X_k]}{\mathbb{E}[X_k]^2} = o(1),$$
And therefore with high probability (**WHP**) $[3,k] \subseteq \mathcal{L}(G(n,p))$.

- Komlós, Szemerédi'83; Bollobás'84: If $np - \log n - \log\log n \to \infty$ then **WHP** $G(n,p)$ is Hamiltonian. Equivalently: WHP $n \in \mathcal{L}(G(n,p))$.

  (Much easier part: $np - \log n - \log\log n \to -\infty$ then **WHP** $G(n,p)$ has vertices of degree <2

  $\Rightarrow$ WHP non-Hamiltonian)

- In fact…

Pancyclicity

Definition: an n-vertex graph G is called **pancyclic** if $\mathcal{L}(G) = [3, n]$.

Theorem (Cooper, Frieze'90):

$$\lim_{n\to\infty} \Pr[G(n,p) \text{ is pancyclic}] = \begin{cases} 0 & ; \quad np - \log n - \log\log n \to -\infty \\ 1 & ; \quad np - \log n - \log\log n \to \infty \end{cases}$$

Preparing for a proof

- 0-statement – trivial in light of Hamiltonicity threshold.
- We will present a proof sketch of the 1-statement.
- This is not the original proof.
- Helpful to demonstrate our tools for proving our results next.
- **Double exposure**:

If $(1-p_1)(1-p_2) = 1-p$ then $G(n,p_1) \cup G(n,p_2) \sim G(n,p)$

- Observe: $p_1 + p_2 \geq p$.

Pancyclicity: proof

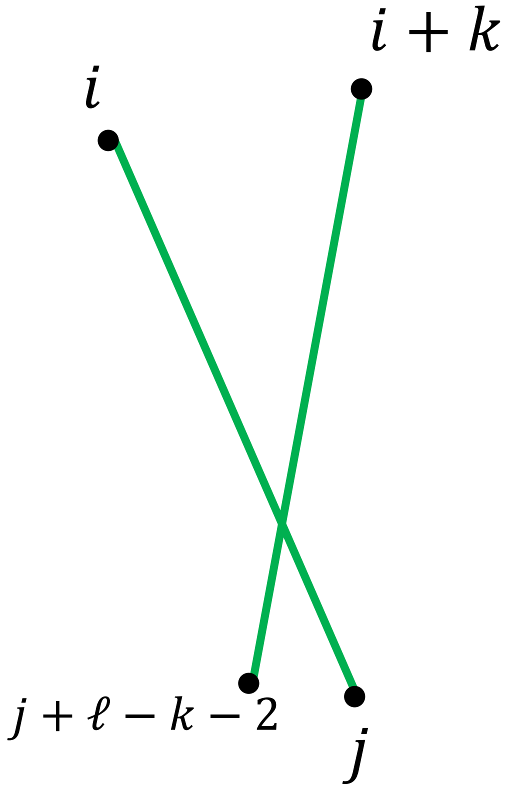
- Proof of the 1-statement (Alon, K., Lubetzky):
- Enough to show that **WHP** $[4, n] \subseteq \mathcal{L}(G(n,p))$.
- Sufficient:

$$\forall \ell \in \left[4, \frac{n}{2}+2\right] : \Pr[\{\ell, n-\ell+4\} \not\subseteq G(n,p)] = \exp(-\omega(1)\cdot \ell)$$

- Say $np - \log n - \log\log n = f(n) \to \infty, G \sim G(n,p)$ and $V(G) = [n]$.
- Double expose G as $G_1 \cup G_2 = G(n,p_1) \cup G(n,p_2)$, where $p_2 = \frac{f(n)}{2n}$, $p_1 \geq p - p_2$.
- With high probability $G(n,p_1)$ contains a Hamilton cycle C_n, WLOG $C_n = (1,2,3,\ldots,n,1)$.

Pancyclicity: proof

If for some $i \neq j \in [n], k \in [\ell - 3]$ we have $\{i, j\}, \{i + k, j + \ell - k - 2\} \in E(G_2)$...



... then $\{\ell, n - \ell + 4\} \subseteq \mathcal{L}(G)$! (double switching)

Pancyclicity: proof

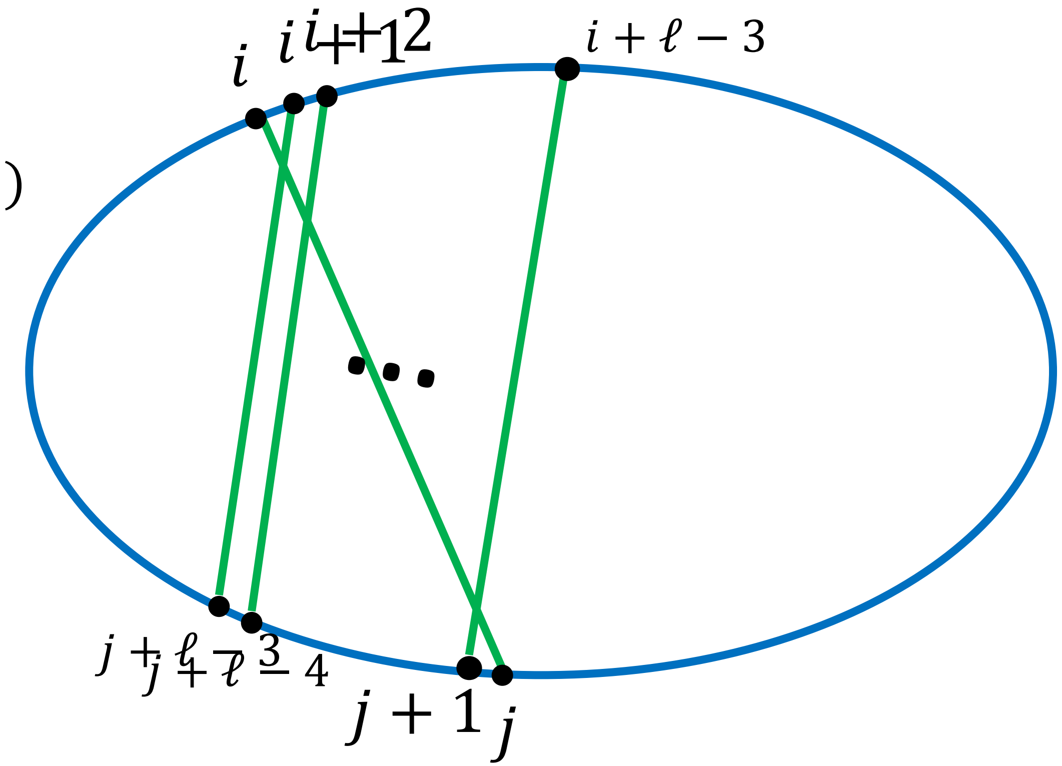
- There are $\Theta(n^2\ell)$ options for the pair i, j and the shift k.
- So $\Pr[\{\ell, n-\ell+4\} \not\subseteq \mathcal{L}(G)]$ is at most about **(*)**

$$\approx (1-p_2^2)^{\Theta(n^2\ell)}$$

$$\leq \exp(-\Theta(f(n)^2\ell))$$

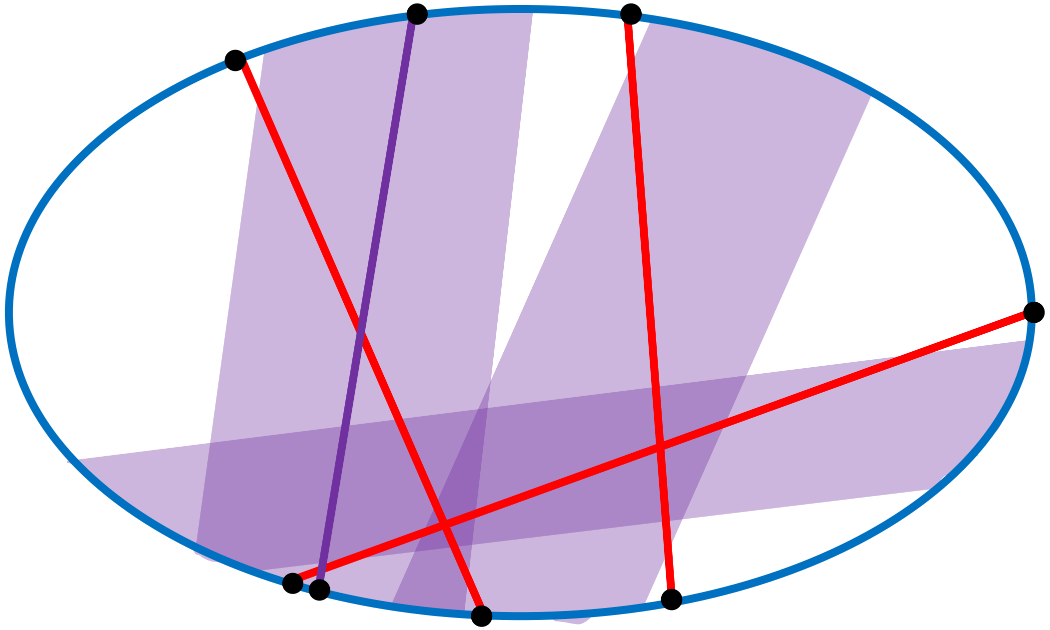
$$= \exp(-\omega(n)\cdot\ell).$$

(*) There are dependencies, so we need to be careful.



Being careful

- We want to overcome possible dependencies…
- Double exposure, yet again!
- Letting $S_{i,j} \coloneqq \{\{i+k, j+\ell-k-2\} \mid k \in [\ell-3]\}$, the "useful" edges WRT $\{i,j\}$.
- Expose G_2 as $G_3 \cup G_4 \approx G(n, p_2/2) \cup G(n, p_2/2)$.
- Show that **WHP** $\left|\bigcup_{\{i,j\}\in E(G_3)} S_{i,j}\right| = \Theta(n^2 p_2 \cdot \ell)$.
- Now the probability G_4 contains no useful edge is $\approx (1-p_2)^{\Theta(n^2 p_2 \ell)} = \exp(-\omega(1)\cdot \ell)$



k-pancyclic graphs

In fact, we have proven:

$G \sim G(n,p), np - \log\ n - \log\log n \to \infty$, is **WHP** 2-pancyclic

– a result of Cooper'91

[Definition: G_n is k-pancyclic if G has a Hamilton cycle C such that for every $3 \le \ell \le n$, G_n contains a cycle C_ℓ with $|E(C_\ell) \setminus E(C)| \le k$.]

An even stronger result:

Theorem (Cooper'92):
The threshold for Hamiltonicity is the threshold for 1-pancyclicity.

What about smaller p's?

Theorem (Łuczak'91):
Let $np \to \infty$, $G \sim G(n,p)$, and let $n_{\leq 1}$ be the number of vertices with degree 0 or 1 in G.
Then for every $\varepsilon > 0$, **WHP**: $[3, n-(1+\varepsilon)n_{\leq 1}] \subseteq \mathcal{L}(G)$.

- This can be proved in the same spirit as pancyclicity (and Łuczak's original proof has some similarities to the proof we presented).
- Relies on:

Theorem (Frieze'86):
Let $np \to \infty$, $G \sim G(n,p)$, and let $n_{\leq 1}$ be the number of vertices with degree 0 or 1 in G.
Then for every $\varepsilon > 0$, **WHP**: $\max\big(\mathcal{L}(G)\big) \geq n-(1+\varepsilon)n_{\leq 1}$.

- Expose $E(G)$ in two parts, find a long cycle in one part and helpful pairs of edges in the other.

What about even smaller p's?

? Can we say something when $np = O(1)$?

- Let's assume $np \geq 1 + \varepsilon$, since otherwise the largest connected component is **WHP** sublinear.
- There are now many cycle lengths we can no longer expect to appear **WHP**!

Theorem (Bollobás 81; Karoński, Ruciński 81):

Let $c > 0,\ k \in \mathbb{N}, \lambda_{c,k} \coloneqq \frac{c^k}{2k}$. Then

$$\#k\text{-cycles in } G\left(n, \frac{c}{n}\right) \xrightarrow{D} \mathrm{Poi}(\lambda_{c,k}).$$

- So for any constant k, with probability bounded away from 0, $k \notin \mathcal{L}(G)$.
- Also, no cycle lengths larger than the size of the giant component (and in fact the 2-core)! This is a linear size interval not in $\mathcal{L}(G)$ **WHP**.

What about even smaller p's? (cont.)

A side remark:

Definition: for some $\beta > 0$, an n-vertex graph G is called a **β-graph** if every disjoint vertex sets $A, B \subseteq V(G)$ of size at least βn are connected by an edge.

- $G\left(n, \frac{c}{n}\right)$ is **WHP** a β-graph for an appropriate $\beta = \beta(c) = o_c(1)$.

Theorem (Friedman, K.'21):
Let $0 < \beta < 0.05$. Then there are $b_1 = b_1(\beta), b_2 = b_2(\beta) = O(\beta)$ such that if G is a β-graph then

$$[b_1 \log n, (1 - b_2)n] \subseteq \mathcal{L}(G).$$

- This implies that, **WHP**, $\mathcal{L}\left(G\left(n, \frac{c}{n}\right)\right)$ contains an interval of size $(1 - o_c(1))n$.
- Since **WHP** $G\left(n, \frac{c}{n}\right)$ contains linearly many isolated vertices, we cannot hope for $(1 - o(1))n$.

Our first result

Theorem 1:

There is $C_0 > 0$ such that for **almost every** $c > C_0$, if $G \sim G\left(n, \frac{c}{n}\right)$, then for every $\varepsilon > 0$ and sequence $\omega_n \to \infty$, **WHP**:

$$\left[\omega_n, (1-\varepsilon)\max\left(\mathcal{L}(G)\right)\right] \subseteq \mathcal{L}(G).$$

- In the lower range this is best possible (we already know how short cycles are distributed).
- In the upper range we capture all but $o(n)$ lengths.
- There are known bounds for $\max\left(\mathcal{L}\left(G(n, c/n)\right)\right)$. (We will mention some (more) soon.)
- Sadly, we can only prove for almost every $c > C_0$…
- But we can still say some things about <u>every</u> $c > 1$.

Proof of Theorem 1

- The proof goes similarly to our proof of pancyclicity.
- First, we prove this main lemma:

Lemma:

Let $\delta > 0$, $G \sim C_n \cup G\left(n, \frac{\delta}{n}\right)$. Then **WHP**, for every sequence $\omega_n \to \infty$,

$$[\omega_n, n - \omega_n] \subseteq \mathcal{L}(G).$$

- You already know the proof…
- This time $np \nrightarrow \infty$, so $(1 - p^2)^{\Theta(n^2 \ell)} \to 0$ only when $\ell \to \infty$ (hence we do not capture the full interval).

Proof of Theorem 1 (cont.)

- **Theorem 1** is a consequence of the lemma and the following result:

Theorem (Anastos, Frieze'21):
There is a monotone non-decreasing function $f: \mathbb{R}^+ \to (0,1)$ and a constant $C_0 > 0$ such that, if $c > C_0$, then

$$\frac{\max\left(\mathcal{L}\left(G\left(n,\frac{c}{n}\right)\right)\right)}{n} \overset{\mathrm{P}}{\to} f(c) \, .$$

- Expose G as $G_1 \cup G_2 \approx G(n,(c-\delta)/n) \cup G(n,\delta/n)$.
- Then **WHP** G_1 contains a cycle C of length $f(c-\delta)n - o(n)$.
- Apply the main lemma on $C \cup G_2[V(C)]$ to get $[\omega_n, f(c-\delta)n - o(n)] \subseteq \mathcal{L}(G)$.
- Choose δ to be small enough so that $f(c-\delta) \geq \left(1-\frac{1}{2}\varepsilon\right) f(c)$, say.
- We can only do the last part in points where f is continuous, hence "almost every" c.

Extensions to Theorem 1

- What we **actually** proved is the following:

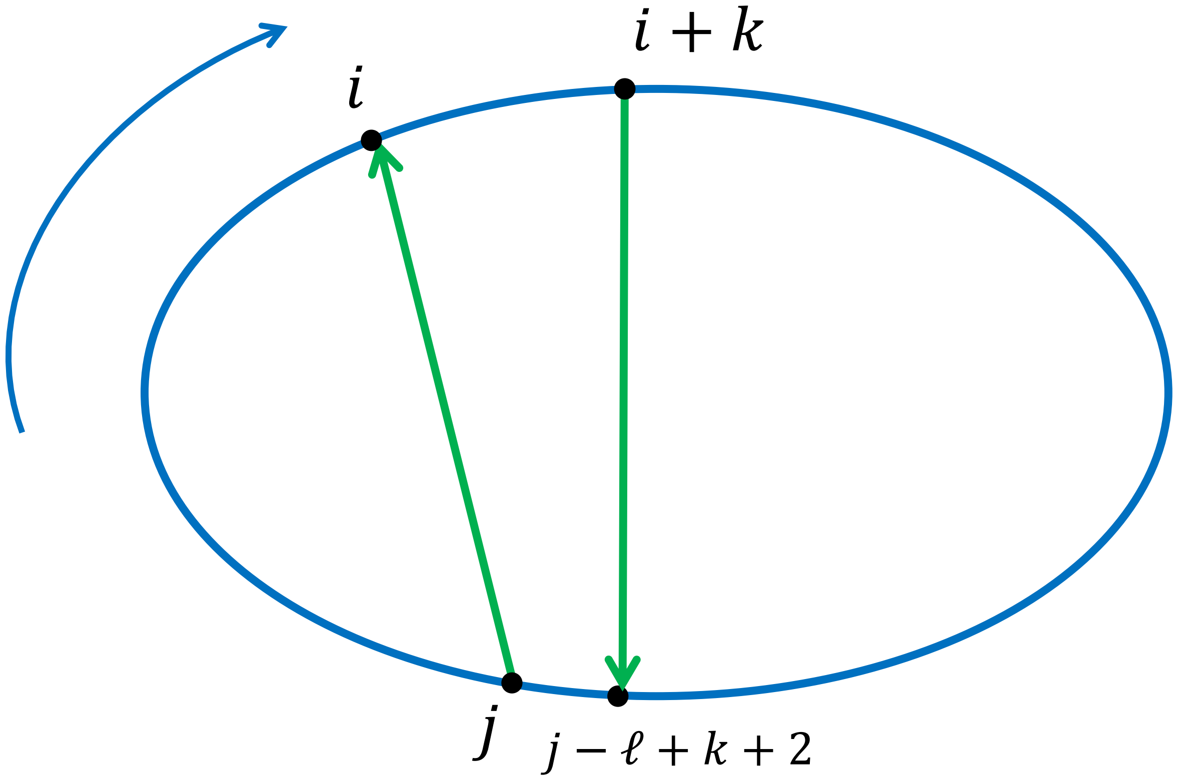
If $B(c,n)$ is a **WHP** lower bound on $\max\left(\mathcal{L}(G(n,c/n))\right)$, then **WHP** for every $c > 1, \delta > 0$, $\omega_n \to \infty$:

$$[\omega_n, B(c-\delta,n) - \omega_n] \subseteq \mathcal{L}\left(G(n,c/n)\right)$$

- So for any $c > 1$, given lower bounds on the maximum length cycle in $G(n,c/n)$ we get some (possibly weaker) statements on $\mathcal{L}\left(G(n,c/n)\right)$.
- For example: if $c = 1 + \varepsilon < 1 + \varepsilon_0$, then it is known that the maximum length is **WHP** at least $\frac{4}{3}\varepsilon^2 n$, and so for all $\gamma < \frac{4}{3}$ **WHP** $[\omega_n, \gamma\varepsilon^2 n] \subseteq \mathcal{L}\left(G(n,c/n)\right)$.
- Everything we said so far is also true for directed cycles in $D(n,p)$!

Extensions to Theorem 1

- Adjustments to $D(n, p)$:



Further applications of the main lemma

Our machinery can be applied in different setups.

Ex.: cycles in randomly perturbed graphs

Theorem (K, Reichman, Samotij'15):

Let T_n be a tree on n vertices with $\Delta(T_n) \leq \Delta = O(1), \delta > 0$, and let $G \sim T_n \cup G\left(n, \frac{\delta}{n}\right)$.

Then **WHP**, G contains a cycle of length $\geq cn$, for $c = c(\Delta, \delta) > 0$.

Invoking the main lemma gives now:

Theorem:

Let T_n be a tree on n vertices with $\Delta(T_n) \leq \Delta = O(1), \delta > 0$, and let $G \sim T_n \cup G\left(n, \frac{\delta}{n}\right)$.

Then **WHP**, $\mathcal{L}(G)$contains an interval of length $\geq cn$, for $c = c(\Delta, \delta) > 0$.

Onward to $G(n,d)$

- What can we say about $\mathcal{L}(G)$ when G is a **random $\boldsymbol{d}$-regular graph** $G(n,d)$?
- When $d = 1$ no cycles, when $d = 2$ likely only logarithmically many. Assume $d \geq 3$.
- In $G(n, c/n)$ we were restricted by the maximum cycle possibly being short (missing isolates)...
- No longer!

Theorem (Robinson, Wormald'92,94):
For every fixed $d \geq 3$, **WHP** $G(n,d)$ is Hamiltonian.

- On the other side of the interval things have not improved…

Theorem (Bollobás'80, Wormald'81):

Let $d \geq 3$, $k \in \mathbb{N}, \lambda_{d,k} \coloneqq \frac{(d-1)^k}{2k}$. Then

$$\#k\text{-cycles in } G(n,d) \xrightarrow{D} \text{Poi}(\lambda_{d,k})$$

Our second result

Theorem 2:
For every integer $d \geq 3$, if $G \sim G(n,d)$, then for every sequence $\omega_n \to \infty$, **WHP**:

$$[\omega_n, n] \subseteq \mathcal{L}(G).$$

- This is best possible.
- We will prove for $d = 3$, which implies the theorem for all $d \geq 3$ (assuming that n is even).
- The odd case can be proved by showing this for $d = 4$ (basically the same, we will not do this here).

Proof of Theorem 2

- The proof for the interval $[\omega_n, n-\omega_n]$ is very similar to $G(n,p)$.
- As a parallel to double exposure, we have the following useful result (contiguity):

Theorem:
If $\mathcal{P}$ is a monotone graph property, then
$$\Pr[G(n,3)\in\mathcal{P}]\to 1 \Leftrightarrow \Pr[C_n\cup G(n,1)\in\mathcal{P}]\to 1.$$

- So we just need to show that, for $\ell\in\left[\omega_n,\frac{n}{2}+2\right]$, a random perfect matching on $[n]$ contains two edges of the form $\{i,j\},\{i+k,j+\ell-k-2\}$ with probability $1-\exp\big(-\Omega(\ell)\big)$.
- Proofs in $G(n,d)$ may look a bit messier than in $G(n,p)$, but trust us this is still true ☺.

Proof of Theorem 2

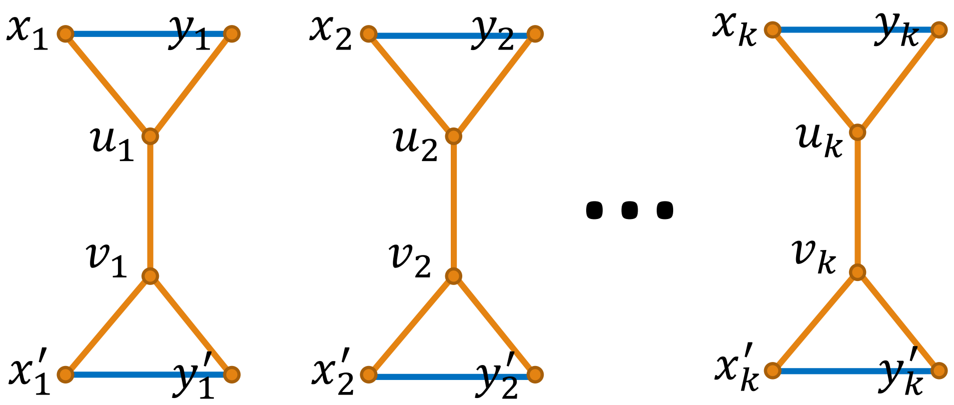
- We still need to take care of $[n-\omega_n, n]$!
- We need the following theorem:

Theorem (Robinson, Wormald'01):
Let $d \geq 3$, $G \sim G(n,d)$, $m_1, m_2 = o(\sqrt{n})$. Set $E_1, E_2 \subseteq E(G)$ to be randomly chosen subsets of sizes m_1, m_2 respectively. Then **WHP** G contains a Hamilton cycle which includes all edges of E_1 and avoids all edges of E_2.

- We want to use this to show that **WHP** $n-\ell \in \mathcal{L}(G)$, for $\ell \in [1, \omega_n]$.
- Since we can assume ω_n grows arbitrarily slowly, this is good enough.
- Prove separately for ℓ even / ℓ odd.

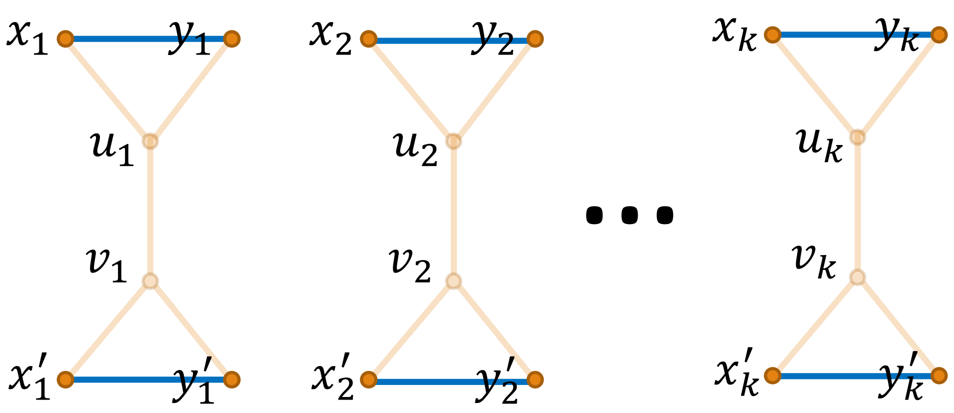
Proof of Theorem 2: $\ell = 2k$

- Let $G \sim G(n,3)$, and pick k edges $\{u_1, v_1\}, \ldots, \{u_k, v_k\}$ of G at random.
- For $i \in [k]$ denote by $\{x_i, y_i\}, \{x_i', y_i'\}$ the other two neighbours of u_i, v_i respectively.
- $G' := G\left([n] \setminus (\bigcup\{u_i, v_i\})\right) \cup \{\text{blue edges}\} \sim G(n-\ell, 3)$, and the blue edges are ℓ random edges in G'.
- A Hamilton cycle in G' which avoids these edges is an $n - \ell$ cycle in G!
- Invoke **Robinson, Wormald**.



Proof of Theorem 2: $\ell = 2k - 1$

- Let $G' \sim G(n - \ell - 1,3)$ and the blue edges be the same as in the case $\ell = 2k$.
- This time, invoke theorem to find Hamilton cycle C in G' which avoids the blue edges, **except** $\{x_1, y_1\}$, which is included in C.
- $C - \{x_1, y_1\} + \{x_1, u_1\} + \{u_1, y_1\}$ is an $n - \ell$-cycle in G!




That's all Folks!