Digraphs

In this chapter we study the random digraph $D_{n,p}$. This has vertex set $[n]$ and each of the $n(n-1)$ possible edges occurs independently with probability p.

We will first study the size of the strong components of $D_{n,p}$.

Case 1: $p = c/n$, $c \ll 1$

We will show that in this case

Theorem 1

Whp

(i) all strong components of $D_{n,p}$ are either cycles or single vertices.

(ii) The number of vertices on cycles is at most ω, for any $\omega = \omega(n) \to \infty$

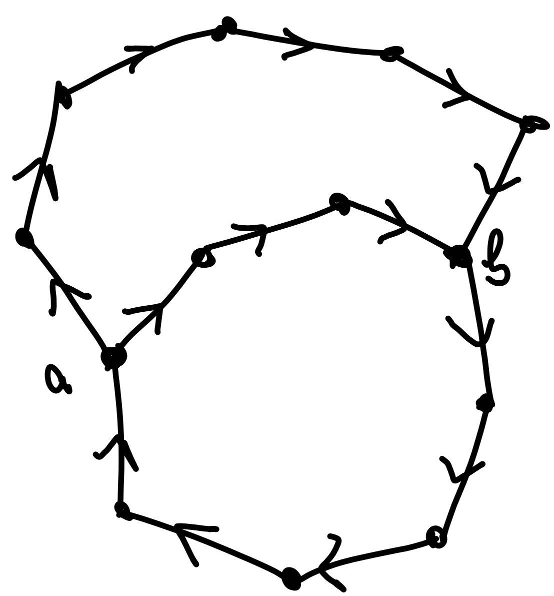
Proof

The expected number of cycles is

$$\sum_{k=2}^{n} \binom{n}{k} (k-1)! \left(\frac{c}{n}\right)^k \leq \sum_{k=2}^{n} \frac{c^k}{k} = O(1).$$

Part (ii) now follows from the Markov inequality

To tackle (i) we argue that if there is a component that is not a cycle or single vertex then there is a cycle C and vertices $a, b \in C$ and a path P from a to b that is internally disjoint from C.



However, the expected number of such sub-graphs is bounded by

$$\sum_{k=2}^{n} \sum_{\ell=1}^{n-k} \binom{n}{k} (k-1)! \left(\frac{c}{n}\right)^{k} \binom{n}{\ell} \ell! \left(\frac{c}{n}\right)^{\ell+1}$$

$$\leq \sum_{k=2}^{\infty} \sum_{\ell=1}^{\infty} \frac{c^{k+\ell+1}}{k\,n} = O\left(\frac{1}{n}\right).$$

Here ℓ is the number of vertices on the path P, excluding a, b.

$\square$

We now consider the case $p = c/n$ where $c > 1$.

We will prove the following theorem that is a directed analogue of the existence of a giant component in $G_{n,p}$.

Theorem 2

Let x be defined by $x < 1$ and $xe^{-x} = ce^{-c}$.

Then whp $D_{n,p}$ contains a unique strong component of size $\approx (1 - \frac{x}{c})^2 n$. All other strong components are of logarithmic size.

General Strategy: For a vertex v let

$$D^+(v) = \{w : \exists \text{ path } v \text{ to } w \text{ in } D_{n,p}\}$$

$$D^-(v) = \{w : \exists \text{ path } w \text{ to } v \text{ in } D_{n,p}\}.$$

We will first prove

Lemma 1

There exist constants α, β (dependent only on c) such that whhp

$\nexists v$ such that $|D^\pm(v)| \in [\alpha \log n, \beta n]$.

Proof

If there is a v such that $|D^+(v)| = s$ then $D_{n,p}$ contains a tree T of size s, rooted at v such that (i) all arcs are oriented away from v and (ii) there are no arcs oriented from $V(T)$ to $[n] \setminus V[T]$.

The expected number of such trees is bounded above by

$$\binom{n}{s} s^{s-2} \left(\frac{c}{n}\right)^{s-1} \left(1-\frac{c}{n}\right)^{s(n-s)} \leq$$

$$\frac{n}{cs^2} \left(c e^{1-c+s/n} \right)^{s} .$$

Now $ce^{1-c} < 1$ for $c \neq 1$ and so there exists β such that when $s \leq \beta n$ we can bound $ce^{1-c+s/n}$ by some constant $\gamma < 1$ (γ depends only on c).
In which case

$$\frac{n}{cs^2} \gamma^s \leq n^{-3} \text{ for } s \geq \frac{4}{\log 1/\gamma} \log n .$$

$\square$

Fix a vertex $v \in [n]$ and consider a directed breadth first search from v.

Let $S_0^+ = \{v\}$ and given $S_0^+, S_1^+, \ldots, S_k^+ \subseteq [n]$ let $T_k^+ = \bigcup_{i=1}^{k} S_i^+$ and let

$$S_{k+1}^+ = \{ w \notin T_k^+ : \exists x \in T_k^+ \text{ s.t. } (x,w) \in E(D_{n,p}) \}.$$

Not surprisingly, we can show that the sub-graph Γ_k induced by T_k^+ is close in distribution to the tree defined by the first $k+1$ levels of a Galton-Watson branching process with $Po(c)$ as the distribution of the number of offspring from a single parent.

<u>Lemma 2</u>

If $\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_k$ and $\hat{T}_k$ are defined with respect to the branching process and if $k \le k_0 = \log^3 n$ and $s_0, s_1, \ldots, s_k \le \log^4 n$ then

$$\Pr(|S_i^+| = s_i,\ 0\le i\le k) = \left(1+O\left(\tfrac{1}{n^{1-o(1)}}\right)\right)\Pr(|\hat{S}_i| = s_i,\ 0\le i\le k).$$

Proof

$$\Pr(|\hat{S}_i| = s_i,\ 0\le i\le k) = \prod_{i=1}^{k} \frac{(cs_{i-1})^{s_i} e^{-cs_{i-1}}}{s_i!},$$

Furthermore, putting $t_i := s_0 + s_1 + \dots + s_i$ we have

$$\Pr(|S_i^+| = s_i,\ 0\le i\le k) = \prod_{i=1}^{k} \binom{s_{i-1}(n-t_i)}{s_i}\left(\frac{c}{n}\right)^{s_i}\left(1-\frac{c}{n}\right)^{s_{i-1}(n-t_i)-s_i}$$

and the lemma follows by simple estimations. $\square$

Lemma 3

(a) $\Pr(|S_i^+| \geq s\log n \mid |S_{i-1}^+| = s) \leq n^{-10}$.

(b) $\Pr(|\hat{S}_i| \geq s\log n \mid |\hat{S}_{i-1}| = s) \leq n^{-10}$.

Proof

(a)

$$\Pr(|S_i^+| \geq s\log n \mid |S_{i-1}^+| = s) \leq$$

$$\Pr\left(B\left(sn, \tfrac{c}{n}\right) \geq s\log n\right) \leq \binom{sn}{s\log n}\left(\frac{c}{n}\right)^{s\log n}$$

$$\leq \left(\frac{sne\,c}{sn\log n}\right)^{s\log n}$$

$$\leq \left(\frac{ec}{\log n}\right)^{\log n}.$$

(b) is similar.

Next let

$$\mathcal{F} = \{ \exists i : |T_i^+| > \log^2 n \}$$

Lemma 4

$$\Pr(\mathcal{F}) = 1 - \frac{x}{c} + o(1)$$

Proof

$$\Pr(\mathcal{F}) = \Pr(\mathcal{F}_1) + o(1)$$

where

$$\mathcal{F}_1 = \{ \exists i \leq \log^2 n : |T_0^+|, \dots, |T_{i-1}^+| < \log^2 n \leq |T_i^+| \}$$

This follows from Lemma 3.

Applying Lemma 2 (on p12) we see that

$$\Pr(\mathcal{F}_1) = \Pr(\hat{\mathcal{F}}_1) + o(1)$$

where $\hat{\mathcal{F}}_1$ is defined w.r.t. the branching process.

Now let $\hat{\mathcal{E}}$ be the event that the branching process becomes extinct.

We write

$$\Pr(\hat{\mathcal{F}}_1) = \Pr(\hat{\mathcal{F}}_1 \mid \neg\hat{\mathcal{E}})\Pr(\neg\hat{\mathcal{E}}) + \Pr(\hat{\mathcal{F}}_1 \wedge \mathcal{E}). \quad (1)$$

To estimate (7) we first define

$$\rho = \Pr(\hat{\mathcal{E}})$$

$$= \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} \rho^k.$$

This is if the origin of the process has k children then each of the processes spawned by them must become extinct for $\mathcal{E}$ to occur.

Thus

$$\rho = e^{c\rho - c}.$$

Substituting $\rho = \frac{\xi}{c}$ proves that

$$\Pr(\hat{\mathcal{E}}) = \frac{\xi}{c} \text{ where } \frac{\xi}{c} = e^{\xi - c}$$

and so $\xi = x$.

The lemma will follow from (1) [p16] and this and $\Pr(\hat{\mathcal{F}} \mid \neg \mathcal{E}) = 1 - o(1)$ (see Lemma 3 [p14]) and

$$\Pr(\hat{\mathcal{F}} \wedge \mathcal{E}) = o(1). \qquad (2)$$

Let us break the first $\log^2 n$ generations of the branching process into $\log n$ rounds of length $\log n$.

If $\neg\mathcal{E}$ occurs then we start each round with a non-zero population.

Claim 1

Each member of this population has a probability of at least $\epsilon > 0$ of producing $\log^2 n$ descendants at depth $\log n$. Here $\epsilon > 0$ depends only on c

and so

$$\Pr(\overline{\mathcal{F}} \wedge \neg\mathcal{E}) \leq (1-\epsilon)^{\log n} = o(1).$$

If the current population of the process is s then the probability that it reaches size at least $\frac{c+1}{2}s$ in the next round is

$$\sum_{k \geq \frac{c+1}{2}s} \frac{(cs)^k e^{-cs}}{k!} \geq 1 - e^{-\alpha s}$$

for some constant $\alpha > 0$ provided $s \geq 100$, say.

Now there is a positive probability ϵ_1 say that a single object spawns at least 100 descendants and so there is a probability of at least

$$\epsilon_1 \left(1 - \sum_{s=100}^{\infty} e^{-\alpha s}\right)$$

that a single object spawns

$$\left(\frac{c+1}{2}\right)^{\log n} \gg \log^2 n$$

descendants at depth $\log n$.

This proves Claim 1 ([p19]) and completes the proof of Lemma 4.

$\square$

We state for future reference that the above argument supports the following claim.

Claim 2

$\Pr(\exists i : |S_i^+| \geq \log^2 n$ and $|T_i^+|$

We must now consider the probability that both $D^+(v)$ and $D^-(v)$ are large.

<u>Lemma 5</u>

$$\Pr\left(|D^-(v)| \geq \log^2 n \mid |D^+(v)| \geq \log^2 n\right) = 1 - \frac{x}{c} + o(1).$$

<u>Proof</u>

Expose $S_0^+, S_1^+ \dots S_k^+$ until either $S_k^+ = \emptyset$ or we see that $|T_k^+| \geq \log^2 n$.

Now let S denote the set of edges/vertices defined by $S_0^+, S_1^+, \dots S_k^+$ we see that (see Lemma 2 [p12])

Let $\mathcal{C}$ be the event that there are no edges from T_ℓ^- to S_k^+ where T_ℓ^- is the set of vertices we reach through our BFS into v, up to the point where we first find that $|D^-(v)| < \log^2 n$ or $\geq \log^2 n$. Then

$$\Pr(\mathcal{C}) = 1 - \frac{1}{n^{1-o(1)}}$$

and

$$\Pr(|S_i^-| = s_i, 0 \leq i \leq k \mid \mathcal{C}) = \prod_{i=1}^{k} \binom{s_{i-1}(n'-t_i)}{s_i} \left(\frac{c}{n}\right)^{s_i} \left(1-\frac{c}{n}\right)^{s_{i-1}(n'-t_i)-s_i}$$

where $n' = n - |T_k^+|$.

Given this we can prove a conditional version of Lemma 2 and continue as before. $\square$

We have now shown that if

$$S = \{ v : |D^+(v)|, |D^-(v)| > \alpha \log n \}$$

then

$$E(|S|) = (1+o(1))\left(1 - \frac{x}{c}\right)^2 n.$$

We also claim that for any two vertices v, w

$$\Pr[v, w \in S] = (1+o(1)) \Pr(v \in S) \Pr(w \in S) \quad (3)$$

and therefore the Chebyshev inequality implies that whp

$$|S| = (1+o(1))\left(1 - \frac{x}{c}\right)^2 n.$$

But (3) follows in a similar manner to the proof of Lemma 5 (p22).

All that remains of the proof of Theorem 2 is to show that

$$\text{whp } S \text{ is a strong component.} \quad (4)$$

(Any $v \notin S$ is in a strong component of size $\leq 2\log n$).

We prove (4) by arguing that

$$\Pr(\exists v, w \in S : w \notin D^+(v)) = o(1). \qquad (5)$$

For this we expose $S_0^+, S_1^+, \ldots, S_k^+$ until we find that $|T_k^+(v)| \geq n^{\frac{1}{2}} \log n$.

At the same time we expose $S_0^-, S_1^-, \ldots, S_\ell^-$ until $|T_\ell^-(w)| \geq n^{\frac{1}{2}} \log n$.

If $w \notin D^+(v)$ then this experiment will have tried at least $(n^{\frac{1}{2}} \log n)^2$ times to find an edge from $D^+(v)$ to $D^-(w)$ and failed every time.

The probability of this is at most

$$\left(1 - \frac{c}{n}\right)^{n \log^2 n} = o(n^{-2}).$$

This completes the proof of Theorem 2.

$\square$

Strong Connectivity Threshold

Here we prove

Theorem 3

Suppose that $p = \frac{\log n + c_n}{n}$. Then

$$\lim_{n\to\infty} \Pr\left(D_{n,p} \text{ is strongly connected}\right) = \begin{cases} 0 & c_n \to -\infty \\ e^{-2e^{-c}} & c_n \to c \\ 1 & c_n \to +\infty \end{cases}$$

$$= \lim_{n\to\infty} \Pr\left(\nexists\, v \text{ such that } d^+(v) = 0 \text{ or } d^-(v) = 0\right).$$

Proof

We leave it as an exercise to prove that

$$\lim_{n\to\infty} \Pr(\nexists\, v \text{ such } d^+(v)=0 \text{ or } d^-(v)=0) = \begin{cases} 1 & c_n \to -\infty \\ 1-e^{-2e^{-c}} & c_n \to c \\ 0 & c_n \to \infty \end{cases}$$

Given this, one only has to show that if $c_n \not\to -\infty$ then whp there does not exist a vertex v such that $2 \le |D^+(v)| \le n/2$ or $2 \le |D^-(v)| \le n/2$.

But, here with $s+1 = |D^+(v)|$,

$$\Pr(\exists v) \le 2n \sum_{s=1}^{n/2} \binom{n}{s} (s+1)^{s-1} \left(\frac{c}{n}\right)^{s} (1-p)^{(s+1)(n-1-s)}$$

$$= o(1). \quad \text{(Exercise)}$$

Hamilton Cycles

Here we prove the following remarkable inequality:

Theorem 4

$$\Pr(D_{n,p} \text{ is Hamiltonian}) \geq \Pr(G_{n,p} \text{ is Hamiltonian})$$

Proof

Remark: This shows that if $p = \frac{\log n + \log\log n + \omega}{n}$ then $D_{n,p}$ is Hamiltonian whp. This result has been strengthened but it requires a much more difficult argument. The $\log\log n$ can be eliminated.

Proof

We consider a sequence of random digraphs $\Gamma_0, \Gamma_1, \Gamma_2, \dots \Gamma_N$, $N = \binom{n}{2}$ defined as follows:

Let $e_1, e_2, \dots e_N$ be an enumeration of the edges of K_n. Each $e_i = (v_i, w_i)$ gives rise to two directed edges $\overrightarrow{e_i} = (v_i, w_i)$ and $\overleftarrow{e_i} = (w_i, v_i)$.

In Γ_i we include $\overrightarrow{e_j}$ and $\overleftarrow{e_j}$ independently of each other, with probability p, for $j \leq i$.

While for $j > i$ we include both or neither with probability p.

Thus Γ_0 is just $G_{n,p}$ with each edge (v,w) replaced by a pair of directed edges (v,w), (w,v) and $\Gamma_N = D_{n,p}$. Theorem 4 follows from

$$\Pr(\Gamma_i \text{ is Hamiltonian}) \geq \Pr(\Gamma_{i-1} \text{ is Hamiltonian})$$

To prove this we condition on the existence or otherwise of directed edges associated with $e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_N$.

Let $\mathcal{C}$ denote this conditioning.

Either $\mathcal{C}$ is such that

(a) $\mathcal{C}$ gives us a Hamilton cycle without arcs associated with e_i or there is no Hamilton cycle even if both $\vec{e}_i, \overleftarrow{e}_i$ occur

Or $\mathcal{C}$ is such that

(b) $\exists$ a Hamilton cycle if at least one of $\vec{e}_i, \overleftarrow{e}_i$ occurs.

In Γ_{i-1} this happens with probability p

In Γ_i this happens with probability $1-(1-p)^2 > p$

[We will never require that <u>both</u> $\vec{e}_i, \overleftarrow{e}_i$ occur.]

$\square$