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: \( p = \frac{c}{n}, \ c < 1 \)

We will show that in this case

**Theorem 1**

Whp

1. all strong components of \( G_{np} \) are either cycles or single vertices.

2. The number of vertices on cycles is at most \( \omega \), for any \( \omega = \omega(n) \to 0 \).
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

\[ \sum_{k=2}^{n} \sum_{l=1}^{n-k} \frac{n}{k} (k-1)! \left( \frac{c}{n} \right)^k (\frac{n}{l}) l! \left( \frac{c}{n} \right)^{l+1} \]

\[ \sum_{k=2}^{\infty} \sum_{l=1}^{\infty} \frac{c^{k+l+1}}{k n} = O\left( \frac{1}{n} \right). \]

Here \( l \) is the number of vertices on the path \( P \), excluding \( a, b \).
We now consider the case $p = \frac{c}{n}$ where $c \geq 1$.

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 $\omega$ be defined by $0 < \omega < 1$ and $\omega \in (0, \frac{1}{n})$. Then with $D_{n,p}$ contains a unique strong component of size $\Omega(n(1-\omega)^2)$. All other strong components are of logarithmic size.
General Strategy: For a vertex \( v \) let
\[
D^+(v) = \{ w : \exists \text{ path } v \to w \in D_{n,p} \}
\]
\[
D^-(v) = \{ w : \exists \text{ path } w \to v \in D_{n,p} \}.
\]

We will first prove

**Lemma 1**
There exist constants \( \alpha, \beta \) (dependent only on \( c \)) such that w.h.p.

\[ \exists v \text{ such that } |D^+(v)| \in [\alpha \log n, \beta n]. \]
Proof

If there is a \( v \) such that \( |D^+(v)| = s \) then \( O_{v,p} \) contains a line \( Q \) 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 \( E(T) \backslash V(T) \).

The expected number \( f \) such lines is bounded above by
\[ \left( \frac{n}{S} \right)^{S-2} \left( \frac{c}{n} \right)^{S-1} (1 - \frac{c}{n})^{S(n-S)} \leq \]
\[ \frac{n}{cS^2} \left( ce^{1-c} + s/n \right)^S. \]

Now \( ce^{1-c} < 1 \) for \( c + 1 \) and so there exists \( \beta \) such that when \( S \leq \beta n \) we can bound \( ce^{1-c} + s/n \) by some constant \( \delta < 1 \) (\( \delta \) depends only on \( c \)). In which case
\[ \frac{n}{cS^2} \delta^S \leq n^{-3} \text{ for } S \geq \frac{4}{\log^{1/3} \log n}. \]
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

\[
S_{k+1}^+ = \{ w \in T_k^+ : \exists x \in T_k^+ \text{ s.t. } (x, w) \in \{(D_{n, p})^2 \} \}.
\]

Not surprisingly, we can show that the sub-graph \( \Gamma_k^+ \) 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 \( \mathcal{P}_0(c) \) as the distribution of the number of offspring from a single parent.

**Lemma 2**

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 \( k < k_0 = \log^3 n \) and \( s_0, s_1, \ldots, s_k \leq \log^{4} n \), then
\[ \Pr\left( |\hat{S}_i^+| = s_i, 0 \leq i \leq k \right) = \left(1 + o\left(\frac{1}{n^{k-1}}\right)\right) \Pr\left( |\hat{S}_i^-| = s_i, 0 \leq i \leq k \right). \]

**Proof**

\[ \Pr\left( |\hat{S}_i^-| = s_i, 0 \leq i \leq k \right) = \prod_{i=1}^{k} \frac{(c_{s_{w-1}})^{s_i} e^{-c_{s_{w-1}}}}{s_i!}. \]

Furthermore, putting \( t_i = s_o + s_1 + \ldots + s_i \) we have

\[ \Pr\left( |\hat{S}_i^+| = s_i, 0 \leq i \leq k \right) = \prod_{i=1}^{k} \left( \frac{\binom{s_{w-1}(n-t_i)}{s_i}}{s_i!} \right) \left( \frac{c_{s_i} s_{w-1}(n-t_i) - s_i}{n} \right) \]

and the lemma follows by simple estimations. \( \square \)
Lemma 3

(a) \( \Pr ( |S_v^+| \geq s \log n \mid |S_{v^-}| = s) \leq n^{-10} \).

(b) \( \Pr ( |S_v^-| \geq s \log n \mid |S_{v^-}| = s) \leq n^{-10} \).

Proof

(a)
\[
\Pr ( |S_v^+| \geq s \log n \mid |S_{v^-}| = s) \leq \Pr ( B(sn, \frac{e}{n}) \geq s \log n ) \leq \left( \frac{sn}{s \log n} \right)^s \log n \leq \left( \frac{en}{s n \log n} \right)^s \log n \leq \left( \frac{e c}{\log n} \right)^s \log n .
\]

(b) is similar.
Next let
\[ \mathcal{F} = \left\{ \exists i : \left| \mathcal{F}_i^+ \right| > \log^2 n \right\} \]

**Lemma 4**
\[ \Pr(\mathcal{F}) = 1 - \frac{n}{n} + o(1) \]

**Proof**
\[ \Pr(\mathcal{F}) = \Pr(\mathcal{F}_1) + o(1) \]
where
\[ \mathcal{F}_1 = \left\{ \exists i : \log^2 n : \left| \mathcal{F}_i^+ \right| > \log^2 n < \left| \mathcal{F}_i^+ \right| \right\} \]

This follows from Lemma 3.
Applying Lemma 2 (on p. 12) we see that
\[ P_r(\hat{F}_1) = P_r(\hat{F}_1^+) + o(1) \]
where \( \hat{F}_1^+ \) is defined w.r.t. the branching process.

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

We write
\[ P_r(\hat{F}_1) = P_r(\hat{F}_1 | \neg \hat{E}) P_r(\neg \hat{E}) + P_r(\hat{F}_1 \land \hat{E}). \tag{1} \]
To estimate (1) we first define
\[ \rho = \rho_S(\mathcal{E}) \]
\[ = \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} \rho^k. \]

This is the origin of the process having \( k \) children if each of the processes spawned by them must become extinct for \( \mathcal{E} \) to occur. Thus
\[ \rho = e^{c \rho - c}. \]

Substituting \( \rho = \frac{\mathcal{E}}{c} \) proves that
\[ P(\hat{E}) = \frac{\delta_0}{c} \quad \text{where} \quad \frac{\delta_0}{c} = e^{-c} \]

and so \( \delta = 2c \).

The lemma will follow from (1) \([\rho 16]\)

and this and \( P(\hat{F} \mid \neg E) = 1 - o(1) \)

(see Lemma 3 \([\rho 14]\)) and

\[ P(\hat{F} \land E) = o(1). \quad (2) \]
Let us break the first \( \log n \) generations of the branching process into \( \log n \) rounds of length \( \log n \).

If \( \overline{E} \) occurs then we start each round with a non-zero population.

Claim 1

Each member of this population has a probability \( \geq \log 2n \) of producing \( \log n \) descendants at depth \( \log n \). Here \( \geq \) depends only on \( C \) and so

\[
P(\overline{E} \cap \overline{E}) \leq (1 - \epsilon)^{\log n} = O(1).
\]
If the current population of the process is $s$ then the probability that it reach size at least $\frac{c+1}{2} s$ in the next round is

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

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

Now there is a positive probability $\varepsilon$, say, that a single object spawns at least 100 descendants and so there is a probability $\varepsilon$ of at least

$$\varepsilon_1 \left(1 - \sum_{s = 100}^{\infty} e^{-\alpha s}\right)$$
That a single object spawns

\[(c+1)^{\log n} \geq \log^2 n\]

descendants at depth \(\log n\).

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

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

Claim 2
\( \Pr(\exists i : |S_i^+| \geq \log^2 n \text{ and } |T_i^-|) \)
We must now consider the probability that both $O^+(v)$ and $O^-(v)$ are large.

**Lemma 5**

$$P( |O^-(v)| \geq \log^2 n | |O^+(v)| \geq \log^2 n) = 1 - \frac{\alpha}{c} + o(1).$$

**Proof**

Expose $S_0^+, S_1^+, \ldots, S_{\rho}^+$ until either $S_{\rho}^+ = \emptyset$ or we see that $|T_k^+| \geq \log^2 n$.

Now let $S$ denote the set of edges/nodes defined by $S_0^+, S_1^+, \ldots, S_{\rho}^+$, we see that (see Lemma 2 [p.127])
Let \( C \) be the event that there are no edges from \( T_e \) to \( S_e \), where \( T_e \) is the set of vertices we reach through our BFS until \( v \), up to the point where we first find that \( |D^-(v)| < \log^2 n \) or \( \geq \log^2 n \). Then

\[
P(C) = 1 - \frac{1}{n^{1 - o(1)}}
\]

end

\[
P_e(1 \leq s_i \leq s_i, 0 \leq k \leq n') = \prod_{i=1}^{k} \left( \frac{s_{i-1}(n'-k_i)}{s_i} \right) \left( \frac{s_i}{n} \right) \left( 1 - \frac{s_{i-1}(n'-k_i) - s_i}{n} \right)
\]

where \( n' = n - |T_e^+| \).

Given this, we can prove a conditional version of Lemma 2 and continue as before.
We have now shown that if
\[ S = \{ v \in V : |D^+(v)|, |D^-(v)| > a \log n \} \]
then
\[ \mathbb{E}(1 \leq |S|) \leq (1 + o(1)) (1 - \frac{n}{c})^2 n. \]

We also claim that for any two vertices \( v, w \)
\[ Pr [ v, w \in S ] = (1 + o(1)) Pr (v \in S) Pr (w \in S) \] (3)

and therefore the Chebyshev inequality implies that when
\[ |S| \leq (1 + o(1)) (1 - \frac{n}{c})^2 n. \]
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

\[ \text{why } S \text{ is a strong component.} \tag{4} \]

(Any \( v \notin S \) is in a strong component of size \( \leq 2 \log n \)).
We prove (4) by arguing that

$$\Pr(\exists \nu, \omega \in S : \omega \in D^+(\nu)) = o(1). \quad (5)$$

For this we expose $S^+_0, S^+_1, \ldots, S^+_k$ until we find that $|T^+_k(\omega)| \geq n^{\frac{1}{2}} \log n$.

At the same time we expose $S^-_0, S^-_1, \ldots, S^-_k$ until $|T^-_k(\omega)| \geq n^{\frac{1}{2}} \log n$.

If $\omega \in D^+(\nu)$ then this experiment will have tried at least $\left( n^{\frac{1}{2}} \log n \right)^2$ lines to find an edge from $D^+(\nu)$ to $D^-(\omega)$ and failed everytime.
The probability is at most
\[(1 - \frac{c}{n})^n \log^2 n = o(n^{-2}).\]

This completes the proof of Theorem 2.
\[\square\]
Strong Connectivity Threshold

Here we prove

**Theorem 3**

Suppose that \( \rho = \frac{\log n + c_n}{n} \). Then

\[
\lim_{n \to \infty} \mathbb{P}(D_{n, \rho} \text{ is strongly connected}) = \begin{cases} 
0 & c_n \to -\infty \\
\exp(-2e^{-c}) & c_n \to c \\
1 & c_n \to +\infty
\end{cases}
\]

\[
= \lim_{n \to \infty} \mathbb{P}(1 \neq v \text{ such that } d^+(v) = 0 \iff d^-(v) = 0).
\]
Proof

We leave it as an exercise to prove that

$$\lim_{n \to \infty} P\left( \exists v \text{ such that } d^+(v) = 0 \lor d^-(v) = 0 \right) = \begin{cases} 1 & \text{if } c_n \to -\infty \\ 1 - e^{-2e^{c_n}} & \text{if } c_n \to c \\ 0 & \text{if } c_n \to \infty \end{cases}$$

Given this, one only has to show that if $c_n \to -\infty$ then why there does not exist a vertex $v$ such that $2 \leq |d^+(v)| \leq \eta_2$ or $2 \leq |d^-(v)| \leq \eta_2$. 
But, here with \( s+1 = 10^2(w) \),

\[
P(\mathcal{F}_w) \leq 2n \sum_{s=1}^{n/2} \left( \binom{n}{s} (s+1) \left( \frac{e}{n} \right)^s (1-p)^{s+1} (n-1-s) \right)
\]

\[= O(1). \quad (Exercise)\]
Hamilton Cycles

Here we prove the following remarkable inequality:

**Theorem 4**

\[ P_r(O_{n,p} \text{ is Hamiltonian}) \geq P_r(G_{n,p} \text{ is Hamiltonian}) \]

**Proof**

Remark: This shows that if \( p = \frac{\log n \cdot \log \log n + 60}{n} \), then \( O_{n,p} \) is Hamiltonian w.h.p. 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 \( \Gamma_0, \Gamma_1, \Gamma_2, \ldots, \Gamma_N \), \( N = \binom{2}{2} \) defined as follows:

Let \( e_1, e_2, \ldots, e_N \) 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 \( \Gamma_i \) 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 $\Gamma_0$ is just $G_{n,p}$ with each edge $(i,j)$ replaced by a pair of directed edges $(i, j)$, $(j, i)$, and $\Gamma_N = G_{n,p}$. Theorem 4 follows from

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

To prove this we condition on the existence of otherwise directed edges associated with $e_i, \ldots, e_{i+1}, e_{i+2}, \ldots, e_N$.

Let $C$ denote this conditioning.
Either $C$ is such that
(a) $C$ gives us a Hamilton cycle without any associated
with $C_i$ or there is no Hamilton cycle even if both
$\overline{C_i}, \overline{C}_i$ occur
or $C$ is such that:
(b) If a Hamilton cycle $y$ at least one of $\overrightarrow{C_i}, \overleftarrow{C}_i$
occurs.
In $\Gamma_{w_i}$ this happens with probability $p$
In $\Gamma_i$ this happens with probability $1 - (1-p)^2 > p$

[We will never require that both $\overline{C_i}, \overline{C}_i$ occur.]