8.2 Asymptotics of Ramsey number $R(3, t)$

$$R(s, t) := \min \{ n : \text{for every } G \text{ on } n \text{ vertices,}$$
$$\omega(G) \geq s \text{ or } \alpha(G) \geq t \}$$

EASY

$$R(s, t) = R(t, s), \quad R(2, t) = t$$

Greenwood & Gleason ('55):

$$R(3, 3) = 6, \quad R(3, 4) = 9,$$
$$R(3, 5) = 14, \quad R(4, 4) = 18$$

MORE:

$$R(3, 6) = 18, \quad R(3, 7) = 23, \quad R(3, 8) = 28,$$
$$R(3, 9) = 36, \quad R(4, 5) = 25,$$
• Ramsey number $R(3, t)$

Erdős (’61)

$$R(3, t) \geq c_1 \frac{t^2}{(\log t)^2},$$

Graver & Yackel (’68)

$$R(3, t) \leq c_2 \frac{t^2 \log \log t}{\log t}.$$

Ajtai, Komlós and Szemerédi (’81)

$$R(3, t) \leq c_3 \frac{t^2}{\log t}$$

(removed the “log log $t$” factor).
Little improvement on lower bound

Spencer (’77), Bollobás (’85),

Erdős, Suen and Winkler (’93)

Krivelevich (’94)

simplified its proof and/or increased constant

**Theorem (K ’95)**

\[ R(3, t) \geq c(1 - o(1)) \frac{t^2}{\log t}, \]

with \( c = 1/162 = 1/(2 \cdot 9^2). \)
Idea of the proof of

\[ R(3, t) \geq c(1 - o(1)) \frac{t^2}{\log t}. \]

Recall

\[ R(3, t) := \min\{n : \text{for every } G_n, \ \omega(G_n) \geq 3 \text{ or } \alpha(G_n) \geq t\} \]

Enough to show

\[ \exists \text{ triangle-free } G_n \text{ for which} \]

\[ \alpha(G_n) \leq 9 \sqrt{n \log n} \]

for sufficiently large \( n \).
• Random Greedy Methods vs. Nibble Methods

  · Random Greedy (or One-by-One) Construction

1. Randomly order all edges in $K_n$
   ($\exists \binom{n}{2}!$ possible ways)

2. Choose edges greedily according to the random order. (An edge cannot be chosen only if it makes a triangle with previously chosen edges.)
• Incremental Random Method (Nibble method)

Let

\[ \Gamma_0 : \text{the set of all } \binom{n}{2} \text{ edges.} \]

Define

a random subset \( X_1 \) of \( \Gamma_0 \):

\[ \Pr[e \in X_1] := \frac{\varepsilon}{\sqrt{n}} \text{ for all } e \in \Gamma_0 \]

independently.

Take any “maximal” (under \( \subseteq \)) family \( \mathcal{F}_1 \) of edge disjoint triangles in \( X_1 \).

Deleting all edges belong to triangles in \( \mathcal{F} \) we obtain a \( \Delta \)-free graph \( G_1 \) on \( n \) vertices.
An edge $e \in \Gamma_0$ survives if $e \notin X_1$ and there no edges $f, g \in Y_1 := X_1$ s.t. $efg$ is a triangle.

Let

$$\Gamma_1$$

be the set of all surviving edges.

Define

a random subset $X_2$ of $\Gamma_1$:

$$\Pr[e \in X_2] := \frac{\varepsilon}{\sqrt{n}}$$

for all $e \in \Gamma_1$ independently.
Forbidden pairs of edges:

\[ \Lambda_2 := \{ e_{uv}e_{vw} : e_{uv}, e_{vw} \subseteq X_2, e_{wu} \in Y_1 \} , \]

where \( e_{vw} := \{v, w\} \).

Forbidden triples of edges:

\[ \Delta_2 := \{ e_{uv}e_{vw}e_{wu} : e_{uv}, e_{vw}, e_{wu} \subseteq X_2 \} \]

Take any “maximal” (under \( \subseteq \)) family \( F_2 \) of edge disjoint forbidden pairs and triples in \( \Lambda_2 \cup \Delta_2 \).

Deleting all edges belong to pairs and triangles in \( F_2 \) we obtain a \( \Delta \)-free graph

\[ G_2 = G_1 \cup (X_2 \setminus \bigcup_{F \in F_2} F) \]

on \( n \) vertices.
An edge $e \in \Gamma_1$ survives
if $e \notin X_2$ and
there no edges $f, g \in Y_2 := Y_1 \cup X_2$ s.t.
$efg$ is a triangle.

Let

$\Gamma_2$ be the set of all surviving edges.

At step $i$, define

a random subset $X_i$ of $\Gamma_{i-1}$:

$$\Pr[e \in X_i] := \varepsilon/\sqrt{n} \quad \text{for all } e \in \Gamma_{i-1}$$

independently.

Forbidden pairs of edges:

$$\Lambda_i := \{e_{uv}e_{vw} : e_{uv}, e_{vw} \subseteq X_i, e_{wu} \in Y_{i-1}\}.$$
Forbidden triples of edges:

$$\Delta_i := \{e_{uv}e_{vw}e_{wu} : e_{uv}, e_{vw}, e_{wu} \subseteq X_i\}$$

Take any “maximal” (under \(\subseteq\)) family \(\mathcal{F}_i\) of edge disjoint forbidden pairs and triples in \(\Lambda_i \cup \Delta_i\).

Deleting all edges belong to pairs and triangles in \(\mathcal{F}_i\), we obtain a \(\Delta\)-free graph

$$G_i = G_{i-1} \cup (X_i \setminus \bigcup_{F \in \mathcal{F}_i} F).$$
An edge $e \in \Gamma_i$ survives
if $e \not\in X_i$ and
there no edges $f, g \in Y_i := Y_{i-1} \cup X_i$ s.t.
$efg$ is a triangle.

Let $\Gamma_{i+1}$ be the set of all surviving edges.

FACT: as $\varepsilon \rightarrow 0$

$$\frac{|\bigcup_{F \in \mathcal{F}_i} F|_i}{|X_i|} \rightarrow 0$$

So, small enough $\varepsilon \Rightarrow Y_i \approx \mathcal{E}(G_i)$

$$\varepsilon = (\log n)^{-2}, \ # \ of \ steps \ \approx n^{1/17}$$
\[ \mathcal{I}_i : \text{the collection of all independent sets in } G_i \text{ of size } t, \text{ i.e.,} \]

\[ \mathcal{I}_i = \{ T : |T| = t, \, T \text{ indepen. in } G_i \}, \]

where \( t := \lceil 9 \sqrt{n \log n} \rceil. \)

STOP when \( |\mathcal{I}_i| < 1. \)

possible ???

Let \( \Gamma_i(T) \) be the set of surviving edges in \( T \). Then WANT Prop. 7.

\[ |\Gamma_i(T)| \geq b_i \mu_i \left( \frac{t}{2} \right). \]
• Probabilities

Suppose $\exists (unknown) \psi$ satisfying

$$\Pr[e \in Y_i] \approx \Pr[e \in \mathcal{E}(G_i)] = \frac{\psi(i\varepsilon)}{\sqrt{n}},$$

for all $i$. Then

$$b_i := \Pr[e \in \Gamma_i] = ???$$

Consider a random graph $G(n, p)$ with edge prob. $p$, where

$$p = \frac{\psi(i\varepsilon)}{\sqrt{n}}.$$
Regarding \( G_i \) as \( G(n, p) \), \( p = \frac{\psi(i\epsilon)}{\sqrt{n}} \),

\[
\Pr[e \in \Gamma_i] = (1 - p^2)^{n-2} = \left(1 - \left(\frac{\psi(i\epsilon)}{\sqrt{n}}\right)^2\right)^{n-2} \approx \exp(-\psi^2(i\epsilon))
\]

On the other hand, we know

\[
|\mathcal{E}(G_{i+1})| \approx |\mathcal{E}(G_i)| + |X_{i+1}|
\]

and, in expectations,

\[
|\mathcal{E}(G_{i+1})| \approx \frac{\psi((i + 1)\epsilon)}{\sqrt{n}} \binom{n}{2} \approx \frac{\psi((i + 1)\epsilon)n^{3/2}}{2},
\]

\[
|\mathcal{E}(G_i)| \approx \frac{\psi(i\epsilon)}{\sqrt{n}} \binom{n}{2} \approx \frac{\psi(i\epsilon)n^{3/2}}{2}
\]
and

\[ |X_{i+1}| \approx \frac{\varepsilon}{\sqrt{n}} |\Gamma_i| \]

\[ \approx \frac{\varepsilon}{\sqrt{n}} \exp(-\psi^2(i\varepsilon)) \binom{n}{2} \]

\[ \approx \frac{\varepsilon \exp(-\psi^2(i\varepsilon)) n^{3/2}}{2}. \]

As \( \mathcal{E}(G_{i+1}) \approx \mathcal{E}(G_i) + |X_i| \), we have

\[ \frac{\psi((i + 1)\varepsilon)n^{3/2}}{2} \approx \frac{\psi(i\varepsilon)n^{3/2}}{2} + \frac{\varepsilon \exp(-\psi^2(i\varepsilon)) n^{3/2}}{2}. \]
In other words,

\[ \psi((i + 1)\epsilon) \approx \psi(i\epsilon) + \epsilon \exp(-\psi^2(i\epsilon)) \]

or

\[ \psi'(x) \approx \exp(-\psi^2(x)) . \]

Define the function \( \psi(x) \):

\[ x = \int_0^\psi(x) e^{\xi^2} \, d\xi . \]

Notice that

\[ \psi(x) \sim \sqrt{\log x} . \]
Let

\[ b_i := \psi'(i\varepsilon) = \exp(-\psi^2(i\varepsilon)), \]

\[ a_i := \sum_{j=0}^{i-1} b_j \varepsilon = \sum_{j=0}^{i-1} \psi'(j\varepsilon) \varepsilon = a_{i-1} + b_{i-1} \varepsilon \]

\[ \approx \psi(i\varepsilon) \]

where \( \varepsilon := (\log n)^{-2} \), and

\[ \mu_i := 1 - 20a_i \varepsilon - \frac{a_i}{3\sqrt{\log n}} \]

For \( A, B \subseteq V \), let

\[ \Gamma_i(A, B) := \{ e_{vw} \in \Gamma_i : v \in A, w \in B \} \text{ and } \Gamma_i(A) := \Gamma_i(A, A). \]
Prop. 1. \( d_{Y_i}(v) \leq a_i \sqrt{n} + in^{1/4}(\log n)^2 \).

Prop. 2. \( d_{\Gamma_i}(v) \leq b_i n \).

Prop. 3. \( |N_{Y_i}(v) \cap N_{Y_i}(w)| \leq 3i \log n \).

Prop. 4. \( d_{\Lambda_i}(e_{vw}, v) \leq b_i (a_i + 5\varepsilon) \sqrt{n} \).

Prop. 5. \( d_{\Delta_i}(e) \leq b_i^2 n \).

Prop. 6. For \( A \cap B = \emptyset \) with \( |A|, |B| \geq \varepsilon^2 b_i^2 \sqrt{n} \),

\[
|\Gamma_i(A, B)| \leq b_i |A||B| .
\]

\[
|\Gamma_i(A)| \leq b_i \binom{|A|}{2} .
\]
Property 7. For all $T \subseteq V$ with $|T| = 9\sqrt{n \log n}$,

$$|\Gamma_i(T)| \geq b_i \mu_i \left(\binom{t}{2}\right).$$

Let $\mathcal{I}_i$ be the set of independent sets of size $9\sqrt{n \log n}$ in $G_i$.

Property 8.

$$|\mathcal{I}_i| \leq n^i \binom{n}{t} \exp \left( - (1 - \varepsilon) \sum_{j=0}^{i-1} \frac{b_j \mu_j \varepsilon}{\sqrt{n}} \binom{t}{2} \right),$$
Definitions

For given \((Y_i, \Gamma_i, G_i)\), set

\[
\Lambda_i := \{ef \subseteq \Gamma_i : \exists g \in Y_i \text{s.t.} efg \in \Delta_0\}
\]

\[
\Delta_i := \{efg \subseteq \Gamma_i : efg \in \Delta_0\},
\]

and

\[
N_{Y_i}(v) := \{w \in V : e_{vw} \in Y_i\}, \quad d_{Y_i}(v) := |N_{Y_i}(v)|.
\]

Given \(v \in V\), let

\[
N_{\Gamma_i}(v) := \{w \in V : e_{vw} \in \Gamma_i\}
\]

\[
\mathcal{M}_{\Gamma_i}(v) := \{e_{vw} : e_{vw} \in \Gamma_i\}
\]

\[
d_{\Gamma_i}(v) := |N_{\Gamma_i}(v)| = |\mathcal{M}_{\Gamma_i}(v)|.
\]
Also, for $e_{vw} \in \Gamma_i$,

\[
N_{\Lambda_i}(e_{vw}, v) := \{u \in V : e_{uv}e_{vw} \in \Lambda_i\}
\]

\[
M_{\Lambda_i}(e_{vw}, v) := \{e_{uv} \in \Gamma_i : e_{uv}e_{vw} \in \Lambda_i\}
\]

\[
d_{\Lambda_i}(e_{vw}, v) := |N_{\Lambda_i}(e_{vw}, v)| = |M_{\Lambda_i}(e_{vw}, v)| ,
\]

and

\[
N_{\Lambda_i}(e_{vw}) := N_{\Lambda_i}(e_{vw}, v) \cup N_{\Lambda_i}(e_{vw}, w)
\]

\[
M_{\Lambda_i}(e_{vw}) := M_{\Lambda_i}(e_{vw}, v) \cup M_{\Lambda_i}(e_{vw}, w) .
\]

Finally,

\[
N_{\Delta_i}(e_{vw}) := \{u \in V : e_{uv}e_{vw}e_{wu} \subseteq \Gamma_i\, \} .
\]
All properties would seem quite natural to expect.

For example, we would expect

\[
d_{\Lambda_i}(e_{vw}, v) := \sum_{u \in V \setminus e_{vw}} 1(e_{wu} \in Y_i \text{ and } e_{uv} \in \Gamma_i)
\]

\[
\approx n \Pr[e_{wu} \in Y_i \text{ and } e_{uv} \in \Gamma_i]
\]

\[
\approx n \Pr[e_{wu} \in Y_i] \Pr[e_{wu} \in \Gamma_i]
\]

\[
\approx n \left( \frac{a_i}{\sqrt{n}} \right) b_i = a_i b_i \sqrt{n}
\]
HOW TO PROVE

Prop. 1. \(d_{Y_i}(v) \leq a_i \sqrt{n} + in^{1/4}(\log n)^2\).

Prop. 2. \(d_{\Gamma_i}(v) \leq b_i n\).

Prop. 3. \(|N_{Y_i}(v) \cap N_{Y_i}(w)| \leq 3i \log n\).

......................

(a) Show the properties at the level of expectations.

(b) Prove that the random variables \(d_{Y_i}(v)\) etc. are highly concentrated near their means. For example,

\[
\Pr \left[ d_{Y_i}(v) E[d_{Y_i}(v)] \geq \delta E[d_{Y_i}(v)] \right] \leq e^{-\left(\log n\right)^2}.
\]

For (b), we need martingale inequalities.
Except Property 7:
For all $T \subseteq V$ with $|T| = 9\sqrt{n\log n}$,

$$|\Gamma_i(T)| \geq b_i \mu_i \left( \frac{t}{2} \right).$$

**Lemma 8.1** The following three conditions hold simultaneously with probability at least $1 - 3/n^2$:

(i) For all $v \in V$, $|N_{X_{i+1}}(v)| \leq b_i \varepsilon \sqrt{n} + n^{1/4} \log n$;

(ii) For all $v \neq w \in V$, $|N_{G_i}(v) \cap N_{X_{i+1}}(w)| \leq \log n$;

(iii) For all $v \neq w \in V$, $|N_{X_{i+1}}(v) \cap N_{X_{i+1}}(w)| \leq \log n$. 
Remark

1. A better constant could be possible:

Setting \( p_i := \frac{\varepsilon}{b_i \sqrt{n}} \).

BUT “More Complicated”.

2. \( R(4, t) =?? \)

- Probably too many properties
- NOT enough independence:
  - HARD to guess the parameters
- ONLY \((\log t)^x\) improvement:

\[
t^{5/2} \lesssim R(4, t) \lesssim t^3
\]
Giant Component of Random graph & 2-SATisfiability Problem
• Random Graph $G(n, p)$:

  each of $\binom{n}{2}$ edges is independently
  in $G(n, p)$ with probability $p$

  $p = 1$: complete graph $p = 0$: empty graph

Expected number of edges

\[
p\left(\binom{n}{2}\right)
\]

For fixed $G = (V, E)$,

\[
\Pr[G(n, p) = G] = p^{|E|}(1 - p)^{\binom{n}{2} - |E|}
\]
$W_i$: the size of the $i^{th}$ largest component of $G(n,p)$

- Erdős & Rényi (’60, ’61)

\[
W_1 \begin{cases} 
\leq c \log n, & p = (1 - \epsilon)/n \\
= \Theta(n^{2/3}), & p \sim 1/n \\
\sim f(\epsilon)n, & p = (1 + \epsilon)/n 
\end{cases}
\]

($\epsilon > 0$), where $f(\epsilon)$ is the positive sol. of

\[
1 - f(\epsilon) = \exp(-(1 + \epsilon)f(\epsilon))
\]

If $\epsilon$ is small,

\[
f(\epsilon) \sim 2\epsilon
\]
What if

\[ p = \frac{1 \pm n^{-\delta}}{n} \]

Bollobás ('84), Łuczak ('90),
Janson, Knuth, Łuczak & Pittel ('94)

For \( p = \frac{1-\lambda n^{-1/3}}{n} \)

\[ W_1 = \Theta(n^{2/3} \log \lambda/\lambda^2) \]
\[ W_2 = \Theta(n^{2/3} \log \lambda/\lambda^2), \]

particularly

\[ \frac{W_1}{n^{2/3}} \to 0, \quad \text{as } \lambda \to \infty \]

For \( p = \frac{1+\lambda n^{-1/3}}{n} \)

\[ W_1 \approx 2\lambda n^{2/3} \]
and hence

\[ W_2 = \Theta(n^{2/3} \log \lambda/\lambda^2) \]

\[ \frac{W_1}{n^{2/3}} \to \infty, \quad \text{as } \lambda \to \infty \]

\[ \frac{W_2}{n^{2/3}} \to 0, \quad \text{as } \lambda \to \infty \]
• Random Cluster Model $H = H(n, p)$:

For fixed $G = (V, E)$,

$$\Pr[G(n, p) = G] \propto p^{|E|}(1 - p)^{\binom{n}{2} - |E|} q^{c(G)}$$

where $q > 0$ and

$$c(G) = \# \text{ connected components of } G$$

(cf. $\Pr[G(n, p) = G] = p^{|E|}(1 - p)^{\binom{n}{2} - |E|}$)
Potts model on $K_n$:

configuration $\sigma : V \rightarrow \{1, 2, ..., q\}$

$$w(\sigma) = \exp \left( \beta \sum_{\substack{i, j \in V \\ i \neq j}} \delta(\sigma(i), \sigma(j)) \right)$$

where

$$\delta(\sigma(i), \sigma(j)) = \begin{cases} 
1 & \text{if } \sigma(i) = \sigma(j) \\
0 & \text{otherwise}
\end{cases}$$
Partition function

\[ Z(\beta) = \sum_{\sigma} w(\sigma) \]

\[ = \sum_{\sigma} \exp \left( \beta \sum_{i, j \in V} \delta(\sigma(i), \sigma(j)) \right) \]
FK (Fortuin & Kasteleyn) Representation:

\[ Z(\beta) = \sum_{\sigma} \prod_{\substack{i,j \in V \atop i \neq j}} e^{\beta \delta(\sigma(i),\sigma(j))} \]

\[ = \sum_{\sigma} \prod_{\substack{i,j \in V \atop i \neq j}} (1 + (e^{\beta \delta(\sigma(i),\sigma(j))} - 1)) \]

\[ = \sum E \sum_{\sigma} \prod_{\substack{i,j \in E \atop i \neq j}} (e^{\beta \delta(\sigma(i),\sigma(j))} - 1) \]

\[ = \sum E \sum_{\sigma} \prod_{\substack{i,j \in E \atop i \neq j}} (e^{\beta \delta(\sigma(i),\sigma(j))} - 1) \]

For \( G = (V, E) \),

\[ \sum_{\sigma} \prod_{\substack{i,j \in E \atop i \neq j}} (e^{\beta \delta(\sigma(i),\sigma(j))} - 1) = (e^\beta - 1)|E| q^c(G) \]
\[ \exists \alpha_c(q) \text{ s.t.} \]
\[
\frac{W_1}{n} \rightarrow \begin{cases} 
0 & \text{if } p \leq (\alpha_c(q) - \varepsilon)/n \\
\frac{f(\varepsilon, q)}{n} & \text{if } p = (\alpha_c(q) + \varepsilon)/n 
\end{cases}
\]

where \(f(\varepsilon, q)\) is the positive sol. of

\[
\frac{1 - f(\varepsilon, q)}{1 + (q - 1)f(\varepsilon, q)} = \exp(-(\alpha_c(q) + \varepsilon)f(\varepsilon, q))
\]

In fact,

\[
\alpha_c(q) = \begin{cases} 
q & 0 < q \leq 2 \\
\frac{2(q-1)\log(q-1)}{q-2} & \text{if } q > 2 
\end{cases}
\]
• Satisfiability

Boolean Variables: $x_1, ..., x_n \in \{0, 1\}$

Negation of $x$: $\bar{x} = 1 - x$

$2n$ literals: $x_1, \bar{x}_1, ..., x_n, \bar{x}_n$

$x$ and $y$ are strictly distinct (s.d.) if $x \neq y$ and $x \neq \bar{y}$

$k$-clause:

$$C = l_1 \lor \cdots \lor l_k$$

where $l_1, ..., l_k$ are s.d. literals

How many $k$-clauses??
Take $k$ Boolean variables out of $n$.

Then $\exists$ two choices (negation or not) for each variable.

$$2^k \binom{n}{k}$$
\( k \)-SAT Formula:

\[
F = F(x_1, \ldots, x_n) = C_1 \land \cdots \land C_m
\]

where \( C_1, \ldots, C_m \) are \( k \)-clauses.

\( F \) is \textit{satisfiable} if

\[
F(x_1, \ldots, x_n) = 1
\]

for some \( x_1, \ldots, x_n \in \{0, 1\} \)

\( k \)-SAT problem: NP-Complete if \( k \geq 3 \)

\((\text{P if } k = 2)\)
• Random $k$-SAT $F_k(n, p)$:

Each $k$-clause appears in $F$
with probability $p$

Expected # of clauses

$$m = 2^k p \binom{n}{k}$$

(Goerdt ’92, Chvátal & Reed ’92, F. de la Vega ’92) For $k = 2$,

$$\Pr[F_2 \text{ is SAT }] \rightarrow \begin{cases} 1 & \text{if } m/n \to c < 1 \\ 0 & \text{if } m/n \to c > 1 \end{cases}$$
Conjecture. For \( k \geq 3 \), \( \exists \alpha(k) \) s.t.

\[
\Pr[F_k \text{ is SAT }] \rightarrow \begin{cases} 
1 & \text{if } m/n \rightarrow c < \alpha(k) \\
0 & \text{if } m/n \rightarrow c > \alpha(k)
\end{cases}
\]

Known

\[3.14 \leq \alpha(3) \leq 4.596\]

\[c_1 2^k / k \leq \alpha(k) \leq c_2 2^k\]

Pittel:

“Y2K Problem”
Friedgut (’97) Let

\[ p_n^{-}(\delta) = \max \{ p : \Pr[F_k(n, p) \text{ is SAT }] \geq 1 - \delta \} \]

and

\[ p_n^{+}(\delta) = \min \{ p : \Pr[F_k(n, p) \text{ is SAT }] \leq \delta \} \]

Then

\[ (p_n^{+}(\delta) - p_n^{-}(\delta))2^k \binom{n}{k} = o(1/n) \]
For $k = 2$, 

$$\Pr[F_2 \text{ is SAT }] \to \begin{cases} 
1 & \text{if } m/n \to c < 1 \\
0 & \text{if } m/n \to c > 1 
\end{cases}$$

What if 

$$\frac{m}{n} = 1 \pm n^{-\delta}$$

(Bollobás, Borgs, Chayes, K, Wilson) For $m/n = 1 - \lambda n^{-1/3}$ 

$$\Pr[F_2 \text{ is SAT }] = 1 - \Theta(1/\lambda^3)$$

For $m/n = 1 + \lambda n^{-1/3}$ 

$$\Pr[F_2 \text{ is SAT }] = \exp(-\Theta(\lambda^3))$$
Spine $S_F$:

A satisfiable formula $F$ fixes a literal $x$ if

$x$ is true (i.e. $x = 1$) in all satisfying assignments.

A literal $x$ is the spine $S_F$ of a formula $F$ iff

$\exists$ a satisfiable subformula which fixes $x$. 
For $m/n = 1 - \lambda n^{-1/3}$

$$\frac{E[|S_F|]}{n^{2/3}} \sim \frac{1}{2\lambda^2}$$

For $m/n = 1 + \lambda n^{-1/3}$

$$\frac{E[|S_F|]}{n^{2/3}} \sim 4\lambda$$
9 Branching Process and Giant Component

$G(n, p)$ undergoes a remarkable change at $p = 1/n$. (Erdős and Rényi, 1960)

- $p = c/n$ with $c < 1$
  - consists of small components, the largest of which is of size $\Theta(\ln n)$.
- $p = c/n$ with $c > 1$
  - forms a “giant component” of size $\Theta(n)$. 

9.1 Branching Process

Imagine the following stochastic process called branching process.

- A unisexual universe
- Initially there is one live organism and no dead ones.
- At each time unit, we select one of the live organisms, it has $Z$ children, and then it dies.
- $Z$ will be Poisson with mean $c$.

We want to study whether or not the process continues forever.
More precisely,

- Let $Z_i$ be the number of children of the organism selected at time $i$.
  - $Z_1, Z_2, \ldots$ be independent random variables, each with distribution $Z$.

- Let $Y_i$ be the number of live organisms at time $i$. Then, $Y_0, Y_1, \ldots$ is given by the recursion

\[
Y_0 = 1, \\
Y_i = Y_{i-1} + Z_i - 1,
\]

for $i \geq 1$. 

• Let $T$ be the least $t$ such that $Y_t = 0$. If no such $t$ exists, we say $T = +\infty$.

• $T$ is the total number of organisms in the process.

• The process stops when $Y_t = 0$ but we define the recursion for all $t$. 

Theorem  When $E[Z] = c < 1$, the process dies out ($T < \infty$) with probability 1.

Proof.
• Since $Y_t = Z_1 + \cdots + Z_t - t + 1$,
  \[ \Pr[T > t] \leq \Pr[Y_t > 0] = \Pr[Z_1 + \cdots + Z_t \geq t]. \]
• $Z_1 + \cdots + Z_t$ has a Poisson distribution with mean $ct$. Then,
  \[ \Pr[Z_1 + \cdots + Z_t \geq t] \leq (ce^{1-c}t). \]
• From the fact that $ce^{1-c} < 1$ for $c < 1$,
  \[ \lim_{t \to \infty} \Pr[T > t] = 0, \]
  which means that $\Pr[T = \infty] = 0$. 
\[ \square \]
Theorem  When \( E[Z] = c > 1 \), there is a nonzero probability that the process goes on forever \((T = \infty)\).

Proof.  
• As in the proof of the previous theorem,  
  \[
  \Pr[Z_1 + \cdots + Z_t \leq t] \leq (1 - \delta)^t,
  \]
  with \( \delta > 0 \).
• As \( \sum_{t=1}^{\infty} (1 - \delta)^t \) converges, there is a \( t_0 \) with  
  \[
  \sum_{t=t_0}^{\infty} \Pr[Z_1 + \cdots + Z_t \leq t] < 1.
  \]
• Then, conditioned on \( Z_1 = t_0 \),  
  \[
  Y_t = t_0 + (Z_2 - 1) + \cdots + (Z_t - 1), \quad \text{for } t \geq 2,
  \]
and so
\[
\sum_{t=2}^{\infty} \Pr[Y_t \leq 0|Z_1 = t_0] = \sum_{t=0}^{\infty} \Pr[t_0 + Z_2 + \cdots Z_t \leq t - 1] \\
\leq \sum_{t=t_0+1}^{\infty} \Pr[Z_2 + \cdots Z_t \leq t - 1] < 1.
\]

Therefore,
\[
\Pr[T = \infty] \geq \Pr[Z_1 = t_0] \left( 1 - \sum_{t=t_0}^{\infty} \Pr[Z_1 + \cdots Z_t \leq t] \right) > 0.
\]
\[\square\]
Analysis using generating functions

Let

\[ p_i = \Pr[Z_1 = i] = e^{-c} \frac{c^i}{i!} \]

and define the generating function

\[ p(x) = \sum_{i=0}^{\infty} p_i x^i = \sum_{i=0}^{\infty} e^{-c} \frac{c^i}{i!} x^i = e^{c(x-1)}. \]
• Let \( q_i = \Pr[T = i] \) and set

\[
q(x) = \sum_{i=0}^{\infty} q_i x^i.
\]

• Conditioning on the first organism having \( s \) children, the generating function for the total number of offspring is

\[
\sum_{i=0}^{\infty} \Pr[T = i \mid Z_1 = s] x^i = \sum_{i=0}^{\infty} \sum_{j_1 + \cdots + j_s = i-1} q_{j_1} \cdots q_{j_s} x^i
\]

\[
= x \sum_{j=0}^{\infty} \sum_{j_1 + \cdots + j_s = j} q_{j_1} \cdots q_{j_s} x^j
\]

\[
= x(q(x))^s.
\]
Hence

\[
q(x) = \sum_{i=0}^{\infty} q_i x^i
\]

\[
= \sum_{i=0}^{\infty} \Pr[T = i] x^i
\]

\[
= \sum_{i=0}^{\infty} \sum_{s=0}^{\infty} \Pr[Z_1 = s] \Pr[T = i | Z_1 = s] x^i
\]

\[
= \sum_{s=0}^{\infty} \Pr[Z_1 = s] \sum_{i=0}^{\infty} \Pr[T = i | Z_1 = s] x^i
\]

\[
= \sum_{s=0}^{\infty} p_s x q(x)^s
\]

\[
= x \sum_{s=0}^{\infty} p_s q(x)^s = xp(q(x)).
\]
• $y_x = q(x)/x$ satisfies the functional equality $y_x = p(xy_x)$, i.e.,

$$y_x = e^{c(x y_x - 1)}.$$  

• The extinction probability

$$y := \Pr[T < \infty] = \sum_{i=0}^{\infty} \Pr[T = i] = \sum_{i=0}^{\infty} q_i = q(1) = q(1)/1 = y_1$$

must satisfy

$$y = e^{c(y - 1)}.$$
• For $c < 1$, $y = e^{c(y-1)}$ has the unique solution $y = 1$, corresponding to the certain extinction.

• For $c > 1$, there are two solutions, $y = 1$ and $y = y^* \in (0, 1)$.

• As $\Pr[T < \infty] < 1$, $\Pr[T < \infty] = y^*$. 
• When a branching process dies, we call $H = (Z_1, \ldots, Z_T)$ the history of the process.

• A sequence $(z_1, \ldots, z_t)$ is a possible history if and only if the sequence $y_i$ given by $y_0 = 1$, $y_i = y_{i-1} + z_i - 1$ has $y_i > 0$ for $0 \leq i < t$ and $y_t = 0$.

• When $Z$ is Poisson with mean $\lambda$,

$$\Pr[H = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \frac{e^{-\lambda} \lambda^{z_i}}{z_i!} = \frac{e^{-\lambda} (\lambda e^{-\lambda})^{t-1}}{\prod_{i=1}^{t} z_i!},$$

since $z_1 + \cdots + z_t = t - 1$. 
• We call $d < 1 < c$ a *conjugate pair* if

$$de^{-d} = ce^{-c}.$$  

• Since $y^* = e^c(y^*-1)$,

$$(cy^*)e^{-cy^*} = ce^{-c},$$

so $cy^*$ and $c$ is a conjugate pair.
• For every history $H = (z_1, \ldots, z_t)$,

$$\Pr_c[H = (z_1, \ldots, z_t) | T < \infty] = \frac{e^{-c}(ce^{-c})^{t-1}}{y^* \prod_{i=1}^{t} z_i!} \frac{e^{-cy^*}(cy^* e^{-cy^*})^{t-1}}{\prod_{i=1}^{t} z_i!} = \Pr_d[H = (z_1, \ldots, z_t)],$$

since $ce^{-c} = (cy^*)e^{-cy^*}$ and $y^* e^{-cy^*} = e^{-c}$.

**Theorem** The branching process with mean $c$, conditional on extinction, has the same distribution as the branching process with mean $d = cy^*$. 
9.2 Giant Component

We define a procedure to find the component $C(v)$ containing a given vertex $v$ in a graph $G = G(n, p)$.

- Vertices will be live, dead, or neutral.
- Originally $v$ is live, all other vertices are neutral, and time $t = 0$.
- Each time $t$, take a live vertex $w$ and check the pairs $\{w, w'\}$ for neutral $w'$:
  - if $\{w, w'\} \in E$, make $w'$ live.
  - otherwise, leave it neutral.

Then, set $w$ dead.

- When there are no live vertices, the process terminates.
  - $C(v)$ is the set of dead vertices.
• Let $Z_t$ be the number of $w'$ with $\{w,w'\} \in E$ at time $t$, and $Y_t$ be the number of live vertices at time $t$. Then,

$$
Y_0 = 1, \\
Y_t = Y_{t-1} + Z_t - 1.
$$

• Since no pair $\{w, w'\}$ is ever examined twice,

$$Z_t \sim Bin[n - (t - 1) - Y_{t-1}, p].$$

• Let $T$ be the least $t$ for which $Y_t = 0$. Then, $T = |C(v)|$.

• We recursively define $Y_t$ for all $0 \leq t \leq n$. 
Lemma For all \( t \),

\[ Y_t \sim Bin[n - 1, 1 - (1 - p)^t] + 1 - t. \]

Proof.
- Let \( N_t = n - t - Y_t \) be the number of neutral vertices at time \( t \).
- Note that
  \[ N_t \sim Bin[n - 1, (1 - p)^t]. \]
- Then,
  \[
  Y_t = n - 1 - N_t + 1 - t \\
  \sim Bin[n - 1, 1 - (1 - p)^t] + 1 - t.
  \]
• Set $p = c/n$.

• For fixed $c$,
  - $Y^*, Z^*, T^*, H^*$: Poisson branching process with mean $c$
  - $Y_t, Z_t, T, H$: random graph process with $G(n, c/n)$

• For any history $(z_1, \ldots, z_t)$,

\[
\Pr[H^* = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \Pr[Z^* = z_i],
\]

where $Z^*$ is Poisson with mean $c$ while

\[
\Pr[H = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \Pr[Z_i = z_i],
\]

where $Z_i \sim Bin[n - 1 - z_1 - \cdots - z_{i-1}, c/n]$. 
For $m = m(n) = n + o(n^{1/4})$ and $z = o(n^{1/4})$,

$$\Pr[\text{Bin}[m, c/n] = z] = \binom{m}{z} \left(\frac{c}{n}\right)^z (1 - \frac{c}{n})^{m-z} = (1 + o(n^{-1/2})) \frac{e^{-c} c^z}{z!}$$

(uniformly).

Hence, for $H = (z_1, \ldots, z_t)$ with $\sum_{i=1}^t z_i = o(n^{1/4})$,

$$\Pr[H = (z_1, \ldots, z_t)] = (1 + o(n^{-1/4})) \Pr[H^* = (z_1, \ldots, z_t)]$$

(uniformly), and so

$$\Pr[T = t] = (1 + o(n^{-1/4})) \Pr[T^* = t],$$

for $t = o(n^{1/4})$. 
Theorem For $c < 1$, $G(n, \frac{c}{n})$ almost always has components all of which have size $O(\ln n)$.

Proof.

• Since $Y_t \sim Bin[n - 1, 1 - (1 - p)^t] + 1 - t$ and $1 - (1 - p)^t \leq tp$,

\[
\Pr[T > t] \leq \Pr[Y_t > 0] = \Pr[Bin[n - 1, 1 - (1 - p)^t] \geq t] \leq \Pr[Bin[n, tc/n] \geq t].
\]

• By (generalized) Chernoff bound,

\[
\Pr[T > t] \leq \Pr[Bin[n, tc/n] \geq t] \leq e^{-\frac{(1-c)^2t^2}{2ct} + \frac{(1-c)^3t^3}{2c^3t^3}} \leq c_1 e^{-c_2 t}
\]

for some constants $c_1, c_2 > 0$. 
• Choose $c_3$ satisfying $c_2c_3 > 1$. Then,

\[ \Pr[T > c_3 \ln n] \leq c_1 e^{-c_2c_3 \ln n} = c_1 n^{-c_2c_3} = o(n^{-1}). \]

• Since there are $n$ choices for initial vertex $v$,

\[ \Pr[\exists v \text{ such that } |C(v)| > c_3 \ln n] \leq n \cdot o(n^{-1}) = o(1). \]
Theorem For $c > 1$, $G(n, \frac{c}{n})$ almost always has a giant component of size $\sim (1 - y)n$ and all other components of size $O(\ln n)$.

Proof.

- Let $t_0 = K \ln n$ for a large constant $K$.

- First, we prove the following fact.

Claim. Let $\varepsilon, \delta > 0$ be arbitrarily small. Then,

\[
y - \varepsilon \leq \Pr[T \leq t_0] \leq y + \varepsilon,
\]

and

\[
1 - y - \varepsilon \leq \Pr[(1 - \delta)(1 - y)n < T < (1 + \delta)(1 - y)n] \leq 1 - y + \varepsilon,
\]

for sufficiently large $n$. 

Proof of Claim.

• Since $\Pr[T = t] = (1 + o(n^{-1/4})) \Pr[T^* = t]$ (uniformly) for $t \leq t_0$ and $\sum_{t=1}^{\infty} \Pr[T^* = t] = y$, there is $N_1 > 0$ such that

$$y - \varepsilon \leq \Pr[T \leq t_0] \leq y + \varepsilon$$

for $n \geq N_1$. 
• Note that $Y_t \sim Bin[n - 1, 1 - (1 - p)^t] + 1 - t$.

• Let $X_t \sim Bin[n - 1, 1 - (1 - p)^t]$.

• For $t = (1 + \delta)(1 - y)n = \alpha n$,

\[
Pr[T \geq \alpha n] \leq Pr[Y_{\alpha n} \geq 0] = Pr[X_{\alpha n} \geq \alpha n - 1].
\]
• From \((1 - x)y = e^{-xy + O(yx^2)}\),

\[
1 - (1 - p)^{\alpha n} = 1 - (1 - \frac{c}{n})^{\alpha n} = 1 - e^{-c\alpha + O(\frac{1}{n})}.
\]

• Since \(\alpha > 1 - e^{-c\alpha}\) for \(\alpha > 1 - y\), by Chernoff bound,

\[
\Pr[T \geq \alpha n] \leq \Pr[X_{\alpha n} \geq \alpha n - 1] \\
\leq \exp\left(-\frac{((\alpha - 1 + e^{-c\alpha - O(\frac{1}{n})})n - 1)^2}{n}\right) \\
\leq e^{-c_1 n}.
\]

for some constant \(c_1 > 0\).

• Hence, we may choose \(N_2\) such that

\[
\Pr[T \geq (1 + \delta)(1 - y)n] \leq \varepsilon
\]

for \(n \geq N_2\).
• For $t = \alpha n$ with $\frac{\ln^2 n}{n} \leq \alpha \leq (1 - \delta)(1 - y)$,

$$
\Pr[Y_{\alpha n} \leq 0] \leq \Pr[X_{\alpha n} \leq \alpha n]
\leq \exp\left(-\frac{\Theta((\alpha - 1 + e^{-c\alpha - O(\frac{1}{n})})^2 n^2)}{2(1 - e^{-c\alpha - O(\frac{1}{n})})n}\right)
\leq \exp\left(-\frac{c_2(\alpha - 1 + e^{-c\alpha - O(\frac{1}{n})})^2 n}{2(1 - e^{-c\alpha - O(\frac{1}{n})})}\right)
$$

for some constant $c_2 > 0$ by Chernoff bound.

• Since, for $0 \leq \alpha \leq (1 - \delta)(1 - y)$,

$$
\alpha - 1 + e^{-c\alpha} \leq \frac{(1 - \delta)(1 - y) - 1 + e^{-c(1-\delta)(1-y)}}{(1 - \delta)(1 - y)}\alpha \leq 0,
$$

we may choose $c_3 > 0$ such that

$$
c_2(\alpha - 1 + e^{-c\alpha - O(\frac{1}{n})})^2 \geq c_3 \alpha^2.
$$
• For $\alpha \geq 0$

\[
(1 - e^{-c\alpha}) \leq (1 - e^{-c\alpha})'\alpha,
\]

so we may choose $c_4 > 0$ such that

\[
2(1 - e^{-c\alpha - O(\frac{1}{n})}) \leq c_4\alpha.
\]

• Set $c_5 = \frac{c_3}{c_4} > 0$, then

\[
\frac{c_2(\alpha - 1 + e^{-c\alpha - O(\frac{1}{n})})^2}{2(1 - e^{-c\alpha - O(\frac{1}{n})})} \geq c_5\alpha \geq c_5 \frac{\ln^2 n}{n}.
\]
• From the above,
  \[ \Pr[Y_{\alpha n} \leq 0] \leq e^{-c_5 K \ln n} = O(n^{-2}), \]
  for sufficiently large \( K \), and so
  \[ \Pr[t_0 \leq T \leq (1 - \delta)(1 - y)n] \leq \Pr[\bigcup_{\alpha} Y_{\alpha n} \leq 0] = O(1/n), \]
  where \( \frac{K \ln n}{n} \leq \alpha \leq (1 - \delta)(1 - y) \) in the union.

• Hence, we may choose \( N_3 \) such that
  \[ \Pr[t_0 \leq T \leq (1 - \delta)(1 - y)n] \leq \varepsilon \]
  for \( n \geq N_3 \).
Therefore, if we let $N = \max\{N_1, N_2, N_3\}$,

$$y - \varepsilon \leq \Pr[T \leq t_0] \leq y + \varepsilon,$$

and

$$1 - y - \varepsilon \leq \Pr[(1 - \delta)(1 - y)n < T < (1 + \delta)(1 - y)n] \leq 1 - y + \varepsilon,$$

for $n \geq N$. 

□
• Start with $G \sim G(n, p)$, select $v = v_1 \in G$, and compute $C(v_1)$.

• Then delete $C(v_1)$, pick $v_2 \in G - C(v_1)$, and iterate.

• Note that, at each stage, the remaining graph has distribution $G(m, p)$ where $m$ is the number of vertices.

• Let $\varepsilon, \delta > 0$ be arbitrarily small.

• Call a component $C(v)$

\[
\begin{cases}
\text{small} & \text{if } |C(v)| \leq t_0, \\
\text{giant} & \text{if } (1 - \delta)(1 - y) < |C(v)| < (1 + \delta)(1 - y), \\
\text{failure} & \text{otherwise}.
\end{cases}
\]
• Let $s = \frac{\ln \varepsilon}{\ln(y + 2\varepsilon)}$. Then,

$$(y + \varepsilon)^s < (y + \varepsilon)^{\frac{\ln \varepsilon}{\ln(y + \varepsilon)}} = e^{\ln(y + \varepsilon) \frac{\ln \varepsilon}{\ln(y + \varepsilon)}} = e^{\ln \varepsilon} = \varepsilon.$$

• Begin the procedure with the full graph and terminate it when
  – a giant component is found,
  – a failure component is found,
  – or $s$ small components are found.

• At each stage, the number of remaining vertices is
  $m = n - O(\ln^2 n) \sim n$.
  – the cond. prob.’s of small, giant, and failure remain asymptotically the same.
• The prob. that the procedure terminates without a giant component is at most

\[ \varepsilon + (y + \varepsilon)\varepsilon + \cdots + (y + \varepsilon)^{s-1}\varepsilon + (y + \varepsilon)^{s} \leq s\varepsilon + \varepsilon = (s + 1)\varepsilon, \]

because \((y + \varepsilon)^{s} < \varepsilon\).

• Since \(\varepsilon \ln \varepsilon \to 0\) as \(\varepsilon \to 0\),

\[ (s + 1)\varepsilon = \left( \frac{\ln \varepsilon}{\ln(y + 2\varepsilon)} + 1 \right)\varepsilon \to 0 \]

as \(\varepsilon \to 0\), so \((s + 1)\varepsilon\) may be made arbitrarily small.

• Hence, we find a giant component with prob. at least

\[ 1 - (s + 1)\varepsilon. \]
• The remaining graph has \( m \sim yn \) vertices.
• Then, \( G(m, p) = G(m, \frac{c}{n}) \sim G(m, \frac{cy}{m}) \).
• As \( cy = d < 1 \), the maximum component size of the remaining graph is \( O(\ln n) \).
Homework 1: Exercises in pages 31, 40, 43, 52, 88, 108, 110, 112, 126 (Due 2/2/07)

List of Papers


