# Connectivity of random graphs

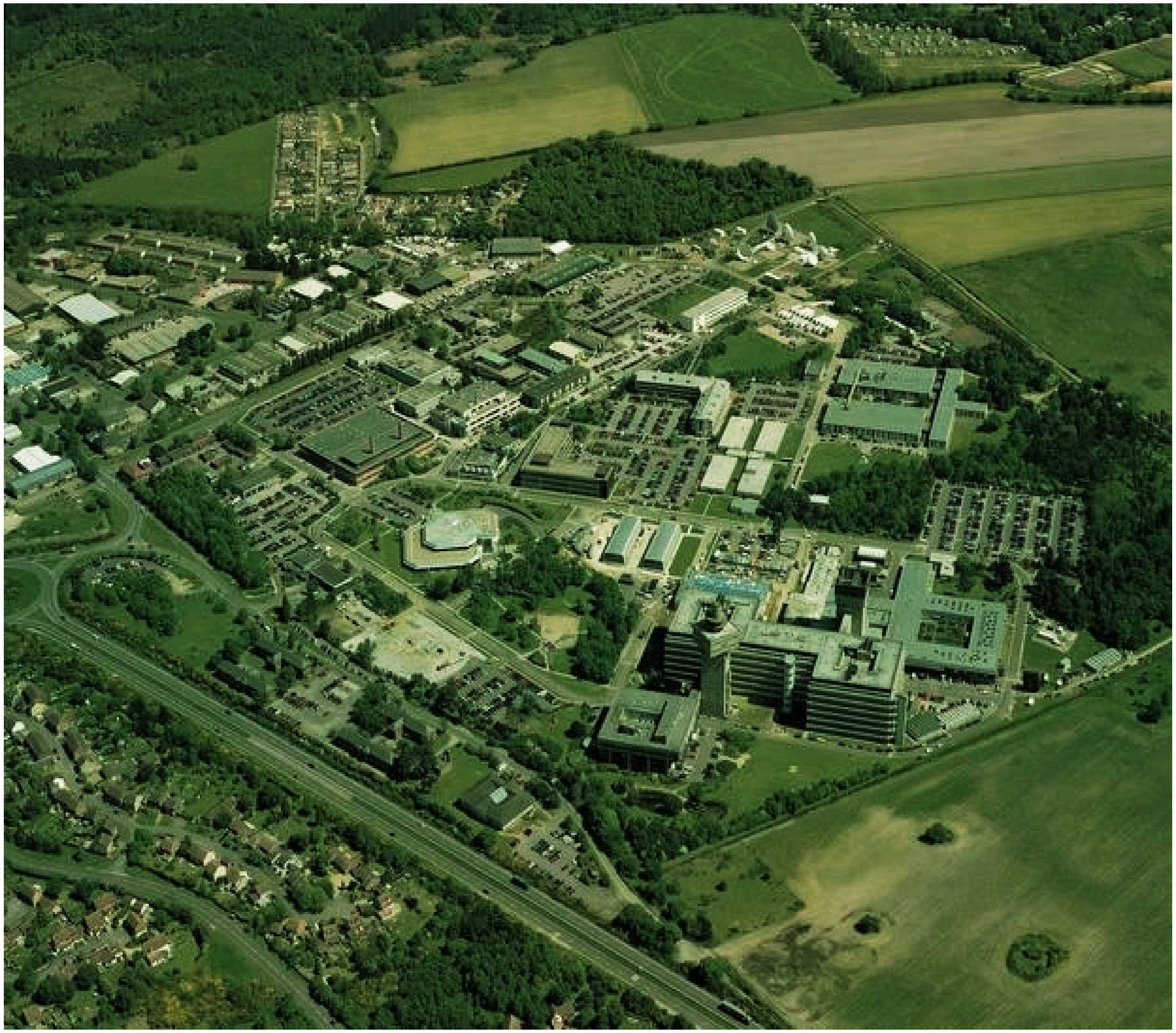
Keith Briggs Keith.Briggs@bt.com

more.btexact.com/people/briggsk2/

CABDyN seminar, Saïd Business School 2004 October 12

typeset 2004 October 14 10:20 in pdfLaTeX on a linux system

# BT Research at Martlesham, Suffolk



- Cambridge-Ipswich high-tech corridor
- 2000 technologists
- 15 companies
- UCL, Univ of Essex

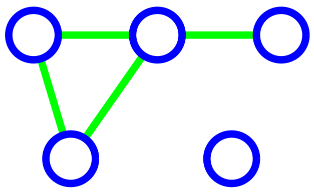
# Graphs

- set of nodes (vertices) $N = \{1, 2, 3, \dots\}$

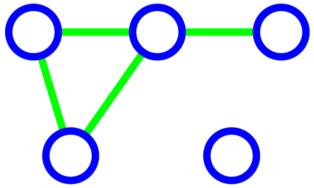
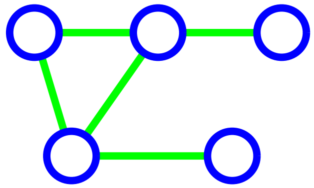
# Graphs

- set of nodes (vertices) $N = \{1, 2, 3, \ldots\}$
- set of edges (links) $E = \{\{1, 2\}, \{1, 4\}, \{2, 5\}, \ldots\}$

# Graphs

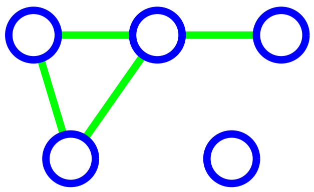
- set of nodes (vertices) $N = \{1, 2, 3, \dots\}$
- set of edges (links) $E = \{\{1, 2\}, \{1, 4\}, \{2, 5\}, \dots\}$
- (simple unlabelled undirected) graph: 

# Graphs

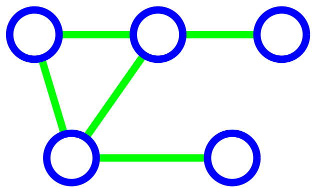
- set of nodes (vertices) $N = \{1, 2, 3, \ldots\}$
- set of edges (links) $E = \{\{1, 2\}, \{1, 4\}, \{2, 5\}, \ldots\}$
- (simple unlabelled undirected) graph: 
- (simple unlabelled undirected) connected graph: 

# Graphs

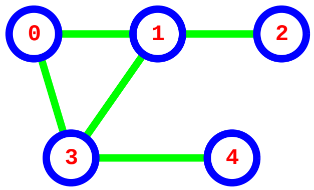
- set of nodes (vertices) $N = \{1, 2, 3, \dots\}$
- set of edges (links) $E = \{\{1, 2\}, \{1, 4\}, \{2, 5\}, \dots\}$
- (simple unlabelled undirected) graph:



- (simple unlabelled undirected) connected graph:



- (simple undirected) labelled graph:



## The Bernoulli random graph model $G\{n,p\}$

- let $G$ be a graph of $n$ nodes

## The Bernoulli random graph model $G\{n, p\}$

- let $G$ be a graph of $n$ nodes
- let $p = 1 - q$ be the probability that each possible edge exists

## The Bernoulli random graph model $G\{n,p\}$

- let $G$ be a graph of $n$ nodes
- let $p = 1 - q$ be the probability that each possible edge exists
- edge events are independent

# The Bernoulli random graph model $G\{n,p\}$

- let $G$ be a graph of $n$ nodes
- let $p = 1 - q$ be the probability that each possible edge exists
- edge events are independent
- let $P(n,p)$ be the probability that $G\{n,p\}$ is connected

## The Bernoulli random graph model $G\{n,p\}$

- let $G$ be a graph of $n$ nodes
- let $p = 1 - q$ be the probability that each possible edge exists
- edge events are independent
- let $P(n,p)$ be the probability that $G\{n,p\}$ is connected
- then $P(1,p) = 1$ and $P(n,p) = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} P(k,p) q^{k(n-k)}$ for $n = 2, 3, 4, \ldots$

# The Bernoulli random graph model $G\{n,p\}$

- let $G$ be a graph of $n$ nodes
- let $p = 1-q$ be the probability that each possible edge exists
- edge events are independent
- let $P(n,p)$ be the probability that $G\{n,p\}$ is connected
- then $P(1,p) = 1$ and $P(n,p) = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} P(k,p) q^{k(n-k)}$ for $n = 2, 3, 4, \ldots$

$$
\begin{aligned}
P(2,p) &= 1-q \\
P(3,p) &= (2\,q+1)\,(q-1)^2 \\
P(4,p) &= \left(6\,q^3+6\,q^2+3\,q+1\right)\,(1-q)^3 \\
P(5,p) &= \left(24\,q^6+36\,q^5+30\,q^4+20\,q^3+10\,q^2+4\,q+1\right)\,(q-1)^4
\end{aligned}
$$

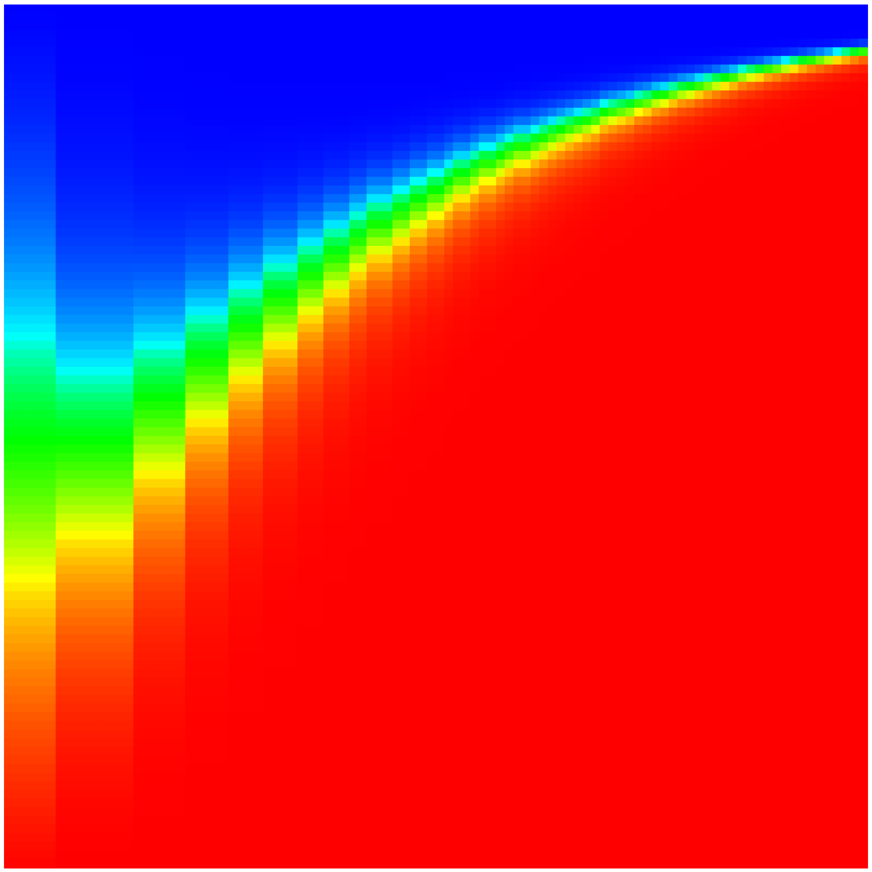
# The Bernoulli random graph model $G\{n,p\}$

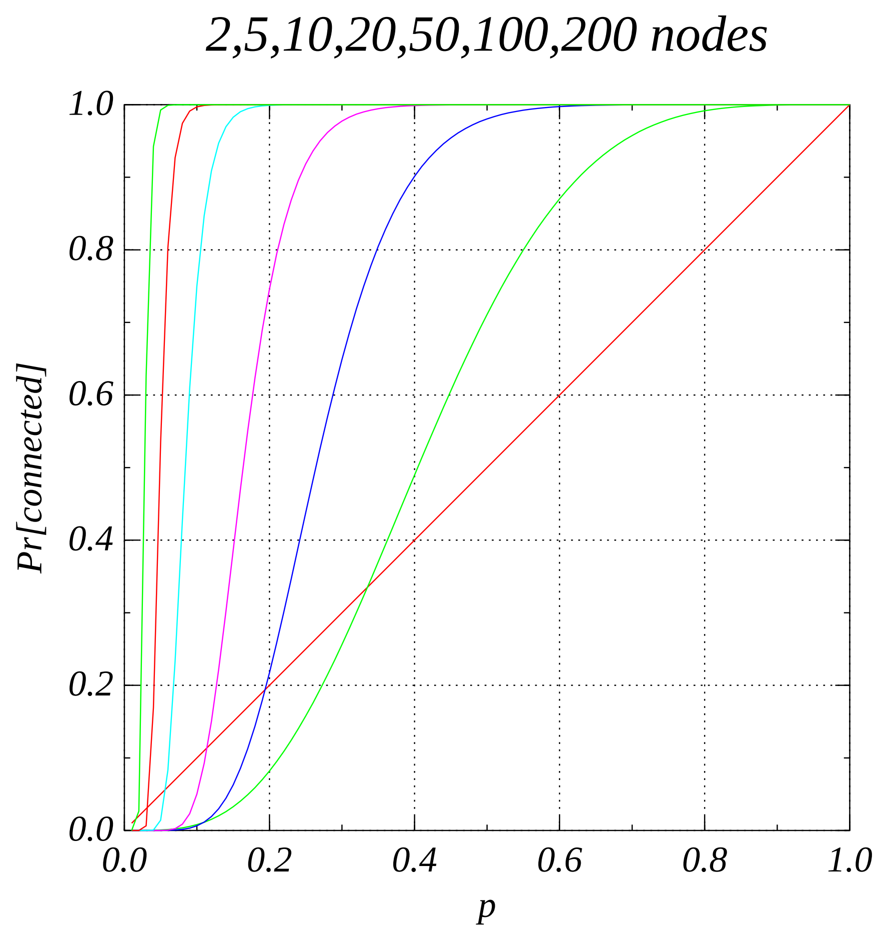
- let $G$ be a graph of $n$ nodes
- let $p=1-q$ be the probability that each possible edge exists
- edge events are independent
- let $P(n,p)$ be the probability that $G\{n,p\}$ is connected
- then $P(1,p)=1$ and $P(n,p)=1-\sum_{k=1}^{n-1}\binom{n-1}{k-1}P(k,p)q^{k(n-k)}$ for $n=2,3,4,\dots$

$$
\begin{aligned}
P(2,p) &= 1-q \\
P(3,p) &= (2\,q+1)\,(q-1)^2 \\
P(4,p) &= \left(6\,q^3+6\,q^2+3\,q+1\right)(1-q)^3 \\
P(5,p) &= \left(24\,q^6+36\,q^5+30\,q^4+20\,q^3+10\,q^2+4\,q+1\right)(q-1)^4
\end{aligned}
$$

- as $n\to\infty$, we have $P(n,p)\to 1-nq^{n-1}$

# Connectivity for the Bernoulli model





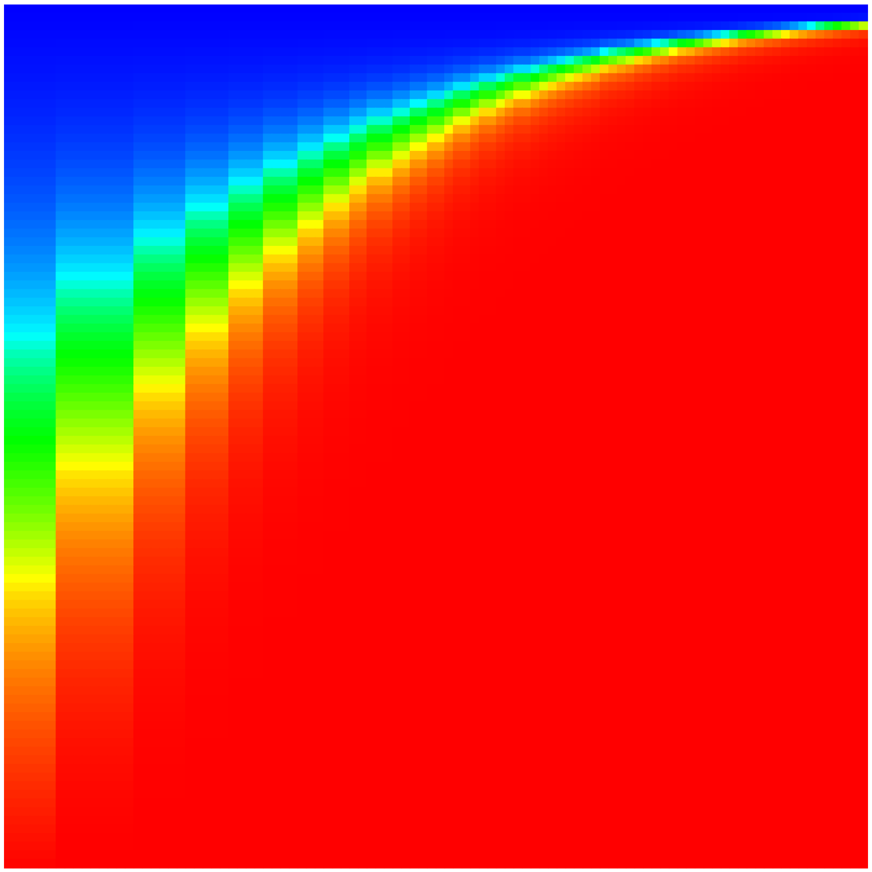

$x$-axis: $\log(n = \text{number of nodes}), n = 2, \ldots, 100$
$y$-axis: $p$, 0 at top, 1 at bottom
blue=0 red=1

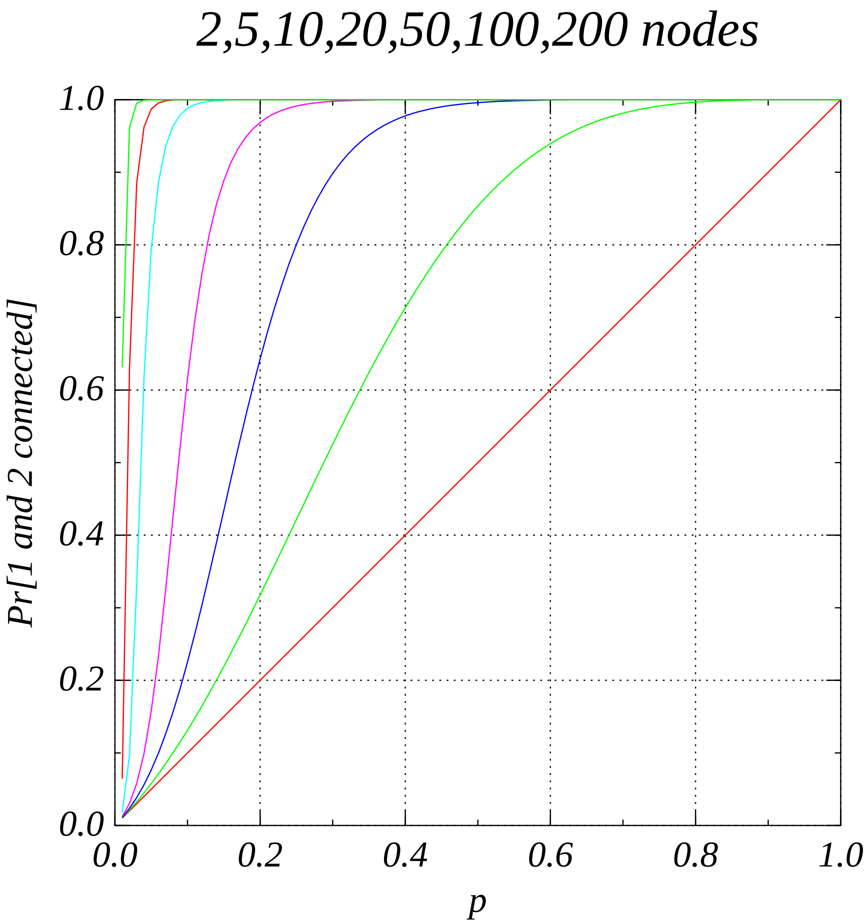
# More on the Bernoulli random graph model

- the probability that two *pre-specified* nodes (not randomly chosen) are connected is different. If we call this $R(n,p)$, then we have $R(1,p)=1$ and:

$$R(n,p)=1-\sum_{k=1}^{n-1}\binom{n-2}{k-1}P(k,p)q^{k(n-k)}, \quad n=2,3,4,\dots$$

# $\Pr[1 \text{ and } 2 \text{ connected}]$ for the Bernoulli model






$x$-axis: $\log(n = \text{number of nodes}), n = 2, \ldots, 100$
$y$-axis: $p$, 0 at top, 1 at bottom
blue=0 red=1

# Probability of connectivity for the $G(n,m)$ model

- problem: compute the numbers of connected labelled graphs with $n$ nodes and $m = n-1, n, n+1, n+2, \ldots$ edges

## Probability of connectivity for the $G(n, m)$ model

- problem: compute the numbers of connected labelled graphs with $n$ nodes and $m = n-1, n, n+1, n+2, \ldots$ edges
- with this information, compute the probability of a randomly chosen labelled graph being connected

# Probability of connectivity for the $G(n,m)$ model

- problem: compute the numbers of connected labelled graphs with $n$ nodes and $m = n-1, n, n+1, n+2, \ldots$ edges
- with this information, compute the probability of a randomly chosen labelled graph being connected
- compute large-$n$ asymptotics for these quantities, for fixed excess $k \equiv m-n$

# Probability of connectivity for the $G(n,m)$ model

- problem: compute the numbers of connected labelled graphs with $n$ nodes and $m=n-1, n, n+1, n+2, \ldots$ edges
- with this information, compute the probability of a randomly chosen labelled graph being connected
- compute large-$n$ asymptotics for these quantities, for fixed excess $k \equiv m-n$
- I have computed the accurate asymptotics and have checked the results against exact numerical data
  - a sketch of the ideas involved follows; full details are available on request

# The idea of generating functions

- generating function (gf):

$$\{a_1, a_2, a_3, \dots\} \leftrightarrow \sum_{k=1}^{\infty} a_k x^k$$

# The idea of generating functions

- generating function (gf):

$$\{a_1, a_2, a_3, \ldots\} \leftrightarrow \sum_{k=1}^{\infty} a_k x^k$$

- exponential generating function (egf):

$$\{a_1, a_2, a_3, \ldots\} \leftrightarrow \sum_{k=1}^{\infty} \frac{a_k}{k!}\, x^k$$

# Some known exponential generating functions

- exponential generating function enumerating labelled graphs ($[z^n]$: number with $n$ nodes; $[w^m]$: number with $m$ edges):

$$g(w,z) = \sum_{k=0}^{\infty} (1+w)^{\binom{k}{2}} z^k / k!$$

# Some known exponential generating functions

- exponential generating function enumerating labelled graphs ($[z^n]$: number with $n$ nodes; $[w^m]$: number with $m$ edges):

$$g(w,z) = \sum_{k=0}^{\infty} (1+w)^{\binom{k}{2}} z^k/k!$$

- exponential generating function enumerating connected labelled graphs:

$$\begin{aligned} c(w,z) &= \log(g(w,z)) \\ &= z + w\frac{z^2}{2} + (3w^2+w^3)\frac{z^3}{6} + (16w^3+15w^4+6w^5+w^6)\frac{z^4}{4!} + \dots \end{aligned}$$

# egfs for labelled graphs [jklp93]

- rooted labelled trees

$$T(z) = z\exp(T(z)) = \sum_{n\geqslant 1} n^{n-1}\frac{z^n}{n!} = z + \tfrac{2}{2!}z^2 + \tfrac{9}{3!}z^3 + \cdots$$

# egfs for labelled graphs [jklp93]

- rooted labelled trees

$$T(z) = z\exp(T(z)) = \sum_{n\geqslant 1} n^{n-1}\frac{z^n}{n!} = z + \tfrac{2}{2!}z^2 + \tfrac{9}{3!}z^3 + \cdots$$

- unrooted labelled trees

$$W_{-1}(z) = T(z) - T(z)^2/2 = z + \tfrac{1}{2!}z^2 + \tfrac{3}{3!}z^3 + \tfrac{16}{4!}z^4 + \ldots$$

# egfs for labelled graphs [jklp93]

- rooted labelled trees

$$T(z) = z\exp(T(z)) = \sum_{n\geqslant 1} n^{n-1}\frac{z^n}{n!} = z + \tfrac{2}{2!}z^2 + \tfrac{9}{3!}z^3 + \cdots$$

- unrooted labelled trees

$$W_{-1}(z) = T(z) - T(z)^2/2 = z + \tfrac{1}{2!}z^2 + \tfrac{3}{3!}z^3 + \tfrac{16}{4!}z^4 + \ldots$$

- unicyclic labelled graphs

$$W_0(z) = \frac{1}{2}\log\left[\frac{1}{1-T(z)}\right] - \frac{1}{2}T(z) - \frac{1}{4}T(z)^2 = \tfrac{1}{3!}z^3 + \tfrac{15}{4!}z^4 + \tfrac{222}{5!}z^5 + \tfrac{3660}{6!}z^6 + .$$

# egfs for labelled graphs [jklp93]

- rooted labelled trees

$$T(z) = z\exp(T(z)) = \sum_{n\geqslant 1} n^{n-1}\frac{z^n}{n!} = z+\tfrac{2}{2!}z^2+\tfrac{9}{3!}z^3+\cdots$$

- unrooted labelled trees

$$W_{-1}(z) = T(z)-T(z)^2/2 = z+\tfrac{1}{2!}z^2+\tfrac{3}{3!}z^3+\tfrac{16}{4!}z^4+\dots$$

- unicyclic labelled graphs

$$W_0(z) = \frac{1}{2}\log\left[\frac{1}{1-T(z)}\right]-\frac{1}{2}T(z)-\frac{1}{4}T(z)^2 = \tfrac{1}{3!}z^3+\tfrac{15}{4!}z^4+\tfrac{222}{5!}z^5+\tfrac{3660}{6!}z^6+\dots$$

- bicyclic labelled graphs

$$W_1(z) = \frac{T(z)^4\big(6-T(z)\big)}{24\big(1-T(z)\big)^3} = \tfrac{6}{4!}z^4+\tfrac{205}{5!}z^5+\tfrac{5700}{6!}z^6+\dots$$

# Introduction to asymptotic expansions

- Stirling:

$$\Gamma(n) \sim \left(\frac{2\pi}{n}\right)^{1/2} \left(\frac{n}{e}\right)^{n} \left[1+\frac{1}{12}n^{-1}+\frac{1}{288}n^{-2}-\frac{139}{51840}n^{-3}+\dots\right]$$

# Introduction to asymptotic expansions

- Stirling:

$$\Gamma(n) \sim \left(\frac{2\pi}{n}\right)^{1/2} \left(\frac{n}{e}\right)^{n} \left[1+\frac{1}{12}n^{-1}+\frac{1}{288}n^{-2}-\frac{139}{51840}n^{-3}+\dots\right]$$

- Taylor series:

$$1/\Gamma(n) = n+0.57721566\dots n-0.65587807\dots n^{2}+\dots$$

# Introduction to asymptotic expansions

- Stirling:

$$\Gamma(n) \sim \left(\frac{2\pi}{n}\right)^{1/2} \left(\frac{n}{e}\right)^n \left[1 + \frac{1}{12}n^{-1} + \frac{1}{288}n^{-2} - \frac{139}{51840}n^{-3} + \ldots\right]$$

- Taylor series:

$$1/\Gamma(n) = n + 0.57721566\ldots n - 0.65587807\ldots n^2 + \ldots$$

- e.g. for $n = 4$, $\Gamma(4) = 6$: 3 terms of asymptotic expansion give an absolute error $< 10^{-6}$

# Introduction to asymptotic expansions

- Stirling:

$$\Gamma(n) \sim \left(\frac{2\pi}{n}\right)^{1/2} \left(\frac{n}{e}\right)^{n} \left[1+\frac{1}{12}n^{-1}+\frac{1}{288}n^{-2}-\frac{139}{51840}n^{-3}+\ldots\right]$$

- Taylor series:

$$1/\Gamma(n) = n+0.57721566\ldots n-0.65587807\ldots n^2+\ldots$$

- e.g. for $n = 4$, $\Gamma(4) = 6$: 3 terms of asymptotic expansion give an absolute error $< 10^{-6}$
- cf. the Taylor series - 3 terms give an absolute error $> 5$

# Introduction to asymptotic expansions

- Stirling:

$$\Gamma(n) \sim \left(\frac{2\pi}{n}\right)^{1/2} \left(\frac{n}{e}\right)^n \left[1+\frac{1}{12}n^{-1}+\frac{1}{288}n^{-2}-\frac{139}{51840}n^{-3}+\dots\right]$$

- Taylor series:

$$1/\Gamma(n) = n+0.57721566\dots n-0.65587807\dots n^2+\dots$$

- e.g. for $n=4$, $\Gamma(4)=6$: 3 terms of asymptotic expansion give an absolute error $<10^{-6}$
- cf. the Taylor series - 3 terms give an absolute error $>5$
- asymptotic expansion diverges for all $n$!

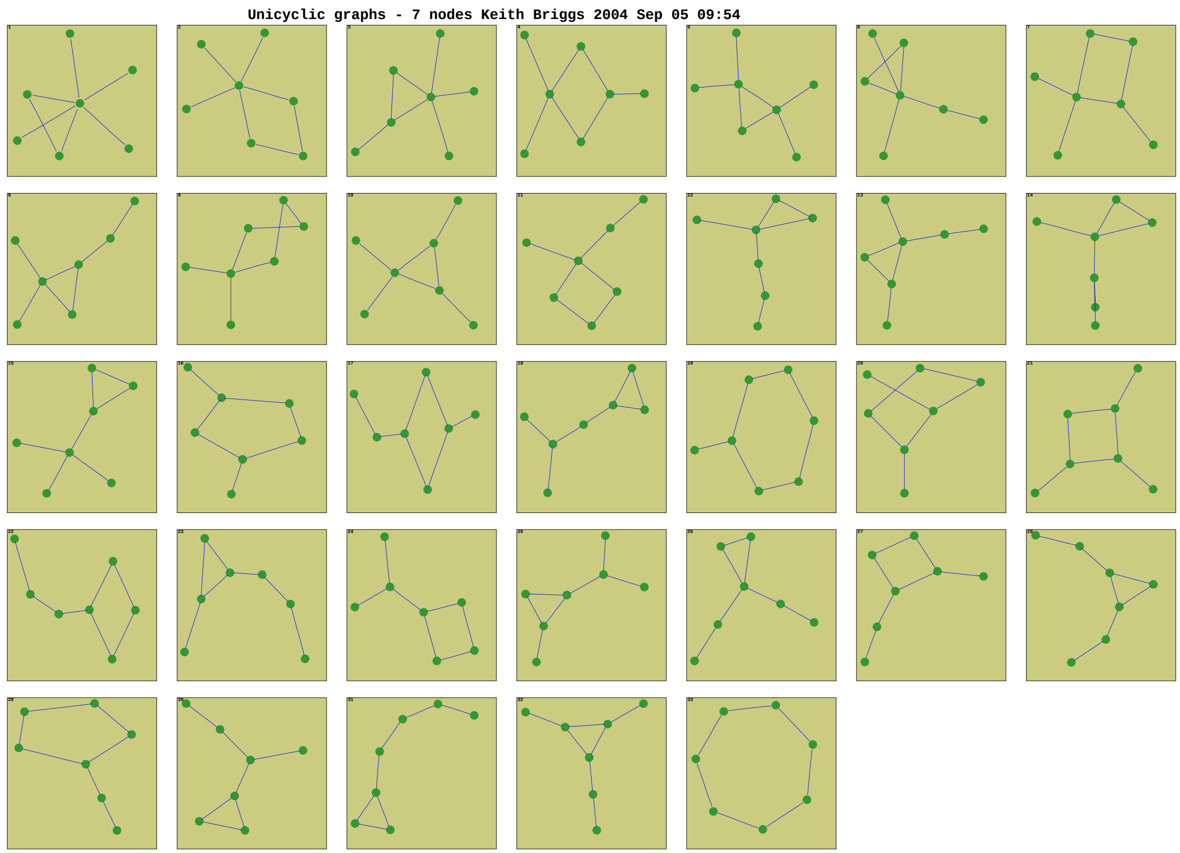
# Asymptotic expansion of $P(n,n+k)$

| $k$ | type | | $[n^0]$ | $[n^{-1/2}]$ | $[n^{-1}]$ | $[n^{-3/2}]$ |
|---|---|---|---|---|---|---|
| $-1$ | tree | $\frac{P(n,n-1)}{2^n e^{2-n} n^{-1/2} \xi}$ | $\frac{1}{2}$ | $0$ | $-\frac{7}{8}$ | $0$ |
| $0$ | unicycle | $\frac{P(n,n+0)}{2^n e^{2-n} \xi}$ | $\frac{1}{4}\xi$ | $-\frac{7}{6}$ | $\frac{1}{3}\xi$ | $-\frac{1051}{1080}$ |
| $1$ | bicycle | $\frac{P(n,n+1)}{2^n e^{2-n} n^{1/2} \xi}$ | $\frac{5}{12}$ | $-\frac{7}{12}\xi$ | $\frac{515}{144}$ | $-\frac{28}{9}\xi$ |

$\xi \equiv \sqrt{2\pi}$

# Unlabelled graphs

- much less is known about the unlabelled case
- the difficulties arise in distinguishing isomorphic graphs



# Asymptotics for unlabelled unicycles

- I get that the number $c(n,n)$ of connected unlabelled unicyclic graphs behaves like

$$\left(\frac{c(n,n)}{d^n}-\frac{1}{4^n}\right)n^{3/2} \sim$$

$$-0.4466410059+0.44311055235n^{-1}+0.91158865326n^{-2}+O(n^{-3})$$

# Geometric random graphs

Consider connectivity of nodes with a radio range $\rho$ placed uniformly and randomly in a bounded region under various models:

# Geometric random graphs

Consider connectivity of nodes with a radio range $\rho$ placed uniformly and randomly in a bounded region under various models:

- *Poisson 1d model*: the nodes exist on all of $\mathbb{R}$ with a exponential distribution of separation with parameter $\lambda$, and a window of unit length is placed over them. The number of nodes visible through the window is Poisson distributed

# Geometric random graphs

Consider connectivity of nodes with a radio range $\rho$ placed uniformly and randomly in a bounded region under various models:

- *Poisson 1d model*: the nodes exist on all of $\mathbb{R}$ with a exponential distribution of separation with parameter $\lambda$, and a window of unit length is placed over them. The number of nodes visible through the window is Poisson distributed
- *fixed-$n$ 1d model*: there are exactly $n$ nodes independently and uniformly placed in $[0, 1]$

# Geometric random graphs

Consider connectivity of nodes with a radio range $\rho$ placed uniformly and randomly in a bounded region under various models:

- *Poisson 1d model*: the nodes exist on all of $\mathbb{R}$ with a exponential distribution of separation with parameter $\lambda$, and a window of unit length is placed over them. The number of nodes visible through the window is Poisson distributed
- *fixed-$n$ 1d model*: there are exactly $n$ nodes independently and uniformly placed in $[0, 1]$
- *Poisson 2d model*: the nodes exist on all of $\mathbb{R}^2$ with a intensity $\lambda$, and a finite-area window is placed over them

# Geometric random graphs

Consider connectivity of nodes with a radio range $\rho$ placed uniformly and randomly in a bounded region under various models:

- *Poisson 1d model*: the nodes exist on all of $\mathbb{R}$ with a exponential distribution of separation with parameter $\lambda$, and a window of unit length is placed over them. The number of nodes visible through the window is Poisson distributed
- *fixed-$n$ 1d model*: there are exactly $n$ nodes independently and uniformly placed in $[0, 1]$
- *Poisson 2d model*: the nodes exist on all of $\mathbb{R}^2$ with a intensity $\lambda$, and a finite-area window is placed over them The number of nodes visible through the window is Poisson distributed
- *fixed-$n$ 2d model*: there are exactly $n$ nodes independently and uniformly placed in a bounded region $R$

# Geometric random graphs

Consider connectivity of nodes with a radio range $\rho$ placed uniformly and randomly in a bounded region under various models:

- *Poisson 1d model*: the nodes exist on all of $\mathbb{R}$ with a exponential distribution of separation with parameter $\lambda$, and a window of unit length is placed over them. The number of nodes visible through the window is Poisson distributed
- *fixed-$n$ 1d model*: there are exactly $n$ nodes independently and uniformly placed in $[0, 1]$
- *Poisson 2d model*: the nodes exist on all of $\mathbb{R}^2$ with a intensity $\lambda$, and a finite-area window is placed over them The number of nodes visible through the window is Poisson distributed
- *fixed-$n$ 2d model*: there are exactly $n$ nodes independently and uniformly placed in a bounded region $R$

Notation:

- ▷ *pdf=probability density function*
- ▷ *cdf=cumulative distribution function*
- ▷ *the notation is sloppy in not distinguishing a RV $X$ and its values $x$*
- ▷ $[\![x]\!]$ *is the indicator function: $1$ if $x$ is true, else $0$*

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$
- we now place a unit length window over $\mathbb{R}$ and assume that $n$ nodes are visible

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$
- we now place a unit length window over $\mathbb{R}$ and assume that $n$ nodes are visible
  - there are $n-1$ internode intervals, and the cdf of the maximum interval is $F_{n-1}(d) = (1 - e^{-\lambda d})^{n-1}$

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$
- we now place a unit length window over $\mathbb{R}$ and assume that $n$ nodes are visible
  - there are $n-1$ internode intervals, and the cdf of the maximum interval is $F_{n-1}(d) = (1-e^{-\lambda d})^{n-1}$
  - the cdf of the minimum interval is $F_1(d) = 1 - e^{-n\lambda d}$

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$
- we now place a unit length window over $\mathbb{R}$ and assume that $n$ nodes are visible
  - there are $n-1$ internode intervals, and the cdf of the maximum interval is $F_{n-1}(d) = (1 - e^{-\lambda d})^{n-1}$
  - the cdf of the minimum interval is $F_1(d) = 1 - e^{-n\lambda d}$
  - the pdf of the minimum interval is $f_1(d) = n\lambda e^{-n\lambda d}$

# Theory for the Poisson 1d model

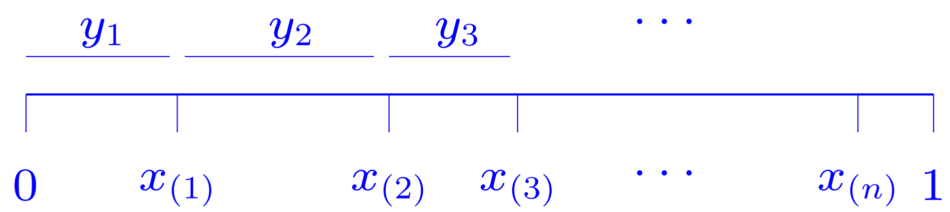
- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$
- we now place a unit length window over $\mathbb{R}$ and assume that $n$ nodes are visible
  - there are $n-1$ internode intervals, and the cdf of the maximum interval is $F_{n-1}(d) = (1-e^{-\lambda d})^{n-1}$
  - the cdf of the minimum interval is $F_1(d) = 1 - e^{-n\lambda d}$
  - the pdf of the minimum interval is $f_1(d) = n\lambda e^{-n\lambda d}$
  - the expectation of the minimum interval is $\mathbb{E}[d_{(1)}] = 1/(2\lambda)$, so is half the expectation of the internode distance

# Theory for the Poisson 1d model

- $\lambda$ is the intensity of nodes per unit length
- the pdf of the internode distance $d$ is $f(d) = \lambda e^{-\lambda d}$
- the cdf of the internode distance is $F(d) = 1 - e^{-\lambda d}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/\lambda$
- we now place a unit length window over $\mathbb{R}$ and assume that $n$ nodes are visible
  - there are $n-1$ internode intervals, and the cdf of the maximum interval is $F_{n-1}(d) = (1-e^{-\lambda d})^{n-1}$
  - the cdf of the minimum interval is $F_1(d) = 1 - e^{-n\lambda d}$
  - the pdf of the minimum interval is $f_1(d) = n\lambda e^{-n\lambda d}$
  - the expectation of the minimum interval is $\mathbb{E}[d_{(1)}] = 1/(2\lambda)$, so is half the expectation of the internode distance
  - the probability of full connectivity for the $n$ nodes is thus approximately (i.e. ignoring correlation and edge effects) $F_{n-1}(\rho) = (1-e^{-\lambda\rho})^{n-1}$

# Exact theory for the fixed-$n$ 1d model

- Let $y_k = x_{(k)} - x_{(k-1)}$ be the gaps ($k = 2, \ldots, n$), with $y_1 = x_{(1)}$




- their joint pdf is (for $1 \leqslant m \leqslant n$ and $\sum_{i=1}^{m} y_i \leqslant 1$)

$$f(y_1, y_2, \ldots, y_m) = \frac{n!}{(n-m)!}\left(1 - \sum_{i=1}^{m} y_i\right)^{n-m}$$

- if $c_i$ are constants such that $\sum_{i=1}^{m} c_i \leqslant 1$, then by integrating the pdf we obtain

$$\Pr\left[y_1 > c_1, y_2 > c_2, \ldots\right] = \left(1 - \sum_{i=1}^{m} c_i\right)^{n-1}$$

- Boole's law for the probability of at least one event $A_i$ of $n$ events $A_1, A_2, \ldots, A_n$ occurring is

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] = \sum_i \Pr\left[A_i\right] - \sum\sum_{i<j} \Pr\left[A_i A_j\right] + \cdots + (-1)^{n-1} \Pr\left[A_1 A_2 \ldots A_n\right]$$

# Exact theory for the fixed-$n$ 1d model (cotd)

- we don't care about $y_1$, so we put $c_1 = 0$

# Exact theory for the fixed-$n$ 1d model (cotd)

- we don't care about $y_1$, so we put $c_1 = 0$
- using Boole's law, the probability that the largest $y_k$ exceeds some constant $\rho$ is

$$\Pr\left[y_{(n)} > \rho\right] = (n-1)\Pr\left[y_1 > \rho\right] - \binom{n-1}{2}\Pr\left[y_1 > c_1, y_2 > c_2\right] + \ldots$$

# Exact theory for the fixed-$n$ 1d model (cotd)

- we don't care about $y_1$, so we put $c_1 = 0$
- using Boole's law, the probability that the largest $y_k$ exceeds some constant $\rho$ is

$$\Pr\left[y_{(n)} > \rho\right] = (n-1)\Pr\left[y_1 > \rho\right] - \binom{n-1}{2}\Pr\left[y_1 > c_1, y_2 > c_2\right] + \ldots$$

- thus

$$\boxed{\Pr\left[\text{connected}\right] = 1 - \sum_{i=1}^{\lfloor 1/\rho \rfloor} (-1)^{i+1} \binom{n-1}{i} (1 - i\rho)^n}$$

# Exact theory for the fixed-$n$ 1d model (cotd)

- we don't care about $y_1$, so we put $c_1 = 0$
- using Boole's law, the probability that the largest $y_k$ exceeds some constant $\rho$ is

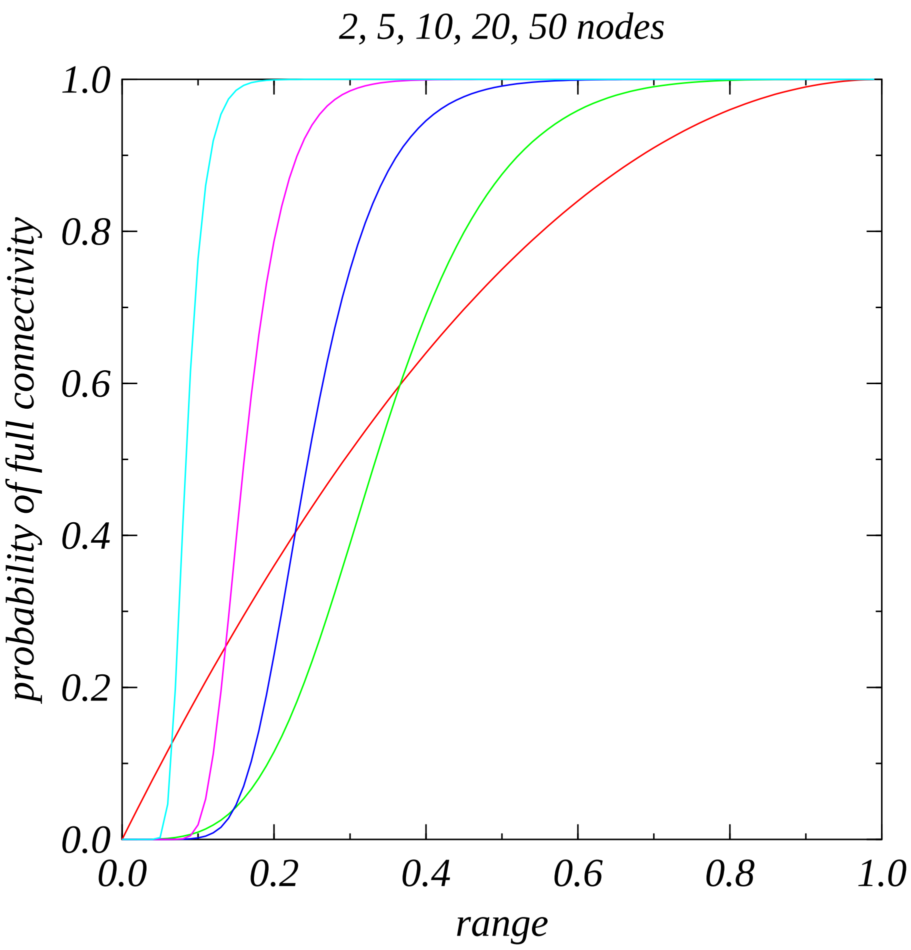
$$\Pr\left[y_{(n)} > \rho\right] = (n-1)\Pr\left[y_1 > \rho\right] - \binom{n-1}{2}\Pr\left[y_1 > c_1, y_2 > c_2\right] + \dots$$

- thus

$$\boxed{\Pr\left[\text{connected}\right] = 1 - \sum_{i=1}^{\lfloor 1/\rho \rfloor} (-1)^{i+1}\binom{n-1}{i}(1-i\rho)^n}$$

- note that for $\rho > 1/2$, this is exactly $1-(n-1)(1-\rho)^n$

# Probability of connectivity for the fixed-$n$ 1d model




2 5 10 20 50

# Theory for the Poisson 2d model

- $\lambda$ is the intensity of nodes per unit area

# Theory for the Poisson 2d model

- $\lambda$ is the intensity of nodes per unit area
- the pdf of the nearest neighbour distance $d$ is $f(d) = 2\pi\lambda d e^{-\lambda\pi d^2}$

# Theory for the Poisson 2d model

- $\lambda$ is the intensity of nodes per unit area
- the pdf of the nearest neighbour distance $d$ is $f(d) = 2\pi\lambda d e^{-\lambda\pi d^2}$
- the cdf of $d$ is $F(d) = 1 - e^{-\pi\lambda d^2}$

# Theory for the Poisson 2d model

- $\lambda$ is the intensity of nodes per unit area
- the pdf of the nearest neighbour distance $d$ is $f(d) = 2\pi\lambda d e^{-\lambda\pi d^2}$
- the cdf of $d$ is $F(d) = 1 - e^{-\pi\lambda d^2}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/(2\lambda^{1/2})$

# Theory for the Poisson 2d model

- $\lambda$ is the intensity of nodes per unit area
- the pdf of the nearest neighbour distance $d$ is $f(d) = 2\pi\lambda d e^{-\lambda\pi d^2}$
- the cdf of $d$ is $F(d) = 1 - e^{-\pi\lambda d^2}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/(2\lambda^{1/2})$
- the variance of $d$ is $(4 - \pi)/(4\pi\lambda)$

# Theory for the Poisson 2d model

- $\lambda$ is the intensity of nodes per unit area
- the pdf of the nearest neighbour distance $d$ is $f(d) = 2\pi\lambda d e^{-\lambda\pi d^2}$
- the cdf of $d$ is $F(d) = 1 - e^{-\pi\lambda d^2}$
- the expectation of $d$ is $\mathbb{E}[d] = 1/(2\lambda^{1/2})$
- the variance of $d$ is $(4-\pi)/(4\pi\lambda)$
- the probability of a node being isolated (i.e. having no neighbour within range $\rho$) is $e^{-\pi\lambda\rho^2}$

# Theory for the Poisson 2d model

we now place a window $R$ of area $A$ over $\mathbb{R}^2$

# Theory for the Poisson 2d model

we now place a window $R$ of area $A$ over $\mathbb{R}^2$

- the number of nodes visible will be Poisson distributed with mean $\lambda A$

# Theory for the Poisson 2d model

we now place a window $R$ of area $A$ over $\mathbb{R}^2$

- the number of nodes visible will be Poisson distributed with mean $\lambda A$
- Conditional on $n$ nodes being visible, and if the nearest neighbour distances were independent (which is *not* the case) the probability of no node being isolated would be $\left(1 - e^{-\pi\lambda\rho^2}\right)^n$

# Theory for the Poisson 2d model

we now place a window $R$ of area $A$ over $\mathbb{R}^2$

- the number of nodes visible will be Poisson distributed with mean $\lambda A$
- Conditional on $n$ nodes being visible, and if the nearest neighbour distances were independent (which is *not* the case) the probability of no node being isolated would be $\left(1 - e^{-\pi\lambda\rho^2}\right)^n$
- there is no simple way to compute the probability of full connectivity. However, since a necessary condition is that no node is isolated, the last expression is an approximate upper bound for the fixed-$n$ model and is plotted in red on the following graph

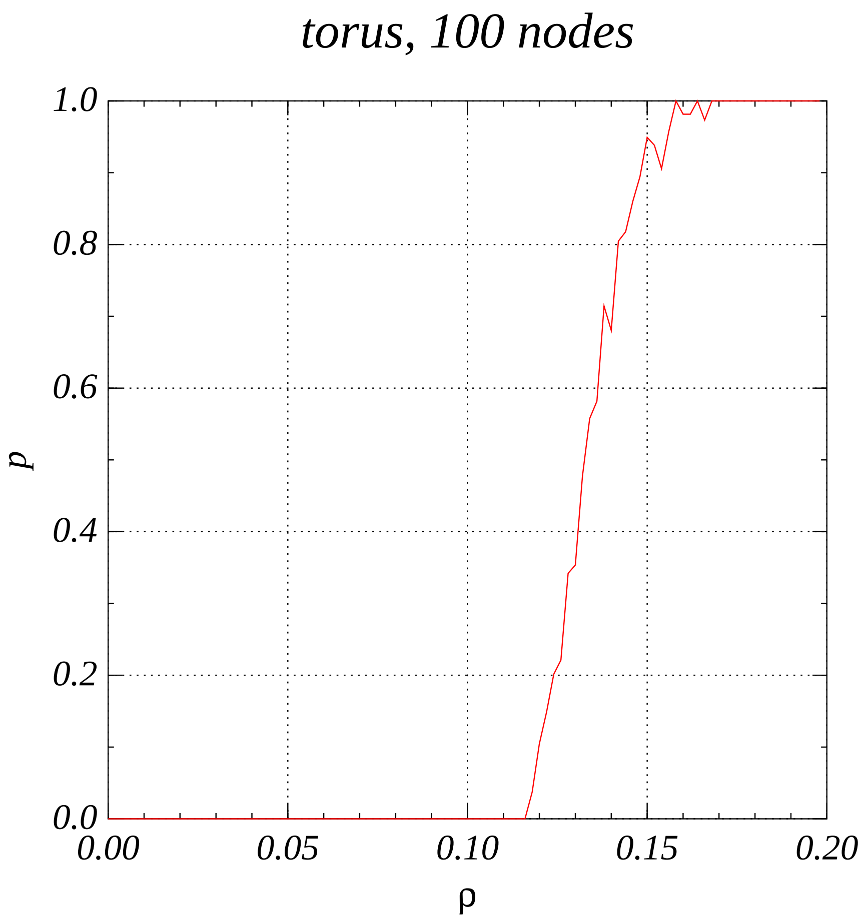
# Theory for the Poisson 2d model

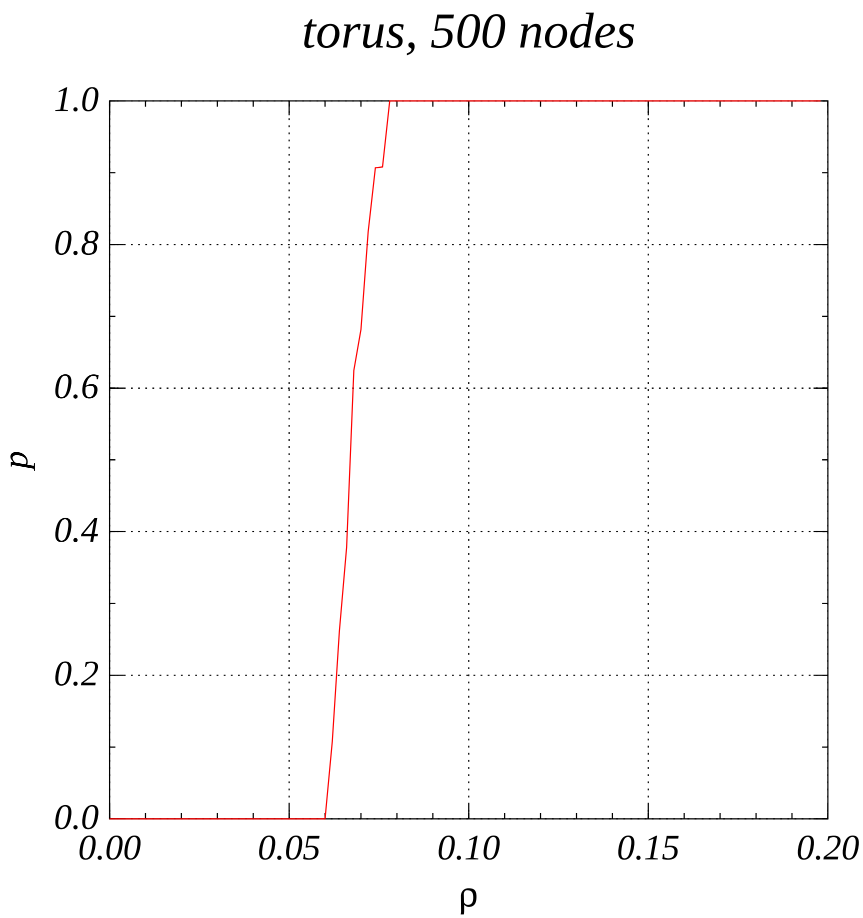
we now place a window $R$ of area $A$ over $\mathbb{R}^2$

- the number of nodes visible will be Poisson distributed with mean $\lambda A$
- Conditional on $n$ nodes being visible, and if the nearest neighbour distances were independent (which is *not* the case) the probability of no node being isolated would be $\left(1-e^{-\pi\lambda\rho^2}\right)^n$
- there is no simple way to compute the probability of full connectivity. However, since a necessary condition is that no node is isolated, the last expression is an approximate upper bound for the fixed-$n$ model and is plotted in red on the following graph
- the blue curve is the asymptotic probability of the whole region $R$ being covered

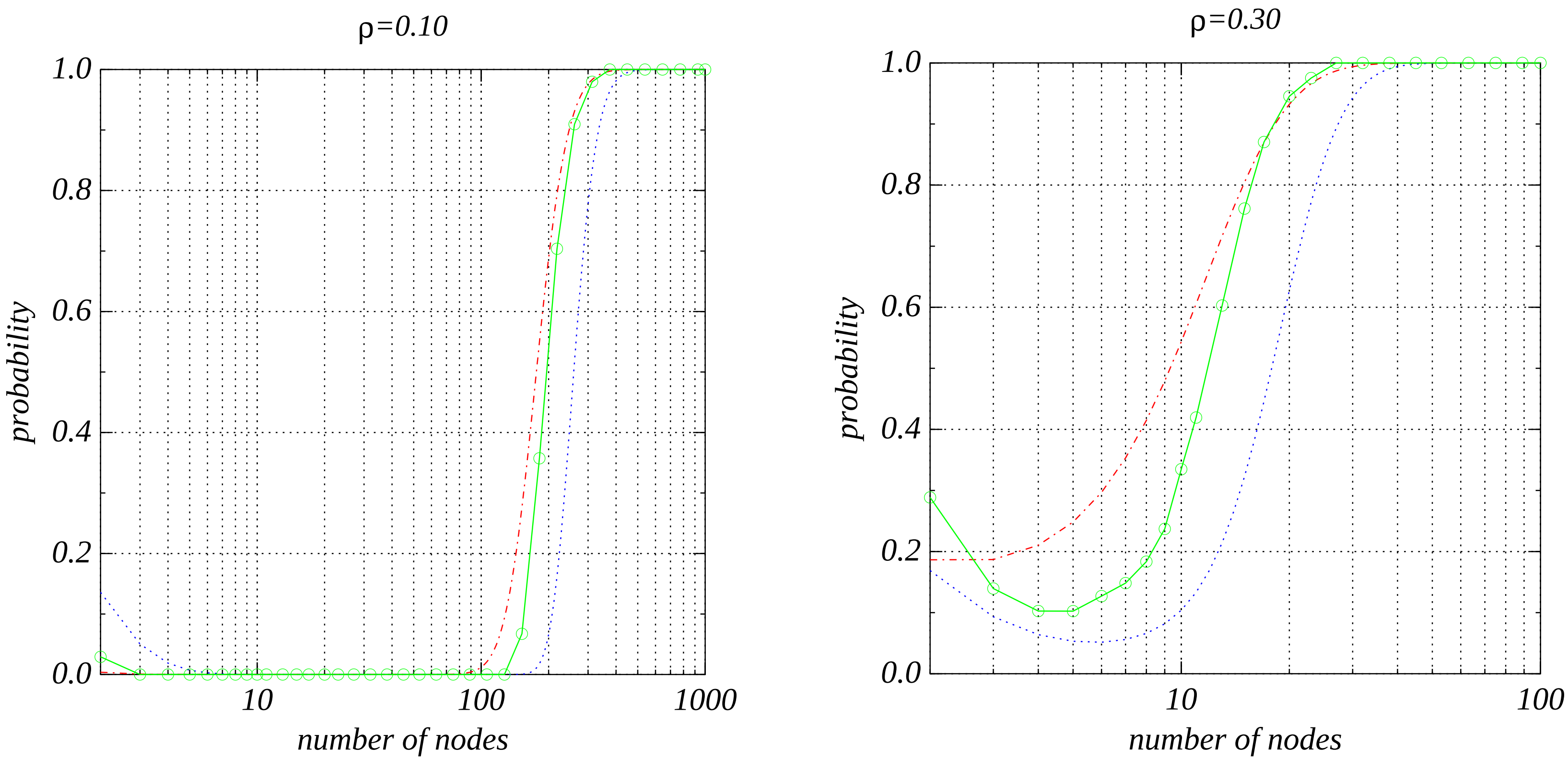
# The phase transition

- the critical radius is $O((\log(n)/(\pi n))^{1/2})$
- for $n = 100$, we estimate $\rho = 0.121$; for $n = 500$, $\rho = 0.0629$



# Simulation results - torus, $\rho = 0.1, 0.3$




$\left[1 - e^{-\pi\lambda\rho^2}\right]^n$ simulation asymptotic

# References

[gil59] E N Gilbert: Random graphs *Ann. Math. Statist.*, **30**, 1141-1144 (1959)

[pyk65] R Pyke, *Spacings*, J. Roy. Stat. Soc. B**27**, 395-449 (1965)

[jklp93] S Janson, D E Knuth, T Łuczak & B G Pittel: The birth of the giant component *Random Structures and Algorithms*, **4**, 233-358 (1993)
www-cs-faculty.stanford.edu/~knuth/papers/bgc.tex.gz

[fss04] Ph Flajolet, B Salvy and G Schaeffer: *Airy Phenomena and Analytic Combinatorics of Connected Graphs*
www.combinatorics.org/Volume_11/Abstracts/v11i1r34.html