Probability of connectedness of labelled graphs

Keith Briggs

Keith.Briggs@bt.com

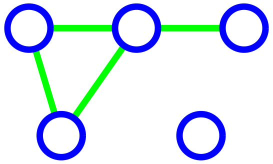
`more.btexact.com/people/briggsk2/`

CRG 2004 Sep 06 1500

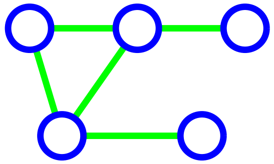
`prob-conn-2004sep06.tex` TYPESET 2004 SEPTEMBER 6 14:16 IN PDFLATEX ON A LINUX SYSTEM

Definitions for graphs

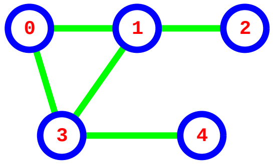
- (simple unlabelled undirected) graph:



- (simple unlabelled undirected) connected graph:



- (simple undirected) labelled graph:



The problem

- ☞ compute the numbers of connected labelled graphs with n nodes and $n-1, n, n+1, n+2, \ldots$ edges
  - ▷ *with this information, we can compute the probability of a randomly chosen labelled graph being connected*
- ☞ compute large-n asymptotics for these quantities, where the number of edges is only slightly larger than the number of nodes
- ☞ I began by reading the paper [fss04], but found some inconsistencies
- ☞ so I did some exact numerical calculations to try to establish the dominant asymptotics
- ☞ I then looked at some earlier papers and found that the required theory to compute exact asymptotics *is* known
- ☞ I computed the exact asymptotics and got perfect agreement with my exact numerical data

The inspirational paper [fss04]

- ☞ Philippe Flajolet, Bruno Salvy and Gilles Schaeffer: *Airy Phenomena and Analytic Combinatorics of Connected Graphs* `www.combinatorics.org/Volume_11/Abstracts/v11i1r34.html`
- ☞ The claim: the number $C(n,n+k)$ of labelled (étiquetés) connected graphs with n nodes and excess (edges-nodes) $=k \geqslant 2$ (why not for $k=1$?) is

$$A_k(1)\sqrt{\pi}\left(\frac{n}{e}\right)^n\left(\frac{n}{2}\right)^{\frac{3k-1}{2}}\left[\frac{1}{\Gamma(3k/2)}+\frac{A_k'(1)/A_k(1)-k}{\Gamma((3k-1)/2)}\sqrt{\frac{2}{n}}+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

☞

k	1	2	3	4	5	6	7
$A_k(1)$	5/24	5/16	1105/1152	565/128	82825/3072	19675/96	1282031525/688128
$A_k'(1)$	19/24	65/48	1945/384	21295/768	603965/3072	10454075/6144	1705122725/98304

- ☞ Airy in Playford:
  `www.ast.cam.ac.uk/~ipswich/History/Airys_Country_Retreat.htm`

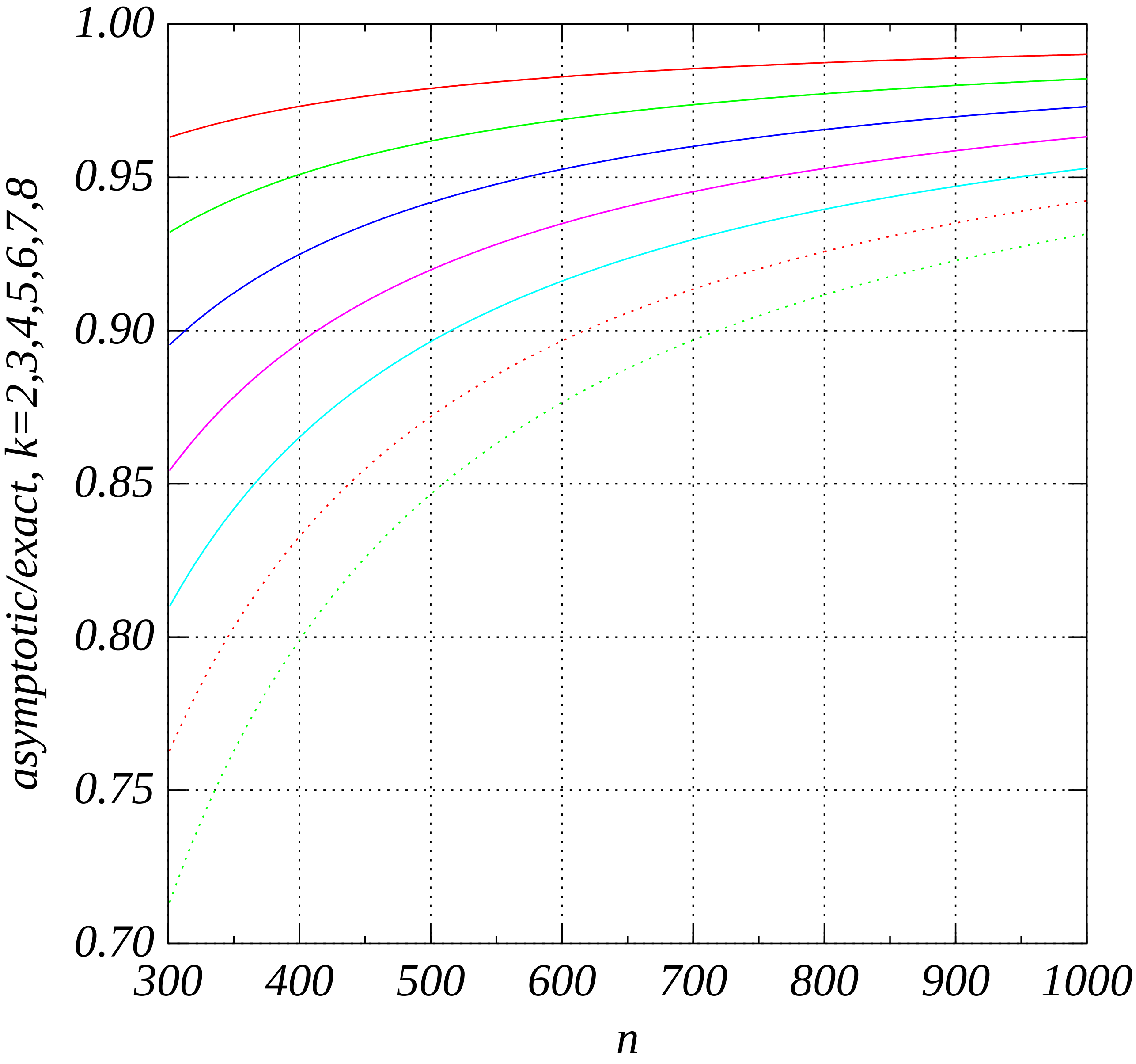
Some problems with the paper

- I did some comparisons with exact counts for up to $n=1000$ nodes and for excess $k=2,3,\ldots,8$
- The exact data was computed from the generating functions using maxima (found to be faster than maple)
- The fit was very bad
- This formula was found to fit the data much better for $k=2$:

$$A_k(1)\sqrt{\pi}n^n\left(\frac{n}{2}\right)^{\frac{3k-1}{2}}\left[\frac{1}{\Gamma(3k/2)}-\frac{A_k'(1)/A_k(1)-k}{\Gamma((3k-1)/2)}\sqrt{\frac{2}{n}}+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

- Also, on pages 4 and 24, I think S should have the expansion $1-(5/4)\alpha+(15/4)\alpha^2+\ldots$

Comparison of exact data with corrected formula



Definitions for generating functions

- ☞ generating function (gf):

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

- ☞ exponential generating function (egf):

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

Exponential generating functions

- ☞ exponential generating function for all labelled graphs:

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

- ☞ exponential generating function for all 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

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

- 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

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

- bicyclic labelled graphs

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

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 $C(n,n+k)/n^{n+\frac{3k-1}{2}}$

$\xi \equiv \sqrt{2\pi}$ green: from [bcm90] red: from [fss04] (with removal of factor e)

k	type	$[n^0]$	$[n^{-1/2}]$	$[n^{-1}]$	$[n^{-3/2}]$
-1	tree	1	0	0	0
0	unicycle	$\xi\frac{1}{4}$			
1	bicycle	$\frac{5}{24}$			
2	tricycle	$\xi\frac{5}{256}$ $\xi\frac{5}{256}$	$-\frac{35}{144}$		
3	quadricycle	$\frac{221}{1512}$ $\frac{221}{24192}$	$-\sqrt{\pi}\frac{35}{96}$		
4	pentacycle	$\xi\frac{113}{196608}$			

blue: conjectured by KMB from numerical experiments

k	type	$[n^0]$	$[n^{-1/2}]$	$[n^{-1}]$	$[n^{-3/2}]$	$[n^{-2}]$	$[n^{-5/2}]$
0	unicycle	$\xi\frac{1}{4}$	$-\frac{7}{6}$	$\xi\frac{1}{48}$	$\frac{131}{270}$	$\xi\frac{1}{1152}$	$-\frac{4}{2835}$?
1	bicycle	$\frac{5}{24}$	$-\xi\frac{7}{24}$	$\frac{25}{36}$	$-\xi\frac{7}{288}$	$-\frac{79}{3240}$?	
2	tricycle	$\xi\frac{5}{256}$	$-\frac{35}{144}$	$\xi\frac{1559}{9216}$	$-\frac{55}{144}$		
3	quadricycle	$\frac{221}{24192}$	$-\xi\frac{35}{10706}$				

Theory 1

The previous observations can be proved using theory available in [jklp93] and [fgkp95]. I sketch the computations.

- Ramanujan's Q-function is defined for $n=1,2,3,\ldots$: $Q(n)\equiv\sum_{k=1}^{\infty}\frac{n^{\underline{k}}}{n^k}=1+\frac{n-1}{n}+\frac{(n-1)(n-2)}{n^2}+\ldots,$
- $\sum_{n=1}^{\infty}Q(n)n^{n-1}\frac{z^n}{n!}=-\log(1-T(z))$, where T is the egf for rooted labelled trees: $T(z)=\sum_{n=1}^{\infty}\frac{n^{n-1}}{n!}z^n$
- $T(z)=z\exp(T(z))$

Theory 2

☞ to get the large-n asymptotics of Q, we first consider the related function $R(n) \equiv 1+\frac{n}{n+1}+\frac{n^2}{(n+1)(n+2)}+\ldots,\ n=1,2,3,\ldots$

- ▷ *we have* $Q(n)+R(n) = n!\,e^n/n^n$
- ▷ *let* $D(n) = R(n)-Q(n)$
- ▷ $\sum_{n=1}^{\infty} D(n) n^{n-1} \frac{z^n}{n!} = \log[\frac{(1-T(z))^2}{2(1-ez)}]$
- ▷ $D(n) \sim \sum_{k=1}^{\infty} c(k)[z^n](T(z)-1)^k$, *where* $c(k) \equiv [\delta^k] \log(\delta^2/2/(1-(1+\delta)e^{-\delta}))$
- ▷ *maple gives* $D(n) \sim \frac{2}{3}+\frac{8}{135}\,n^{-1}-\frac{16}{2835}\,n^{-2}-\frac{32}{8505}\,n^{-3}+\frac{17984}{12629925}\,n^{-4}+\frac{668288}{492567075}\,n^{-5}+O\left(n^{-6}\right)$

☞ now using $Q(n) = (n!\,e^n/n^n - D(n))/2$, we get

- ▷ $Q(n) \sim \frac{1}{2}n^{1/2}\sqrt{2\pi}-\frac{1}{3}+\frac{1}{24}\sqrt{2\pi}n^{-1/2}-\frac{4}{135}\,n^{-1}+\frac{1}{576}\sqrt{2\pi}n^{-3/2}+\frac{8}{2835}\,n^{-2}-\frac{139}{103680}\sqrt{2\pi}n^{-5/2}+\frac{16}{8505}\,n^{-3}-\frac{571}{4976640}\sqrt{2\pi}n^{-7/2}-\frac{8992}{12629925}\,n^{-4}+\frac{163879}{418037760}\sqrt{2\pi}n^{-9/2}-\frac{334144}{492567075}\,n^{-5}+\frac{5246819}{150493593600}\sqrt{2\pi}n^{-11/2}+O\left(n^{-6}\right)$

Theory 3

- Let W_k be the egf for connected labelled $(k+1)$-cyclic graphs
  - *for unrooted trees* $W_{-1}(z)=T(z)-T^2(z)/2$, $[z^n]W_{-1}(z)=n^{n-2}$
  - *for unicycles* $W_0(z)=-(\log(1-T(z))+T(z)+T^2(2)/2)/2$
  - *for bicycles* $W_1(z)=\frac{6T^4(z)-T^5(z)}{24(1-T(z))^3}$
  - *for* $k\geqslant 1$, $W_k(z)=\frac{A_k(T(z))}{(1-T(z))^{3k}}$, *where* A_k *are polynomials computable from results in [jklp93]*
- Knuth and Pittel's tree polynomials $t_n(y)$ ($y\neq 0$) are defined by $(1-T(z))^{-y}=\sum_{n=0}^{\infty}t_n(y)\frac{z^n}{n!}$
  - *we can compute these for* $y>0$ *from*
    $t_n(1)=1;\quad t_n(2)=n^n(1+Q(n));\quad t_n(y+2)=n\,t_n(y)/y+t_n(y+1)$
- thanks to this recurrence, the asymptotics for t_n follow from the known asymptotics of Q

Theory 4

Let $\xi=\sqrt{2\pi}$. All results agree with numerical estimates on this page.

- the number of connected unicycles is $C(n,n)=n![z^n]W_0(z)=\frac{1}{2}Q(n)n^{n-1}+3/2+t_n(-1)-t_n(-2)/4$
  - ▷ $\frac{C(n,n)}{n^n}\sim\frac{1}{4}\,\xi n^{-1/2}-\frac{7}{6}\,n^{-1}+\frac{1}{48}\,\xi n^{-3/2}+\frac{131}{270}\,n^{-2}+\frac{1}{1152}\,\xi n^{-5/2}+\frac{4}{2835}\,n^{-3}-\frac{139}{207360}\,\xi n^{-7/2}+\frac{8}{8505}\,n^{-4}-\frac{571}{9953280}\,\xi\left(n^{-1}\right)^{9/2}-\frac{4496}{12629925}\,n^{-5}+\frac{163879}{836075520}\,\xi n^{-11/2}+O\left(n^{-6}\right)$
- the number of connected bicycles is $C(n,n+1)=n![z^n]W_1(z)=\frac{5}{24}t_n(3)-\frac{19}{24}t_n(2)+\frac{13}{12}t_n(1)-\frac{7}{12}t_n(0)+\frac{1}{24}t_n(-1)+\frac{1}{24}t_n(-2)$
  - ▷ $\frac{C(n,n+1)}{n^n}\sim\frac{5}{24}\,n-\frac{7}{24}\,\xi n^{1/2}+\frac{25}{36}-\frac{7}{288}\,\xi n^{-1/2}-\frac{79}{3240}\,n^{-1}-\frac{7}{6912}\,\xi n^{-3/2}-\frac{413}{4860}\,n^{-2}+\frac{973}{1244160}\,\xi n^{-5/2}-\frac{4}{3645}\,n^{-3}+\frac{3997}{59719680}\,\xi n^{-7/2}+\frac{2248}{5412825}\,n^{-4}-\frac{163879}{716636160}\,\xi n^{-9/2}+\frac{83536}{211100175}\,n^{-5}-\frac{5246819}{257989017600}\,\xi n^{-11/2}+O\left(n^{-6}\right)$
- similarly, for the number of connected tricycles we get
  - ▷ $\frac{C(n,n+2)}{n^n}\sim\frac{5}{256}\,\xi n^{5/2}-\frac{35}{144}\,n^{2}+\frac{1559}{9216}\,\xi n^{3/2}-\frac{55}{144}\,n+\frac{33055}{221184}\,\xi n^{1/2}-\frac{41971}{136080}+\frac{31357}{2654208}\,\xi n^{-1/2}+\frac{1129}{81648}\,n^{-1}+O\left(n^{-3/2}\right)$

Probability of connectness 1

- ☞ we now have all the results needed to calculate the asymptotic probability $P(n, n+k)$ that a randomly chosen graph with n nodes and $n+k$ edges is connected (for $n \to \infty$ and small fixed k)

- ☞ the total number of graphs is $g(n, n+k) \equiv \binom{\binom{n}{2}}{n+k}$. This can be asymptotically expanded:
  - ▷ $\frac{g(n,n-1)}{\sqrt{\frac{2}{\pi}}e^{n-2}\left(\frac{n}{2}\right)^n n^{-3/2}} \sim 1+\frac{7}{4}n^{-1}+\frac{259}{96}n^{-2}+\frac{22393}{5760}n^{-3}+\frac{54359}{10240}n^{-4}+\frac{52279961}{7741440}n^{-5}+\frac{777755299}{103219200}n^{-6}+O\left(n^{-7}\right)$
  - ▷ $\frac{g(n,n+0)}{\sqrt{\frac{2}{\pi}}e^{n-2}\left(\frac{n}{2}\right)^n n^{-1/2}} \sim \frac{1}{2}-\frac{5}{8}n^{-1}-\frac{53}{192}n^{-2}-\frac{4067}{11520}n^{-3}-\frac{9817}{20480}n^{-4}-\frac{10813867}{15482880}n^{-5}-\frac{217565701}{206438400}n^{-6}-\frac{11591924473}{7431782400}n^{-7}+O\left(n^{-8}\right)$
  - ▷ $\frac{g(n,n+1)}{\sqrt{\frac{2}{\pi}}e^{n-2}\left(\frac{n}{2}\right)^n n^{3/2}} \sim \frac{1}{4}-\frac{21}{16}n^{-1}+\frac{811}{384}n^{-2}-\frac{43187}{23040}n^{-3}+\frac{159571}{73728}n^{-4}-\frac{55568731}{30965760}n^{-5}+\frac{2867716177}{1238630400}n^{-6}-\frac{3215346127}{2123366400}n^{-7}+\frac{1317595356557}{475634073600}n^{-8}+O\left(n^{-9}\right)$
  - ▷ $\ldots$
  - ▷ $g(n, n+k) \sim \sqrt{\frac{2}{\pi}}e^{n-2}\left(\frac{n}{2}\right)^n n^{k-1/2}\left(2^{-k-1}+O(n^{-1})\right)$

Probability of connectness 2

☞ $$\frac{P(n,n-1)}{2^n e^{2-n} n^{-1/2} \xi} \sim \frac{1}{2} - \frac{7}{8} n^{-1} + \frac{35}{192} n^{-2} + \frac{1127}{11520} n^{-3} + \frac{5189}{61440} n^{-4} + \frac{457915}{3096576} n^{-5} + \frac{570281371}{1857945600} n^{-6} + \frac{291736667}{495452160} n^{-7} + O\left(n^{-8}\right)$$

- ▷ *check:* $n = 10$, *exact*=0.1128460393, *asymptotic*=0.1128460359

☞ $$\frac{P(n,n+0)}{2^n e^{2-n} \xi} \sim \frac{1}{4}\xi - \frac{7}{6} n^{-1/2} + \frac{1}{3}\xi n^{-1} - \frac{1051}{1080} n^{-3/2} + \frac{5}{9}\xi n^{-2} + O\left(n^{-3}\right)$$

- ▷ *check:* $n = 10$, *exact*=0.276, *asymptotic*=0.319

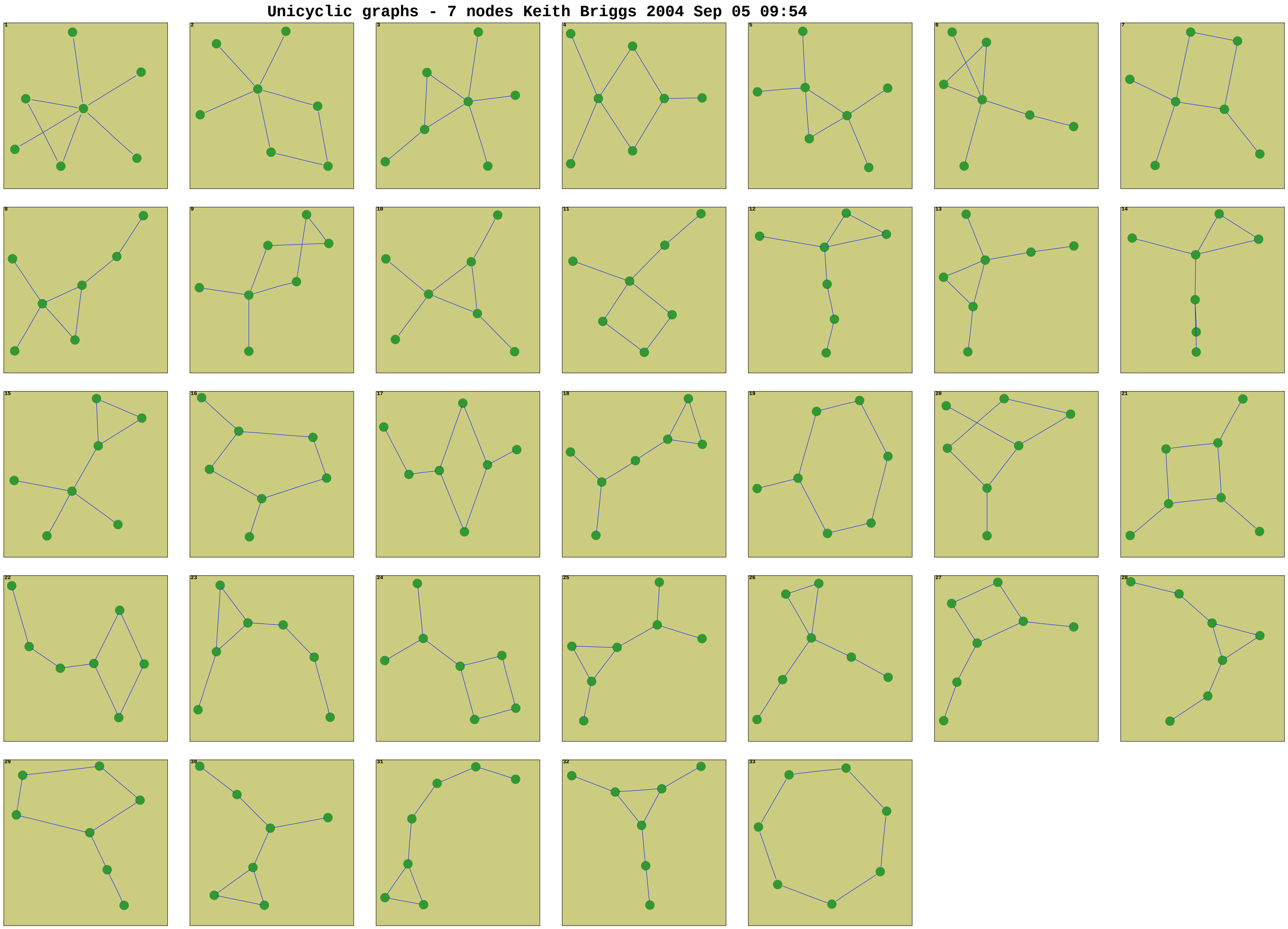
☞ $$\frac{P(n,n+1)}{2^n e^{2-n} n^{1/2} \xi} \sim \frac{5}{12} - \frac{7}{12}\xi n^{-1/2} + \frac{515}{144} n^{-1} - \frac{28}{9}\xi n^{-3/2} + \frac{788347}{51840} n^{-2} - \frac{308}{27}\xi n^{-5/2} + O\left(n^{-3}\right)$$

- ▷ *check:* $n = 10$, *exact*=0.437, *asymptotic*=0.407
- ▷ *check:* $n = 20$, *exact*=0.037108, *asymptotic*=0.037245
- ▷ *check:* $n = 100$, *exact*=2.617608×10^{-12}, *asymptotic*=2.617596×10^{-12}

The unlabelled case - unicycles NEW

- A connected unicyclic graph is an undirected cycle of 3 or more rooted trees. Start with a single undirected cycle (or polygon) graph. It must have at least 3 nodes. Hanging from each node in the cycle is a tree (a tree is of course a connected acyclic graph). The node where the tree intersects the cycle is the root, thus it is (combinatorially) a rooted tree.
    - *A001429 is undirected cycles of 3 or more rooted trees*
    - *A068051 is undirected cycles of 1 or more rooted trees*
    - *A027852 is undirected cycles of exactly 2 rooted trees*
    - *A000081 is undirected cycles of exactly 1 rooted tree*

The unlabelled case - unicycles for $n = 7$



The unlabelled case - asymptotics for unicycles

- $C(n,n) \sim 2.955765286^n\, n^{-1}(1/4 - 0.44689 n^{-1/2} + 0.02197 n^{-1} + \dots)$

n	1 term	2 terms	3 terms
10	0.516328	1.187715447	1.164181370
100	0.823154	1.002325806	1.001254380
500	0.920261	1.000220890	1.000029852
1000	0.943559	1.000092238	0.999999092
2000	0.960070	1.000042796	0.999997026
5000	0.974737	1.000017220	0.999999188

References

[bcm90] E A Bender, E R Canfield & B D McKay: *The asymptotic number of labeled connected graphs with a given number of vertices and edges*, Random structures and algorithms **1**, 127-169 (1990)

[slo03] N J A Sloane, ed. (2003), The on-line encyclopedia of integer sequences `www.research.att.com/~njas/sequences/`

[jklp93] S Janson, D E Knuth, T Luczak & 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`

[fgkp95] Ph Flajolet, P Grabner, P Kirschenhofer & H Prodinger: *On Ramanujan's Q-function* Journal of Computational and Applied Mathematics **58** 103-116 (1995) `www.inria.fr/rrrt/rr-1760.html`

[fss04] Philippe Flajolet, Bruno Salvy and Gilles Schaeffer: *Airy Phenomena and Analytic Combinatorics of Connected Graphs* `www.combinatorics.org/Volume_11/Abstracts/v11i1r34.html`