The diameter of random Cayley digraphs

Mark C. Wilson
www.cs.auckland.ac.nz/~mcw/

Department of Computer Science
University of Auckland

Alden Biesen, 2006-07-06

The University of Auckland
NEW ZEALAND

The genesis of this paper

- The second half was done by Manuel Lladser (Boulder) and Mark Wilson (Auckland) with major input from Robin Pemantle (Philadelphia) . . .

The genesis of this paper

- The second half was done by Manuel Lladser (Boulder) and Mark Wilson (Auckland) with major input from Robin Pemantle (Philadelphia) . . .
- who heard from Herb Wilf (Philadelphia) . . .

The genesis of this paper

- The second half was done by Manuel Lladser (Boulder) and Mark Wilson (Auckland) with major input from Robin Pemantle (Philadelphia) . . .
- who heard from Herb Wilf (Philadelphia) . . .
- who was asked by Marko Petkovšek (Ljubljana) . . .

The genesis of this paper

- The second half was done by Manuel Lladser (Boulder) and Mark Wilson (Auckland) with major input from Robin Pemantle (Philadelphia) . . .
- who heard from Herb Wilf (Philadelphia) . . .
- who was asked by Marko Petkovšek (Ljubljana) . . .
- who was asked by Primož Potočnik (Ljubljana) and Jana Šiagiová (Bratislava) . . .

The genesis of this paper

- The second half was done by Manuel Lladser (Boulder) and Mark Wilson (Auckland) with major input from Robin Pemantle (Philadelphia) ...
- who heard from Herb Wilf (Philadelphia) ...
- who was asked by Marko Petkovšek (Ljubljana) ...
- who was asked by Primož Potočnik (Ljubljana) and Jana Šiagiová (Bratislava) ...
- who had completed the first half while visiting Jozef Širáň (Auckland). (!)

Background and motivation

- Random graphs and digraphs have diameter 2 with high probability as long as they are not too sparse.

Background and motivation

- Random graphs and digraphs have diameter 2 with high probability as long as they are not too sparse.
- We wish to sharpen this for various families of graphs. In particular, we investigate the diameter of random Cayley digraphs.

Background and motivation

- Random graphs and digraphs have diameter 2 with high probability as long as they are not too sparse.
- We wish to sharpen this for various families of graphs. In particular, we investigate the diameter of random Cayley digraphs.
- Cayley graphs are often used as models for communications networks. The diameter is the number of rounds needed to send a message across the graph.

Background and motivation

- Random graphs and digraphs have diameter 2 with high probability as long as they are not too sparse.
- We wish to sharpen this for various families of graphs. In particular, we investigate the diameter of random Cayley digraphs.
- Cayley graphs are often used as models for communications networks. The diameter is the number of rounds needed to send a message across the graph.
- Cayley graphs are also useful for studying groups: the diameter is the maximum length of words in the generators required to generate G as a semigroup.

Background and motivation

- Random graphs and digraphs have diameter 2 with high probability as long as they are not too sparse.
- We wish to sharpen this for various families of graphs. In particular, we investigate the diameter of random Cayley digraphs.
- Cayley graphs are often used as models for communications networks. The diameter is the number of rounds needed to send a message across the graph.
- Cayley graphs are also useful for studying groups: the diameter is the maximum length of words in the generators required to generate G as a semigroup.
- Many combinatorial generation algorithms amount to finding Hamilton cycles in Cayley graphs.

Definitions

- Let G be a finite group and S a set of non-identity elements of G. The Cayley digraph $\Gamma = \mathrm{Cay}(G, S)$ has vertex set G and arcs of the form (g, gs) where $g \in G, s \in S$.

Definitions

- Let G be a finite group and S a set of non-identity elements of G. The Cayley digraph $\Gamma = \mathrm{Cay}(G, S)$ has vertex set G and arcs of the form (g, gs) where $g \in G, s \in S$.
- The distance $\partial(v, w)$ between v and w in G is the minimal number of arcs in a path from v to w.

Definitions

- Let G be a finite group and S a set of non-identity elements of G. The Cayley digraph $\Gamma = \mathrm{Cay}(G, S)$ has vertex set G and arcs of the form (g, gs) where $g \in G, s \in S$.
- The distance $\partial(v, w)$ between v and w in G is the minimal number of arcs in a path from v to w.
- The diameter $\mathrm{diam}(\Gamma)$ is the minimal d such that all distances between pairs of elements of Γ are at most d.

Definitions

- Let G be a finite group and S a set of non-identity elements of G. The Cayley digraph $\Gamma = \mathrm{Cay}(G, S)$ has vertex set G and arcs of the form (g, gs) where $g \in G, s \in S$.
- The distance $\partial(v, w)$ between v and w in G is the minimal number of arcs in a path from v to w.
- The diameter $\mathrm{diam}(\Gamma)$ is the minimal d such that all distances between pairs of elements of Γ are at most d.
- By vertex-transitivity of Γ, $\mathrm{diam}(\Gamma) = \max_v \partial(1, v)$.

The main question

- How does $\operatorname{diam}\operatorname{Cay}(G, S)$ behave asymptotically as $n \to \infty$? What relationship between $k := |S|$ and $n := |G|$ must hold in order that the diameter is equal to 2 with high probability?

The main question

- How does $\operatorname{diam}\operatorname{Cay}(G, S)$ behave asymptotically as $n \to \infty$? What relationship between $k := |S|$ and $n := |G|$ must hold in order that the diameter is equal to 2 with high probability?
- (Lower bound) The Moore bound shows that if $1 + k^2 < n$, then $\operatorname{diam}\operatorname{Cay}(G, S) > 2$.

The main question

- How does $\operatorname{diam} \operatorname{Cay}(G, S)$ behave asymptotically as $n \to \infty$? What relationship between $k := |S|$ and $n := |G|$ must hold in order that the diameter is equal to 2 with high probability?
- (Lower bound) The Moore bound shows that if $1 + k^2 < n$, then $\operatorname{diam} \operatorname{Cay}(G, S) > 2$.
- (Upper bound) If $k \geq n/2$ then $\operatorname{diam} \operatorname{Cay}(G, S) = 2$.

The main question

- How does $\operatorname{diam}\operatorname{Cay}(G, S)$ behave asymptotically as $n \to \infty$? What relationship between $k := |S|$ and $n := |G|$ must hold in order that the diameter is equal to 2 with high probability?
- (Lower bound) The Moore bound shows that if $1 + k^2 < n$, then $\operatorname{diam}\operatorname{Cay}(G, S) > 2$.
- (Upper bound) If $k \geq n/2$ then $\operatorname{diam}\operatorname{Cay}(G, S) = 2$.
- What about the region between $\sqrt{n}$ and $n/2$?

The probability model

- For each k with $1 \leq k < n$, define $\mathbb{P}(G, k)$ to consist of all possible generating sets (as above) that have size k. Give $\mathbb{P}$ the uniform measure.

The probability model

- For each k with $1 \leq k < n$, define $\mathbb{P}(G, k)$ to consist of all possible generating sets (as above) that have size k. Give $\mathbb{P}$ the uniform measure.
- Let $\mathrm{Diam}_{n,k}$ be the random variable on $\mathbb{P}$ equal to $\mathrm{diam}\,\mathrm{Cay}(G, S)$.

The probability model

- For each k with $1 \leq k < n$, define $\mathbb{P}(G, k)$ to consist of all possible generating sets (as above) that have size k. Give $\mathbb{P}$ the uniform measure.
- Let $\mathrm{Diam}_{n,k}$ be the random variable on $\mathbb{P}$ equal to $\operatorname{diam} \operatorname{Cay}(G, S)$.
- We seek the asymptotics of $\Pr(\mathrm{Diam}_{n,k} > 2)$ as $n \to \infty$ and k varies with n, say $k = f(n)$.

The probability model

- For each k with $1 \leq k < n$, define $\mathbb{P}(G, k)$ to consist of all possible generating sets (as above) that have size k. Give $\mathbb{P}$ the uniform measure.
- Let $\mathrm{Diam}_{n,k}$ be the random variable on $\mathbb{P}$ equal to $\operatorname{diam} \operatorname{Cay}(G, S)$.
- We seek the asymptotics of $\Pr(\mathrm{Diam}_{n,k} > 2)$ as $n \to \infty$ and k varies with n, say $k = f(n)$.
- As far as we know even the linear case $f(n) = cn, 0 < c < 1/2$ is unexplored. Other interesting special cases: $k = \lfloor n^{\alpha} \rfloor$ for $1/2 < \alpha < 1$.

Overview of results of this section

- For $2t \leq n, k \leq n$ define

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

Overview of results of this section

- For $2t \leq n, k \leq n$ define

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

- For general groups::

$$\Pr(\mathrm{Diam} > 2) \leq (n-1)p\left(n-1, k, \lfloor \frac{n-4}{12} \rfloor\right).$$

Overview of results of this section

- For $2t \le n, k \le n$ define

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

- For general groups::

$$\Pr(\mathrm{Diam} > 2) \le (n-1)p\left(n-1, k, \lfloor \frac{n-4}{12} \rfloor\right).$$

- For elementary abelian 2-groups:

$$p(n-1, k, \frac{n-1}{2}) - \frac{k}{n-1} \le \Pr(\mathrm{Diam} > 2) \le (n-1)p\left(n-1, k, \frac{n-1}{2}\right).$$

Overview of results of this section

- For $2t \leq n, k \leq n$ define

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

- For general groups::

$$\Pr(\text{Diam} > 2) \leq (n-1)p\left(n-1, k, \lfloor \frac{n-4}{12} \rfloor\right).$$

- For elementary abelian 2-groups:

$$p(n-1,k,\frac{n-1}{2}) - \frac{k}{n-1} \leq \Pr(\text{Diam} > 2) \leq (n-1)p\left(n-1,k,\frac{n-1}{2}\right).$$

- We therefore want to know the asymptotics of $p(n,k,t)$ for the given values of t, and for various k depending on n.

A basic estimate

- Let $T(y)$ be the event that there exists a path of length 2 from 1 to y, and let $M = \max_y \Pr \overline{T(y)}$. Then

$$M - \frac{k}{n-1} \leq \Pr(\text{Diam} > 2) \leq (n-1)M.$$

A basic estimate

- Let $T(y)$ be the event that there exists a path of length 2 from 1 to y, and let $M = \max_y \Pr \overline{T(y)}$. Then

$$M - \frac{k}{n-1} \leq \Pr(\text{Diam} > 2) \leq (n-1)M.$$

- Details:

A basic estimate

- Let $T(y)$ be the event that there exists a path of length 2 from 1 to y, and let $M = \max_y \Pr \overline{T(y)}$. Then

$$M - \frac{k}{n-1} \leq \Pr(\text{Diam} > 2) \leq (n-1)M.$$

- Details:
  - If $\operatorname{diam} \operatorname{Cay}(G, S) > 2$, there is y with $S \in \overline{T(y)}$.

A basic estimate

- Let $T(y)$ be the event that there exists a path of length 2 from 1 to y, and let $M = \max_y \Pr \overline{T(y)}$. Then

$$M - \frac{k}{n-1} \leq \Pr(\mathrm{Diam} > 2) \leq (n-1)M.$$

- Details:
  - If $\operatorname{diam} \operatorname{Cay}(G, S) > 2$, there is y with $S \in \overline{T(y)}$.
  - If $\operatorname{diam} \operatorname{Cay}(G, S) \leq 2$ then for every y, $y \in S$ or $S \in T(y)$.

A basic estimate

- Let $T(y)$ be the event that there exists a path of length 2 from 1 to y, and let $M = \max_y \Pr \overline{T(y)}$. Then

$$M - \frac{k}{n-1} \leq \Pr(\mathrm{Diam} > 2) \leq (n-1)M.$$

- Details:
  - If $\operatorname{diam} \operatorname{Cay}(G, S) > 2$, there is y with $S \in \overline{T(y)}$.
  - If $\operatorname{diam} \operatorname{Cay}(G, S) \leq 2$ then for every y, $y \in S$ or $S \in T(y)$.
  - Hence

$$\Pr \overline{T(y)} - \frac{k}{n-1} \leq \Pr(\mathrm{Diam} > 2) \leq \Pr \bigcup_{y \in G^*} \overline{T(y)}.$$

A more detailed estimate

- Let
$$T(x, y) = \{S \mid \{x, x^{-1}y\} \subseteq S\}.$$
be the event that there is a path $1 \to x \to y$.

A more detailed estimate

- Let
$$T(x,y) = \{S \mid \{x, x^{-1}y\} \subseteq S\}.$$
be the event that there is a path $1 \to x \to y$.
- For each $I \subseteq J$ with $|I| = i$, we have
$$\Pr \bigcap_{x \in I} T(x,y) = \binom{n-1}{k}^{-1} \binom{n-1-2i}{k-2i}.$$

A more detailed estimate

- Let
$$T(x,y) = \{S \mid \{x, x^{-1}y\} \subseteq S\}.$$
be the event that there is a path $1 \to x \to y$.
- For each $I \subseteq J$ with $|I| = i$, we have
$$\Pr \bigcap_{x \in I} T(x,y) = \binom{n-1}{k}^{-1} \binom{n-1-2i}{k-2i}.$$
- Suppose we have a set J of t such x's such that the pairs $\{x, x^1 y\}$ are all distinct. Then by inclusion-exclusion
$$\Pr \overline{T(y)} = 1 - \Pr \bigcup_{x \in G^*} T(x,y) \leq 1 - \Pr \bigcup_{x \in J} T(x,y)$$
$$= \binom{n-1}{k}^{-1} \sum_{i=1}^{t} (-1)^{i-1} \binom{t}{i} \binom{n-1-2i}{k-2i}.$$

How big can t be?

Let

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

We know that $M \le p(n-1,k,t)$, where $t = |J|$, and we want to maximize t.

- For elementary abelian 2-groups, $M = p(n-1,k,t)$ and can take $t = \frac{n-1}{2}$.

How big can t be?

Let

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

We know that $M \le p(n-1,k,t)$, where $t = |J|$, and we want to maximize t.

- For elementary abelian 2-groups, $M = p(n-1,k,t)$ and can take $t = \frac{n-1}{2}$.
- For general groups, can take $t = \lfloor \frac{n-1-s}{3} \rfloor$ and s is the number of square roots of y in G.

How big can t be?

Let

$$p(n,k,t) = \binom{n}{k}^{-1} \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i}.$$

We know that $M \le p(n-1,k,t)$, where $t = |J|$, and we want to maximize t.

- For elementary abelian 2-groups, $M = p(n-1,k,t)$ and can take $t = \frac{n-1}{2}$.
- For general groups, can take $t = \lfloor \frac{n-1-s}{3} \rfloor$ and s is the number of square roots of y in G.
- Fact: no nonidentity element in a finite group has more than $3n/4$ square roots. Thus for general groups we have

$$M \le p(n-1,k,t) \qquad \text{where } t = \lfloor \frac{n-4}{12} \rfloor.$$

Overview of results in this section

- We want asymptotics of $p(n, k, \lfloor \frac{n-4}{12} \rfloor)$. The first step is the exponential rate, namely the asymptotics as $n \to \infty$ of $\mathrm{rate} := n^{-1} \log a(n, k, t)$.

Overview of results in this section

- We want asymptotics of $p(n, k, \lfloor \frac{n-4}{12} \rfloor)$. The first step is the exponential rate, namely the asymptotics as $n \to \infty$ of $\mathrm{rate} := n^{-1} \log a(n, k, t)$.
- The linear case $k \sim cn$ is automatically handled by the general multivariate GF asymptotics machinery of Pemantle and Wilson (reported on in Strobl).

Overview of results in this section

- We want asymptotics of $p(n, k, \lfloor \frac{n-4}{12} \rfloor)$. The first step is the exponential rate, namely the asymptotics as $n \to \infty$ of $\mathrm{rate} := n^{-1} \log a(n, k, t)$.
- The linear case $k \sim cn$ is automatically handled by the general multivariate GF asymptotics machinery of Pemantle and Wilson (reported on in Strobl).
- For other growth rates of k we derive uniform asymptotics using methods of Manuel Lladser's thesis (reported on in San Miniato).

Overview of results in this section

- We want asymptotics of $p(n, k, \lfloor \frac{n-4}{12} \rfloor)$. The first step is the exponential rate, namely the asymptotics as $n \to \infty$ of $\text{rate} := n^{-1} \log a(n, k, t)$.
- The linear case $k \sim cn$ is automatically handled by the general multivariate GF asymptotics machinery of Pemantle and Wilson (reported on in Strobl).
- For other growth rates of k we derive uniform asymptotics using methods of Manuel Lladser's thesis (reported on in San Miniato).
- Result: if $k = \omega(\sqrt{n \log n})$ then

$$\Pr(\text{Diam}_{n,k} > 2) \to 0 \quad \text{as } n \to \infty.$$

Convergence is exponentially fast if k is linear in n and superpolynomial otherwise.

The generating function

- Let $a(n,k,t) = \binom{n}{k}^{-1} p(n,k,t)$.

The generating function

- Let $a(n,k,t) = \binom{n}{k}^{-1} p(n,k,t)$.
- Combinatorial interpretation: $a(n,k,t)$ is the number of subsets of $[n]$ of size k not containing any of a fixed collection of t disjoint pairs from $[n]$.

The generating function

- Let $a(n,k,t) = \binom{n}{k}^{-1} p(n,k,t)$.
- Combinatorial interpretation: $a(n,k,t)$ is the number of subsets of $[n]$ of size k not containing any of a fixed collection of t disjoint pairs from $[n]$.
- Note that $2t \leq n, k+t \leq n$ in this interpretation.

The generating function

- Let $a(n,k,t) = \binom{n}{k}^{-1} p(n,k,t)$.
- Combinatorial interpretation: $a(n,k,t)$ is the number of subsets of $[n]$ of size k not containing any of a fixed collection of t disjoint pairs from $[n]$.
- Note that $2t \leq n, k+t \leq n$ in this interpretation.
- The trivariate GF assuming $2t \leq n, k+t \leq n$ is easily derived:

$$\sum_{n,k,t} a(n,k,t) x^n y^k z^t = \frac{1}{1-x(1+y)} \frac{1}{1-zx^2(1+2y)}.$$

Reminder of mvGF techniques

- Assume that locally $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z}) = \sum a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is a quotient of analytic functions.

Reminder of mvGF techniques

- Assume that locally $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z}) = \sum a_{\mathbf{r}}\mathbf{z}^{\mathbf{r}}$ is a quotient of analytic functions.
- Asymptotics of $F(\mathbf{z})$ in direction $\mathbf{r}$ are determined by contributing critical points of the singular variety $\mathcal{V}$ of F.

Reminder of mvGF techniques

- Assume that locally $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z}) = \sum a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is a quotient of analytic functions.
- Asymptotics of $F(\mathbf{z})$ in direction $\mathbf{r}$ are determined by contributing critical points of the singular variety $\mathcal{V}$ of F.
- For generic combinatorial problems, there is exactly one contributing point for each direction. These points satisfy $\mathbf{r} \in \mathrm{dir}(\mathbf{z})$ where $\mathrm{dir}(\mathbf{z})$ is a certain cone defined geometrically.

Reminder of mvGF techniques

- Assume that locally $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z}) = \sum a_{\mathbf{r}}\mathbf{z}^{\mathbf{r}}$ is a quotient of analytic functions.
- Asymptotics of $F(\mathbf{z})$ in direction $\mathbf{r}$ are determined by contributing critical points of the singular variety $\mathcal{V}$ of F.
- For generic combinatorial problems, there is exactly one contributing point for each direction. These points satisfy $\mathbf{r} \in \operatorname{dir}(\mathbf{z})$ where $\operatorname{dir}(\mathbf{z})$ is a certain cone defined geometrically.
- The exponential rate of the coefficients of F in direction $\mathbf{r}$ is given by $-\mathbf{r} \log \mathbf{z}$ where $\mathbf{z}$ is a contributing point for that direction. A full asymptotic expansion can be obtained when the local geometry of $\mathcal{V}$ is nice enough.

Reminder of mvGF techniques

- Assume that locally $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z}) = \sum a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is a quotient of analytic functions.
- Asymptotics of $F(\mathbf{z})$ in direction $\mathbf{r}$ are determined by contributing critical points of the singular variety $\mathcal{V}$ of F.
- For generic combinatorial problems, there is exactly one contributing point for each direction. These points satisfy $\mathbf{r} \in \operatorname{dir}(\mathbf{z})$ where $\operatorname{dir}(\mathbf{z})$ is a certain cone defined geometrically.
- The exponential rate of the coefficients of F in direction $\mathbf{r}$ is given by $-\mathbf{r} \log \mathbf{z}$ where $\mathbf{z}$ is a contributing point for that direction. A full asymptotic expansion can be obtained when the local geometry of $\mathcal{V}$ is nice enough.
- Analyticity means expansions are uniform in large cones.

Details of this mvGF computation

- Here $\mathcal{V}$ consists of of two intersecting smooth hypersurfaces $\mathcal{V}_1, \mathcal{V}_2$ in $\mathbb{C}^3$.

Details of this mvGF computation

- Here $\mathcal{V}$ consists of of two intersecting smooth hypersurfaces $\mathcal{V}_1, \mathcal{V}_2$ in $\mathbb{C}^3$.
- The contributing points all lie on the curve $\mathcal{V}_1 \cap \mathcal{V}_2$. The point $\mathbf{z}$ determines asymptotics in direction $\mathbf{r}$ if and only if $\mathbf{z} \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $\mathbf{r} \in \mathrm{dir}(\mathbf{z})$.

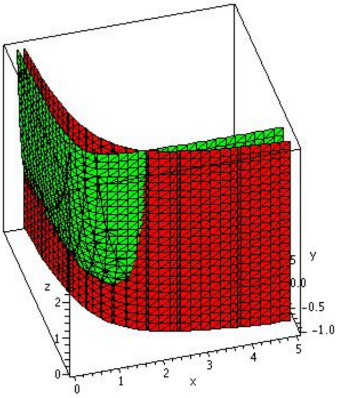
Details of this mvGF computation

- Here $\mathcal{V}$ consists of of two intersecting smooth hypersurfaces $\mathcal{V}_1, \mathcal{V}_2$ in $\mathbb{C}^3$.
- The contributing points all lie on the curve $\mathcal{V}_1 \cap \mathcal{V}_2$. The point $\mathbf{z}$ determines asymptotics in direction $\mathbf{r}$ if and only if $\mathbf{z} \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $\mathbf{r} \in \mathrm{dir}(\mathbf{z})$.
- The contributing point for direction $\mathbf{r}$ is found by solving a system saying that $H_1 = 0, H_2 = 0$ and $\mathbf{r}$ is in the span of $\mathrm{dir}_1(\mathbf{z}), \mathrm{dir}_2(\mathbf{z})$. There is a unique positive solution to this system of 3 polynomial equations in 3 unknowns.

Details of this mvGF computation

- Here $\mathcal{V}$ consists of of two intersecting smooth hypersurfaces $\mathcal{V}_1, \mathcal{V}_2$ in $\mathbb{C}^3$.
- The contributing points all lie on the curve $\mathcal{V}_1 \cap \mathcal{V}_2$. The point $\mathbf{z}$ determines asymptotics in direction $\mathbf{r}$ if and only if $\mathbf{z} \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $\mathbf{r} \in \operatorname{dir}(\mathbf{z})$.
- The contributing point for direction $\mathbf{r}$ is found by solving a system saying that $H_1 = 0, H_2 = 0$ and $\mathbf{r}$ is in the span of $\operatorname{dir}_1(\mathbf{z}), \operatorname{dir}_2(\mathbf{z})$. There is a unique positive solution to this system of 3 polynomial equations in 3 unknowns.
- Upshot: to find asymptotics in direction (n, k, t) we use the contributing point $(1/(1+y), y, (1+y)^2/(1+2y))$ where $y > 0$ and $2(n-k-t)y^2 + (n-3k)y + k = 0$. In the case $k \sim cn, t \sim n/12$, the exponential rate is readily computed.

A picture of the singular variety



Outline of approach in the case of sublinear k

- The contributing points converge to a coordinate axis and the above method requires extension.
- Reduce to a 1-parameter problem: t and k are determined by n, and t is linear in n.
- Use Cauchy's formula in a circle of radius r and convert to a saddle point/stationary phase integral.
- Tune the radius r of the circle of integration in order to capture the correct exponential order.
- Need uniform estimates, obtained by analyticity of the original GF.
- Extract subexponential factors by Laplace's method or similar.

Reduction to a 1-dimensional Fourier-Laplace integral

By expanding the GF, applying Cauchy's integral formula, writing the complex variable in polar form and normalizing we obtain

$$\begin{aligned} a(n,k,t) &= [x^n y^k z^t] \\ &= [y^k](1+y)^{n-2t}(1+2y)^t \\ &= \frac{r^{-k}}{2\pi}\int_{-\pi}^{\pi}(1+re^{i\theta})^{n-2t}(1+2re^{i\theta})^t e^{-ik\theta}\,d\theta \\ &=: (2\pi)^{-1}E(r;n,k,t)I(r;n,k,t) \end{aligned}$$

where

$$E(r;n,k,t) := r^{-k}(1+r)^{n-2t}(1+2r)^t$$

$$I(r;n,k,t) := \int_{-\pi}^{\pi}\left(\frac{1+re^{i\theta}}{1+r}\right)^{n-2t}\left(\frac{1+2re^{i\theta}}{1+2r}\right)^t e^{-ik\theta}\,d\theta.$$

Dealing with $I(r; n, k, t)$

- Write
$$I(r; n, k, t) = \int_{-\pi}^{\pi} e^{-nF(\theta; r, d_1, d_2, d_3)}\, d\theta$$
where
$$F(\theta; r, d_1, d_2, d_3) := id_3\theta - d_1 \log \frac{1 + re^{i\theta}}{1 + r} - d_2 \log \frac{1 + 2re^{i\theta}}{1 + 2r}$$
and $d_1 := (n - 2t)/n, d_2 := t/n, d_3 := k/n$.

Dealing with $I(r; n, k, t)$

- Write

$$I(r; n, k, t) = \int_{-\pi}^{\pi} e^{-nF(\theta; r, d_1, d_2, d_3)}\, d\theta$$

where

$$F(\theta; r, d_1, d_2, d_3) := i d_3 \theta - d_1 \log \frac{1 + re^{i\theta}}{1 + r} - d_2 \log \frac{1 + 2re^{i\theta}}{1 + 2r}$$

and $d_1 := (n - 2t)/n, d_2 := t/n, d_3 := k/n$.

- For each (n, k, t) there is a unique $r = r^* > 0$ for which $\theta = 0$ is a strict local maximum for F.

Dealing with $I(r; n, k, t)$

- Write
$$I(r; n, k, t) = \int_{-\pi}^{\pi} e^{-nF(\theta; r, d_1, d_2, d_3)}\, d\theta$$
where
$$F(\theta; r, d_1, d_2, d_3) := i d_3 \theta - d_1 \log \frac{1 + re^{i\theta}}{1 + r} - d_2 \log \frac{1 + 2re^{i\theta}}{1 + 2r}$$
and $d_1 := (n - 2t)/n, d_2 := t/n, d_3 := k/n$.
- For each (n, k, t) there is a unique $r = r^* > 0$ for which $\theta = 0$ is a strict local maximum for F.
- In the linear case the classical Laplace approximation hypotheses hold uniformly. and $I(r^*; n, k, t) \sim n^{-1/2}$. In the sublinear case $r^* \to 0$, but we get a similar result eventually.

Dealing with $I(r; n, k, t)$

- Write
$$I(r; n, k, t) = \int_{-\pi}^{\pi} e^{-nF(\theta; r, d_1, d_2, d_3)}\, d\theta$$
where
$$F(\theta; r, d_1, d_2, d_3) := id_3\theta - d_1 \log \frac{1 + re^{i\theta}}{1 + r} - d_2 \log \frac{1 + 2re^{i\theta}}{1 + 2r}$$
and $d_1 := (n - 2t)/n, d_2 := t/n, d_3 := k/n$.
- For each (n, k, t) there is a unique $r = r^* > 0$ for which $\theta = 0$ is a strict local maximum for F.
- In the linear case the classical Laplace approximation hypotheses hold uniformly. and $I(r^*; n, k, t) \sim n^{-1/2}$. In the sublinear case $r^* \to 0$, but we get a similar result eventually.
- Upshot: $\operatorname{rate} p(n, k, t) = \operatorname{rate} \binom{n}{k}^{-1} E(r^*; n, k, t)$.

The University of Auckland NEW ZEALAND

Dealing with $E(r; n, k, t)$

- The rate in question is equal to

$$d_3 \log d_3 + (1 - d_3) \log(1 - d_3) - d_3 \log r^* + (1 - 2d_2) \log(1 + r^*) + d_2 \log(1 + 2r^*).$$

Dealing with $E(r; n, k, t)$

- The rate in question is equal to

$$d_3 \log d_3 + (1 - d_3) \log(1 - d_3) - d_3 \log r^* \\ + (1 - 2d_2) \log(1 + r^*) + d_2 \log(1 + 2r^*).$$

- In the linear case $k \sim cn$ this converges to a constant $R(c)$ which is negative for $c > 0$.

Dealing with $E(r; n, k, t)$

- The rate in question is equal to

$$d_3 \log d_3 + (1 - d_3) \log(1 - d_3) - d_3 \log r^* + (1 - 2d_2) \log(1 + r^*) + d_2 \log(1 + 2r^*).$$

- In the linear case $k \sim cn$ this converges to a constant $R(c)$ which is negative for $c > 0$.
- In the sublinear case this is asymptotic to $-d_3^2/12$ as $n \to \infty$.

Dealing with $E(r; n, k, t)$

- The rate in question is equal to

$$d_3 \log d_3 + (1 - d_3) \log(1 - d_3) - d_3 \log r^* \\ + (1 - 2d_2) \log(1 + r^*) + d_2 \log(1 + 2r^*).$$

- In the linear case $k \sim cn$ this converges to a constant $R(c)$ which is negative for $c > 0$.
- In the sublinear case this is asymptotic to $-d_3^2/12$ as $n \to \infty$.
- Putting it all together with the subexponential factors we obtain the advertised result.

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.
- Result: $p(n, k, \frac{n-2}{2})$ has the same exponential growth rate as

$$b(t, k) := 2^k \binom{t}{k} \binom{2t}{k}^{-1}.$$

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.
- Result: $p(n, k, \frac{n-2}{2})$ has the same exponential growth rate as

$$b(t, k) := 2^k \binom{t}{k} \binom{2t}{k}^{-1}.$$

- Stirling's approximation gives the first-order asymptotics.

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.
- Result: $p(n, k, \frac{n-2}{2})$ has the same exponential growth rate as

$$b(t, k) := 2^k \binom{t}{k} \binom{2t}{k}^{-1}.$$

- Stirling's approximation gives the first-order asymptotics.
- We have

$$\text{rate} = (2 - \lambda) \log(1 - \lambda/2) - (1 - \lambda) \log(1 - \lambda).$$

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.
- Result: $p(n, k, \frac{n-2}{2})$ has the same exponential growth rate as

$$b(t, k) := 2^k \binom{t}{k} \binom{2t}{k}^{-1}.$$

- Stirling's approximation gives the first-order asymptotics.
- We have

$$\mathrm{rate} = (2 - \lambda) \log(1 - \lambda/2) - (1 - \lambda) \log(1 - \lambda).$$

  - If $k \sim cn$ with $0 < c < 1/2$, then $\lambda = c$ and $\mathrm{rate} < 0$. Note that $\mathrm{rate} \to 0$ as $\lambda \to 0$.

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.
- Result: $p(n, k, \frac{n-2}{2})$ has the same exponential growth rate as

$$b(t, k) := 2^k \binom{t}{k} \binom{2t}{k}^{-1}.$$

- Stirling's approximation gives the first-order asymptotics.
- We have

$$\mathrm{rate} = (2-\lambda)\log(1-\lambda/2) - (1-\lambda)\log(1-\lambda).$$

  - If $k \sim cn$ with $0 < c < 1/2$, then $\lambda = c$ and $\mathrm{rate} < 0$. Note that $\mathrm{rate} \to 0$ as $\lambda \to 0$.
  - If $\lambda = o(1)$ as $n \to \infty$ then $\mathrm{rate} \sim -\lambda^2/4 + O(\lambda^3)$.

The elementary abelian 2-group case

- We can solve exactly for M, and we have a lower bound. We can also use a simpler generating function.
- Result: $p(n, k, \frac{n-2}{2})$ has the same exponential growth rate as

$$b(t,k) := 2^k \binom{t}{k} \binom{2t}{k}^{-1}.$$

- Stirling's approximation gives the first-order asymptotics.
- We have

$$\text{rate} = (2-\lambda)\log(1-\lambda/2) - (1-\lambda)\log(1-\lambda).$$

  - If $k \sim cn$ with $0 < c < 1/2$, then $\lambda = c$ and $\text{rate} < 0$. Note that $\text{rate} \to 0$ as $\lambda \to 0$.
  - If $\lambda = o(1)$ as $n \to \infty$ then $\text{rate} \sim -\lambda^2/4 + O(\lambda^3)$.
  - Thus we see that $\Pr_{n,k}(\text{Diam} > 2)$ converges to 0 as $n \to \infty$ provided $k = \omega(\sqrt{n \log n})$.

The threshold

- If $k = \omega(\sqrt{n \log n})$ then $\Pr(\mathrm{Diam}_{n,k} > 2)$ converges to zero and if $k = o(\sqrt{n \log n})$ then our upper bound does not. We conjecture the existence of a sharp phase transition.
- Our lower bound even in the abelian case is too weak to prove this.
- Robin Pemantle has indicated an argument based on Poissonization that confirms the conjecture. We await its appearance!

What next?

- The bounds are fairly crude and general - refine them and specialize for various classes of groups.
- Study the phase transition analytically in much more detail.
- Study the behaviour of Diam when $k \sim c\sqrt{n}$ for c close to 1 (the Moore bound).
- Extend to higher values of diameter?
- Generalize and automate the asymptotic analysis used here in the sublinear case.