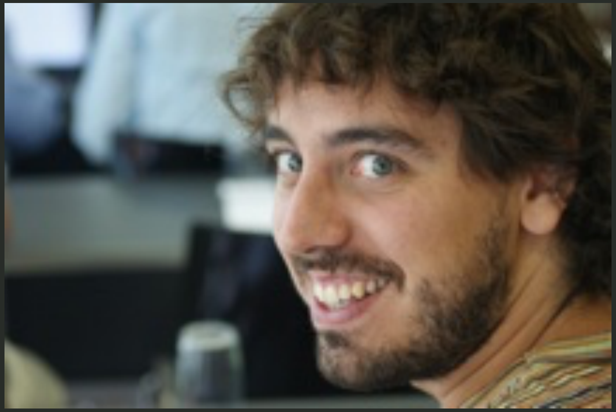
Extreme Statistics of Random Digraphs

Louigi Addario-Berry
McGill / Oxford



Borja Balle
McGill/Lancaster

Guillem Perarnau
McGill/Birmingham

Oxford, May 2015

Model

$r \geq 2$

r-out digraph $D(n,r)$

Vertices $[n] := \{1, 2, \ldots, n\}$

Each vertex has r unif. random out-edges

Model

$r \geq 2$

r-out digraph $D(n,r)$

Vertices $[n] := \{1, 2, \ldots, n\}$

Each vertex has r unif. random out-edges

Concrete Description

$(H_{i,j},\ i \in [n], j \in [r])$ indep. uniform on $[n]$.

$D(n,r)$ has edges

$$\{(i, H_{i,j}),\ i \in [n], j \in [r]\}$$

Model

$r \geq 2$

r-out digraph $D(n,r)$

Vertices $[n] := \{1, 2, \dots, n\}$

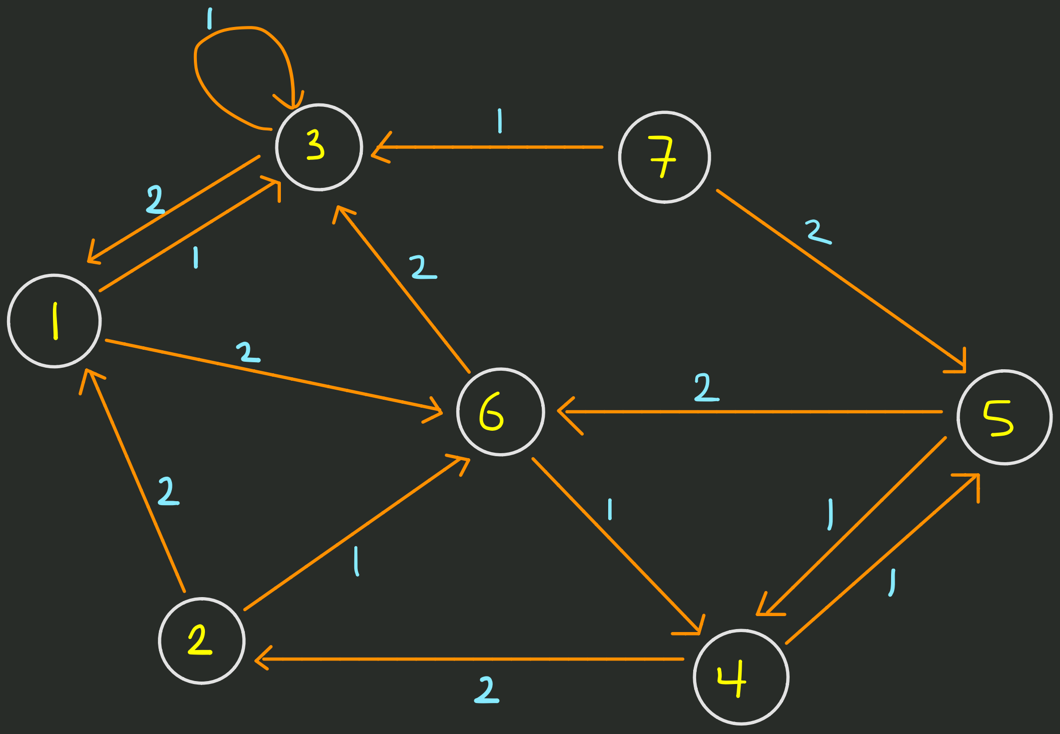
Each vertex has r unif. random out-edges

Concrete Description

$(H_{i,j},\ i \in [n], j \in [r])$ indep. uniform on $[n]$.

$D(n,r)$ has edges

$$\{(i, H_{i,j}),\ i \in [n], j \in [r]\}$$



Model

$r \geq 2$

r-out digraph $D(n,r)$

Vertices $[n] := \{1,2,\dots,n\}$

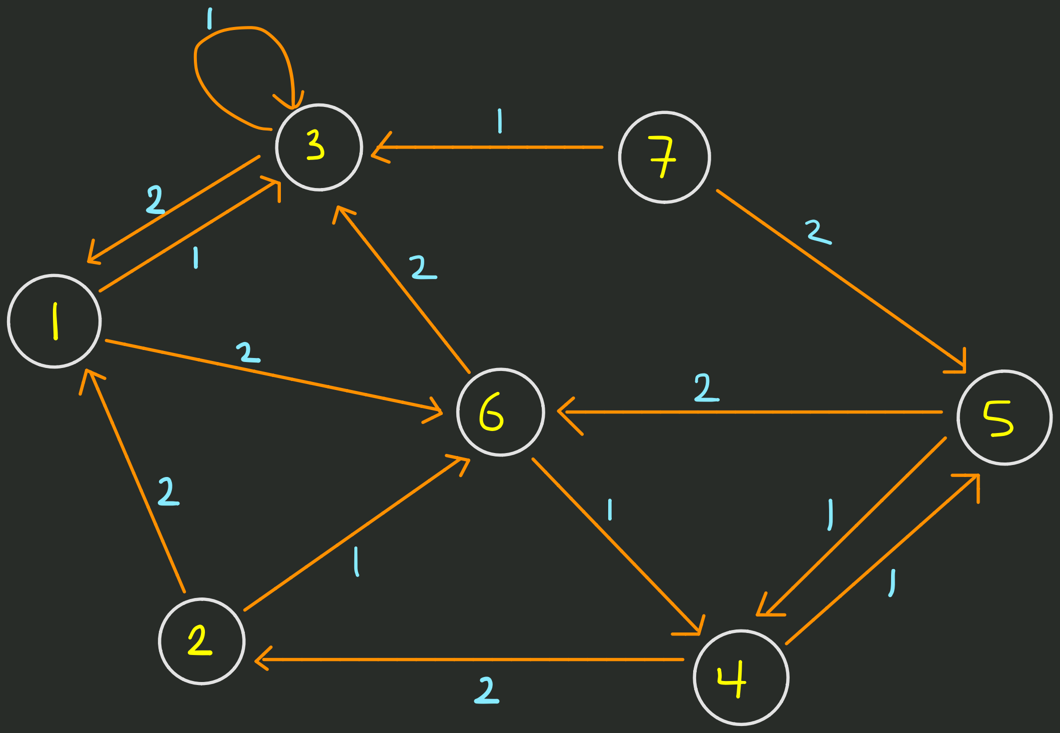
Each vertex has r unif. random out-edges

Concrete Description

$(H_{i,j}\,,\ i\in[n], j\in[r])$ indep. uniform on $[n]$.

$D(n,r)$ has edges $\{(i, H_{i,j}),\ i\in[n], j\in[r]\}$

Aim: Detailed understanding of metric, spectral properties of $D(n,r)$, for large n



Model

$r \geq 2$

r-out digraph $D(n,r)$

Vertices $[n] := \{1,2,\dots,n\}$

Each vertex has r unif. random out-edges

Concrete Description

$(H_{i,j},\ i\in[n], j\in[r])$ indep. uniform on $[n]$.

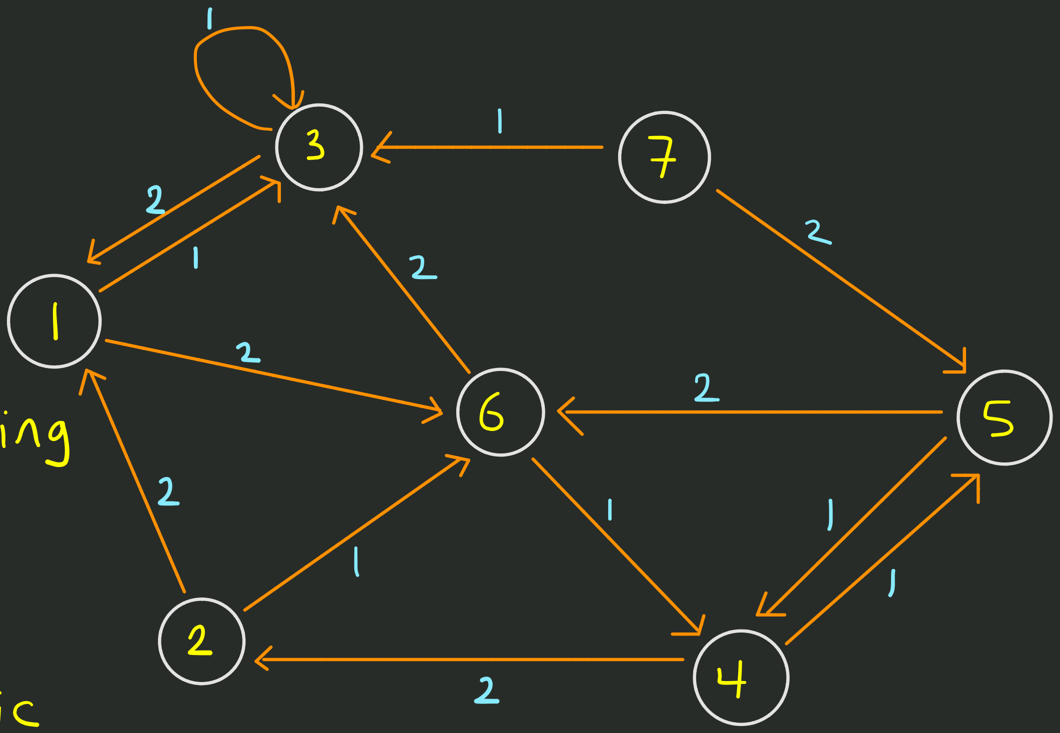
$D(n,r)$ has edges $\{(i, H_{i,j}),\ i\in[n], j\in[r]\}$

Aim: Detailed understanding of metric, spectral properties of $D(n,r)$, for large n

Motivation:

① Probabilistic Machine Learning of Deterministic Finite Automata/Regular Languages

② Directed graphs more realistic than undirected! We should be studying them.



Results 1: Diameter

Def: For a digraph D and $u,v \in D$,

$$\text{dist}(u,v) = \text{min \# edges on directed } u\text{-}v \text{ path}$$
$$= \infty \text{ if no } uv \text{ path.}$$

Results 1: Diameter

Def: For a digraph D and $u, v \in D$,

$\mathrm{dist}(u,v) =$ min # edges on directed u-v path
$\quad = \infty$ if no uv path.

$\mathrm{diam}(D) = \max(\mathrm{dist}(u,v) : uv \in D, \mathrm{dist}(u,v) < \infty)$

Results 1: Diameter

Def: For a digraph D and $u, v \in D$,

$$\mathrm{dist}(u,v) = \text{min \# edges on directed } u\text{-}v \text{ path}$$
$$= \infty \text{ if no } uv \text{ path.}$$

$$\mathrm{diam}(D) = \max(\mathrm{dist}(u,v) : uv \in D,\ \mathrm{dist}(u,v) < \infty)$$

$$D_0(n,r) = \text{Largest } \underbrace{\text{strongly-connected}}_{\exists \text{ path from any vertex to any other}} \text{ component of } D(n,r)$$

Results 1: Diameter

Def: For a digraph D and $u,v \in D$,

$$\begin{aligned} \mathrm{dist}(u,v) &= \min \#\text{ edges on directed } u\text{-}v \text{ path} \\ &= \infty \text{ if no } uv \text{ path.} \\ \mathrm{diam}(D) &= \max\left(\mathrm{dist}(u,v) : uv \in D,\ \mathrm{dist}(u,v) < \infty\right) \end{aligned}$$

$D_0(n,r)$ = Largest $\underbrace{\text{strongly-connected}}_{\exists \text{ path from any vertex to any other}}$ component of $D(n,r)$

Theorem (A-B, Balle, Perarnau 2015):

As $n \to \infty$,

$$\mathrm{diam}(D(n,r)) = (1 + c_r + o_p(1)) \log_r(n)$$

$$\mathrm{diam}(D_0(n,r)) = (1 + c_r + o_p(1)) \log_r(n)$$

Results 1: Diameter

Def: For a digraph D and $u, v \in D$,

$$\mathrm{dist}(u,v) = \min \# \text{ edges on directed } u\text{-}v \text{ path}$$
$$= \infty \text{ if no } uv \text{ path.}$$
$$\mathrm{diam}(D) = \max\left(\mathrm{dist}(u,v) : uv \in D,\ \mathrm{dist}(u,v) < \infty\right)$$

$D_0(n,r)$ = Largest $\underbrace{\text{strongly-connected}}_{\exists \text{ path from any vertex to any other}}$ component of $D(n,r)$

Theorem (A-B, Balle, Perarnau 2015):

As $n \to \infty$,

$$\mathrm{diam}(D(n,r)) = (1 + c_r + o_p(1)) \log_r(n)$$
$$\mathrm{diam}(D_0(n,r)) = (1 + c_r + o_p(1)) \log_r(n)$$

This means $\forall\, \varepsilon > 0$ $\mathbb{P}\left(\left|\mathrm{diam}(D(n,r)) - (1+c_r)\log_r n\right| > \varepsilon \log_r n\right) \to 0$

Results 1: Diameter

Def: For a digraph D and $u,v \in D$,

$$\mathrm{dist}(u,v) = \min \#\text{ edges on directed } u\text{-}v \text{ path}$$
$$= \infty \text{ if no } uv \text{ path.}$$
$$\mathrm{diam}(D) = \max(\mathrm{dist}(u,v) : u,v \in D,\ \mathrm{dist}(u,v) < \infty)$$

$D_0(n,r) =$ Largest $\underbrace{\text{strongly-connected}}_{\exists \text{ path from any vertex to any other}}$ component of $D(n,r)$

Theorem (A-B, Balle, Perarnau 2015):

As $n \to \infty$,

$$\mathrm{diam}(D(n,r)) = (1 + c_r + o_p(1)) \log_r(n)$$
$$\mathrm{diam}(D_0(n,r)) = (1 + c_r + o_p(1)) \log_r(n)$$

This means $\forall \varepsilon > 0$ $\mathbb{P}(|\mathrm{diam}(D(n,r)) - (1+c_r)\log_r n| > \varepsilon \log_r n) \to 0$

where $c_r = (\log r)/(s r - \log r)$,

$s = \mathbb{P}(\text{Poisson}(r) \text{ Galton-Watson proc. survives})$

$1 - s = e^{-rs}$

Random walks on graphs

Simple random walk on an undirected graph

$G=(V,E) \longrightarrow \deg(v) = \#$ neighbours of $v = \#\{w : \{v,w\} \in E\}$

$X = (X_i, i \geq 0)$ Markov chain,

$$p_{vw} = \mathbb{P}(X_{i+1}=w \mid X_i = v) = \begin{cases} 1/\deg(v) & \text{if } \{v,w\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Random walks on graphs

Simple random walk on an undirected graph

$G=(V,E) \longrightarrow \deg(v) = \#$ neighbours of $v = \#\{w : \{v,w\} \in E\}$

$X=(X_i, i \geq 0)$ Markov chain,

$$p_{vw} = \mathbb{P}(X_{i+1}=w \mid X_i = v) = \begin{cases} 1/\deg(v) & \text{if } \{v,w\}\in E \\ 0 & \text{otherwise} \end{cases}$$

Stationary dist: G connected $\Rightarrow$ stat. dist. π unique, $\pi(v) = \deg(v)/2|E|$.

Markov chain reversible

Random walks on graphs

Simple random walk on an undirected graph

$G=(V,E) \longrightarrow \deg(v) = \#$ neighbours of $v = \#\{w : \{v,w\}\in E\}$

$X=(X_i, i\geq 0)$ Markov chain,

$$p_{vw} = \mathbb{P}(X_{i+1}=w \mid X_i=v) = \begin{cases} 1/\deg(v) & \text{if } \{v,w\}\in E \\ 0 & \text{otherwise} \end{cases}$$

Stationary dist: G connected $\Rightarrow$ stat. dist. π unique, $\pi(v) = \deg(v)/2|E|$.

Markov chain reversible

Simple random walk on a directed graph

$D=(V,E) \rightarrow \deg^+(v) = \#\{w : vw\in E\}$ $\quad \deg^-(v) = \#\{u : uv\in E\}$

$X=(X_i, i\geq 0)$ $\quad p_{vw} = \begin{cases} 1/\deg^+(v) & \text{if } vw\in E \\ 0 & \text{otherwise} \end{cases}$

Random walks on graphs

Simple random walk on an undirected graph

$G=(V,E) \longrightarrow \deg(v) = \#$ neighbours of $v = \#\{w : \{v,w\}\in E\}$

$X=(X_i, i\geq 0)$ Markov chain,

$$p_{vw} = \mathbb{P}(X_{i+1}=w \mid X_i=v) = \begin{cases} 1/\deg(v) & \text{if } \{v,w\}\in E \\ 0 & \text{otherwise} \end{cases}$$

Stationary dist: G connected $\Rightarrow$ stat. dist. π unique, $\pi(v) = \deg(v)/2|E|$.

Markov chain reversible

Simple random walk on a directed graph

$D=(V,E) \rightarrow \deg^+(v) = \#\{w : vw\in E\}$ $\quad \deg^-(v) = \#\{u : uv\in E\}$

$X=(X_i, i\geq 0)$ $\quad p_{vw} = \begin{cases} 1/\deg^+(v) & \text{if } vw\in E \\ 0 & \text{otherwise} \end{cases}$

Stationary dist: G strongly connected $\Rightarrow$ stat. dist. π unique, $\pi(v) = ??$

Random walks on graphs

Simple random walk on an undirected graph

$G=(V,E) \longrightarrow \deg(v) = \#$ neighbours of $v = \#\{w : \{v,w\} \in E\}$

$X = (X_i, i \geq 0)$ Markov chain,

$$p_{vw} = \mathbb{P}(X_{i+1}=w \mid X_i = v) = \begin{cases} 1/\deg(v) & \text{if } \{v,w\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Stationary dist: G connected $\Rightarrow$ stat. dist. π unique, $\pi(v) = \deg(v)/2|E|$.

Markov chain reversible

Simple random walk on a directed graph

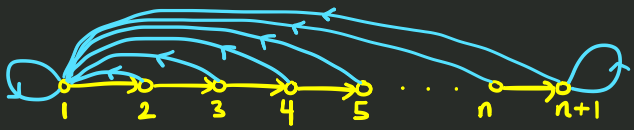
$D=(V,E) \rightarrow \deg^+(v) = \#\{w : vw \in E\}$ $\quad \deg^-(v) = \#\{u : uv \in E\}$

$X = (X_i, i \geq 0)$ $\quad p_{vw} = \begin{cases} 1/\deg^+(v) & \text{if } vw \in E \\ 0 & \text{otherwise} \end{cases}$

Stationary dist: G strongly connected $\Rightarrow$ stat. dist. π unique, $\pi(v) = ??$

⚠ No simple description. Natural guess: $\pi(v) = \deg^-(v)/|E|$.

(Counter) Example:



Guess: $\pi(1) = \frac{1}{2}$, $\pi(i) = \frac{1}{2n}$, $i \geq 2$

Truth: $\pi(n+1) = 2^{-n}$, $\pi(i) = 2^{-i}$, $i \leq n$

Results 2: Stationary distribution

Proposition (Grusho, 1973): The largest component $D_0(n,r)$ of $D(n,r)$ is with high probability the unique recurrent component.

Corollary: Simple random walk has unique stat. dist π whp.

Results 2: Stationary distribution

Proposition (Grusho, 1973): The largest component $D_0(n,r)$ of $D(n,r)$ is with high probability the unique recurrent component.

<u>Corollary</u>: Simple random walk has unique stat. dist π whp.

Theorem (A-B, Balle, Perarnau):

$$\pi_{max} = n^{-1+o_p(1)}, \quad \pi_{min} = n^{-(1+c_r)+o_p(1)}$$

Results 2: Stationary distribution

Proposition (Grusho, 1973): The largest component $D_0(n,r)$ of $D(n,r)$ is with high probability the unique recurrent component.

Corollary: Simple random walk has unique stat. dist π whp.

Theorem (A-B, Balle, Perarnau):

$$\pi_{max} = n^{-1+o_p(1)}, \quad \pi_{min} = n^{-(1+C_r)+o_p(1)}$$

i.e. $\forall \varepsilon > 0$, $\mathbb{P}(\pi_{max} > n^{-1+\varepsilon} \text{ or } \pi_{max} < n^{-1-\varepsilon}) \to 0$

$\mathbb{P}(\pi_{min} > n^{-(1+C_r)+\varepsilon} \text{ or } \pi_{min} < n^{-(1+C_r)-\varepsilon}) \to 0$

Results 2: Stationary distribution

Proposition (Grusho, 1973): The largest component $D_0(n,r)$ of $D(n,r)$ is with high probability the unique recurrent component.

Corollary: Simple random walk has unique stat. dist π whp.

Theorem (A-B, Balle, Perarnau):

$$\pi_{max} = n^{-1+o_p(1)}, \quad \pi_{min} = n^{-(1+C_r)+o_p(1)}$$

i.e. $\forall \varepsilon > 0$, $\mathbb{P}(\pi_{max} > n^{-1+\varepsilon} \text{ or } \pi_{max} < n^{-1-\varepsilon}) \to 0$

$\mathbb{P}(\pi_{min} > n^{-(1+C_r)+\varepsilon} \text{ or } \pi_{min} < n^{-(1+C_r)-\varepsilon}) \to 0$

Remark: C_r same as in theorem on diameter

A People's History of the Diameter of Random Graphs

$D(n,r)$ (1973) Trakhtenbrot & Bardin: $\mathrm{diam}(D(n,r)) = O_p(\log n)$

$G(n,r)$ (1982) Bollobás & de la Vega: $\mathbb{P}(|\mathrm{diam}(G(n,r) - \log_{r-1}(\log n)| \in [-5,5]) \to 1.$

$G(n,\frac{\lambda}{n})$ (1974) Burtin: $\lambda(n)/\log n \to \infty$: Two-point concentration.

$$\mathbb{P}(\mathrm{diam}(G(n,\tfrac{\lambda}{n}) \in \{\lfloor \log_\lambda n \rfloor, \lfloor \log_\lambda n \rfloor + 1\}) \to 1$$

(1981) Bollobas: $\lambda(n) \geq \log^3 n$: "Explicit" two point concentration.

$$\mathbb{P}(\mathrm{diam}(G(n,\tfrac{\lambda}{n}) = \lfloor \log_\lambda n \rfloor) = p_1(n), \quad \mathbb{P}(\mathrm{diam}(G(n,\tfrac{\lambda}{n}) = \lfloor \log_\lambda n \rfloor + 1) = p_2(n),$$

$$p_1(n) + p_2(n) \to 1.$$

(2010) Riordan & Wormald: $G(n,\frac{\lambda}{n})$

$\lambda > 1$ fixed: $\mathrm{diam}(G(n,\lambda/n)) = (1 + 2c_\lambda)\log_\lambda n + O_p(1)$

$\lambda \to \infty$: 2-point concentration.

$\lambda(n) = 1 + o(1)$: $\mathrm{diam}(G(n,\lambda/n)) = d(n) + O_p(\frac{1}{\lambda - 1})$

$$d(n) = (1 + o(1)) \frac{3}{\lambda - 1} \log((\lambda - 1)^3 n).$$

(see also Ding-Kim-Lubetzky-Peres 2010)

Random edge weights

- Bhamidi & van der Hofstad → Conv. in dist., complete graph
- Amini & Lelarge → $G(n,\vec{r})$, $C \log n + o_p(\log n)$
- Ding-Kim-Lubetzky-Peres → $G(n,r)$, $C \log n + O_p(\log\log n)$

A People's History of the Diameter of Random Graphs

$D(n,r)$ (1973) Trakhtenbrot & Bardin: $\mathrm{diam}(D(n,r)) = O_p(\log n)$

$G(n,r)$ (1982) Bollobás & de la Vega: $\mathbb{P}(|\mathrm{diam}(G(n,r) - \log_{r-1}(\log n)| \in [-5,5]) \to 1.$

$G(n,\frac{\lambda}{n})$ (1974) Burtin: $\lambda(n)/\log n \to \infty$: Two-point concentration.

$$\mathbb{P}(\mathrm{diam}(G(n,\tfrac{\lambda}{n}) \in \{\lfloor \log_\lambda n \rfloor, \lfloor \log_\lambda n \rfloor + 1\}) \to 1$$

(1981) Bollobas: $\lambda(n) \geq \log^3 n$: "Explicit" two point concentration.

$$\mathbb{P}(\mathrm{diam}(G(n,\tfrac{\lambda}{n}) = \lfloor \log_\lambda n \rfloor) = p_1(n),\ \mathbb{P}(\mathrm{diam}(G(n,\tfrac{\lambda}{n}) = \lfloor \log_\lambda n \rfloor + 1) = p_2(n),$$

$$p_1(n) + p_2(n) \to 1.$$

(2010) Riordan & Wormald: $G(n,\frac{\lambda}{n})$

$\lambda > 1$ fixed: $\mathrm{diam}(G(n,\lambda/n)) = (1 + 2c_\lambda)\log_\lambda n + O_p(1)$

$\lambda \to \infty$: 2-point concentration.

$\lambda(n) = 1 + o(1)$: $\mathrm{diam}(G(n,\lambda/n)) = d(n) + O_p(\frac{1}{\lambda - 1})$

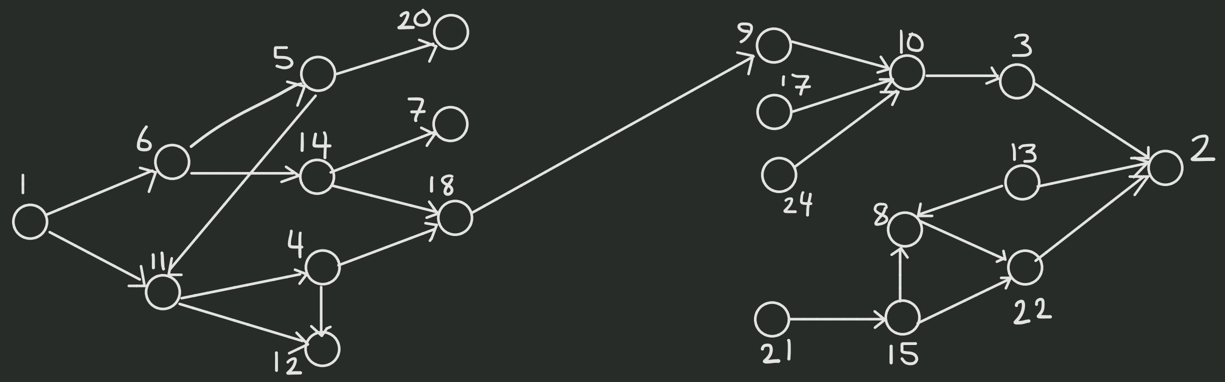
$$d(n) = (1 + o(1)) \frac{3}{\lambda - 1} \log((\lambda - 1)^3 n).$$

(see also Ding-Kim-Lubetzky-Peres 2010)

Random edge weights

- Bhamidi & van der Hofstad → Conv. in dist., complete graph
- Amini & Lelarge → $G(n,\vec{r})$, $C\log n + o_p(\log n)$
- Ding-Kim-Lubetzky-Peres → $G(n,r)$, $C\log n + O_p(\log\log n)$

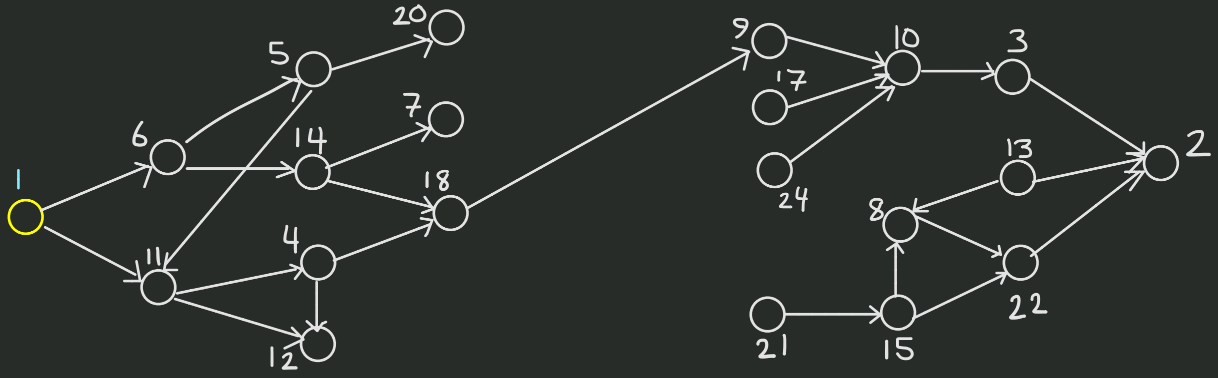
Breadth-first search



Aim: Find a shortest path from 1 to 2.

Breadth-first search

Outward

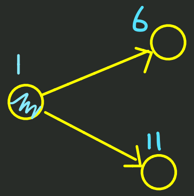


Breadth-first search

Outward

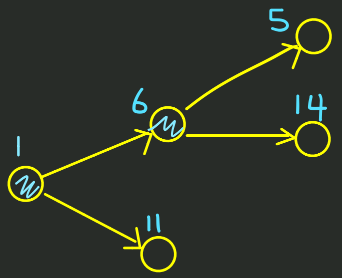
Breadth-first search

Outward



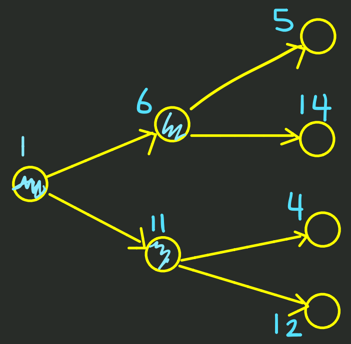
Breadth-first search

Outward



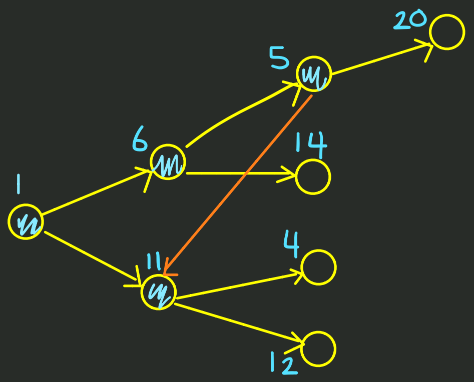
Breadth-first search

Outward



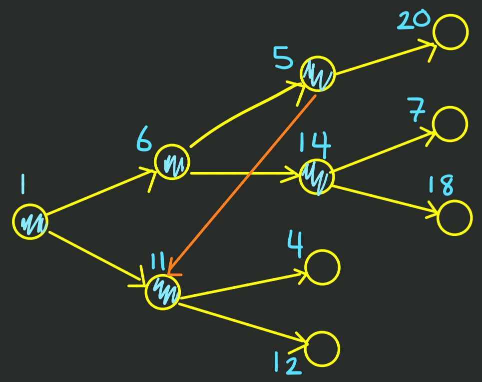
Breadth-first search

Outward



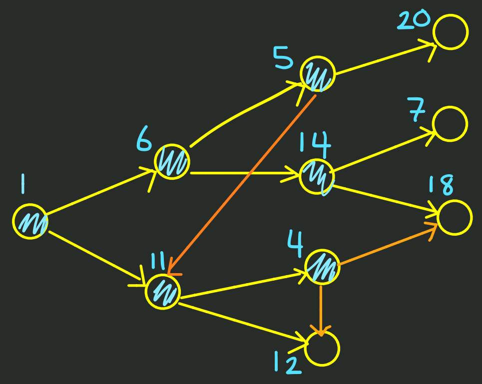
Breadth-first search

Outward



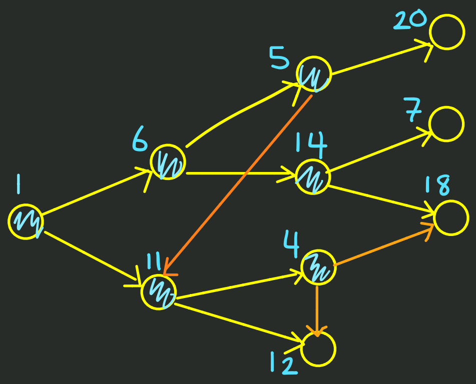
Breadth-first search

Outward



Breadth-first search

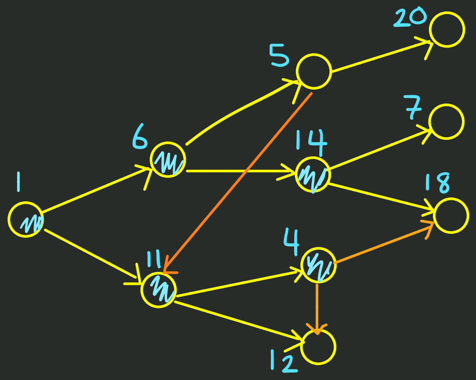
Outward

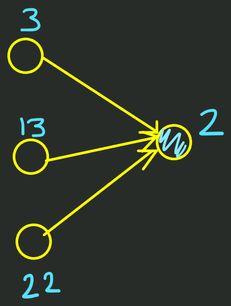


Inward

Breadth-first search

Outward

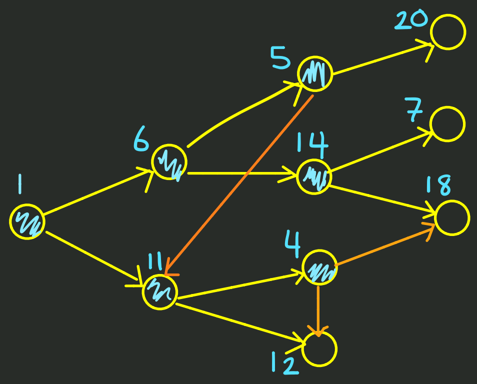


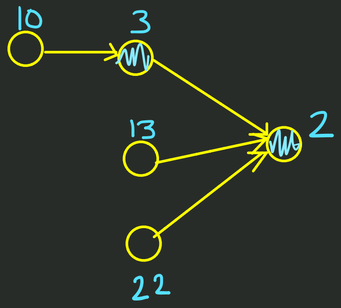


Inward

Breadth-first search

Outward

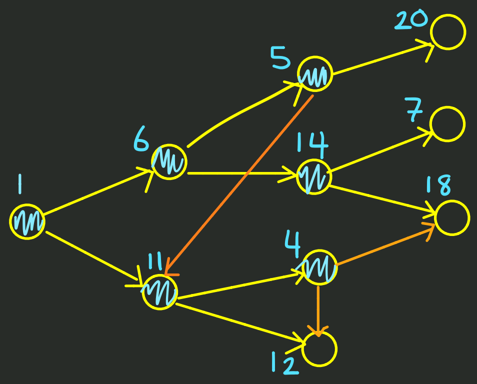


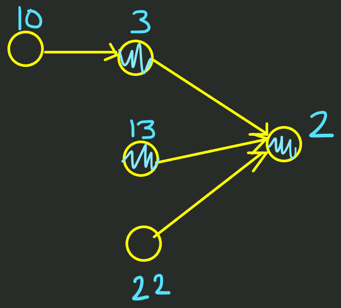


Inward

Breadth-first search

Outward

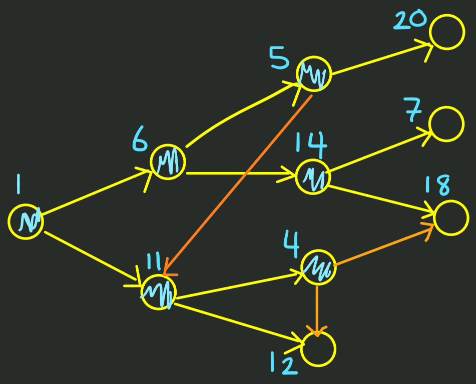


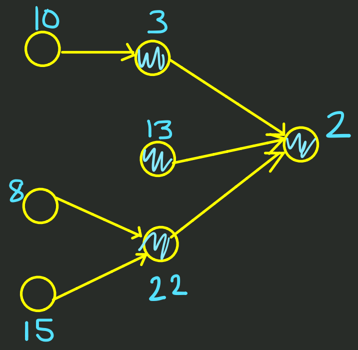


Inward

Breadth-first search

Outward

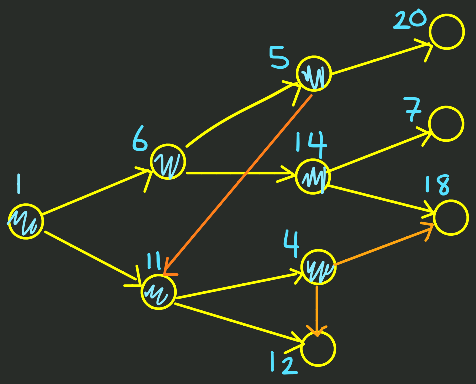


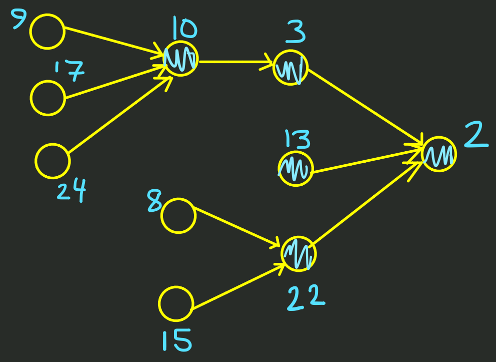


Inward

Breadth-first search

Outward

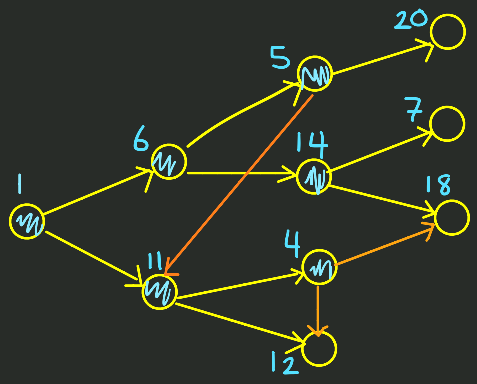


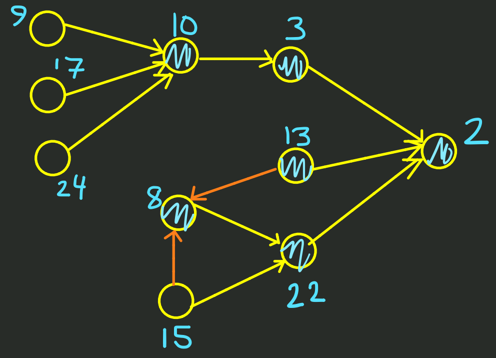


Inward

Breadth-first search

Outward

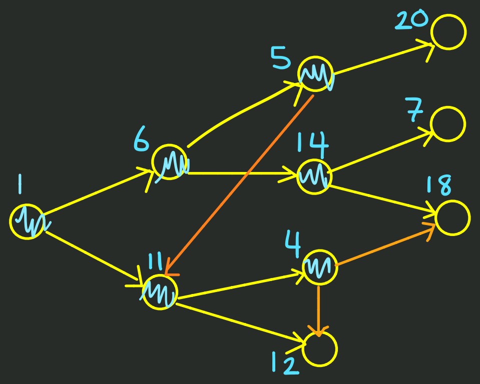


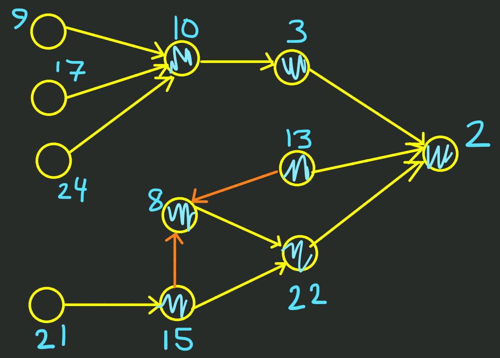


Inward

Breadth-first search

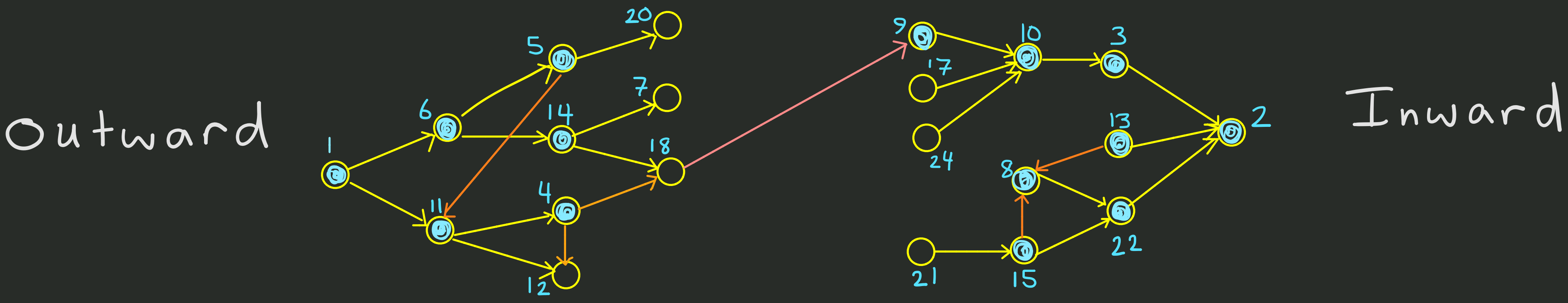
Outward



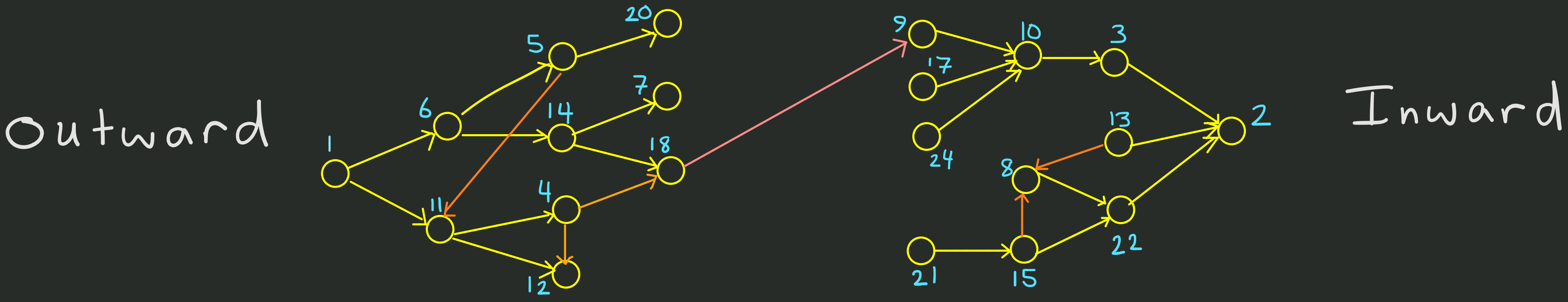


Inward

Breadth-first search

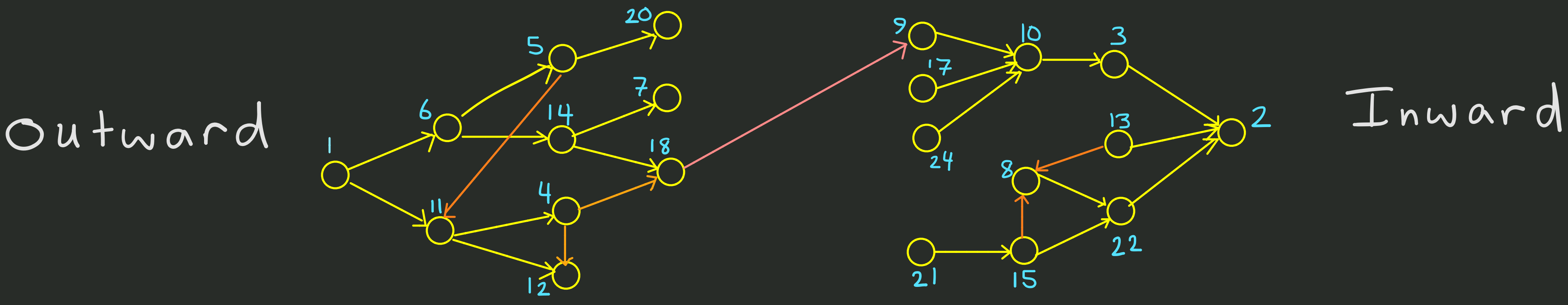


Breadth-first search



Observation : $\boxed{\mathrm{dist}(u,v) = \min\{\, j+k : N_j^+(u) \cap N_k^-(v) \neq \phi \,\}}$

Breadth-first search

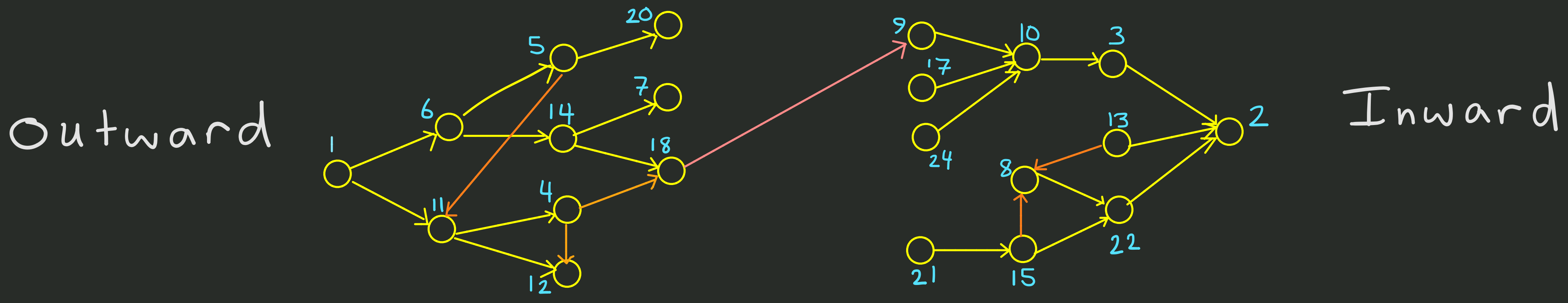


Observation: $\boxed{\mathrm{dist}(u,v) = \min\{ j+k : N_j^+(u) \cap N_k^-(v) \neq \phi \}}$

Heuristics:

1) oBFS looks like BFS on complete r-ary tree
2) iBFS looks like BFS on a Poisson(r) Galton-Watson tree.
3) out-nbhds much more regular than in-nbhds so in-nbhds cause any fluctuations in distances.

Breadth-first search



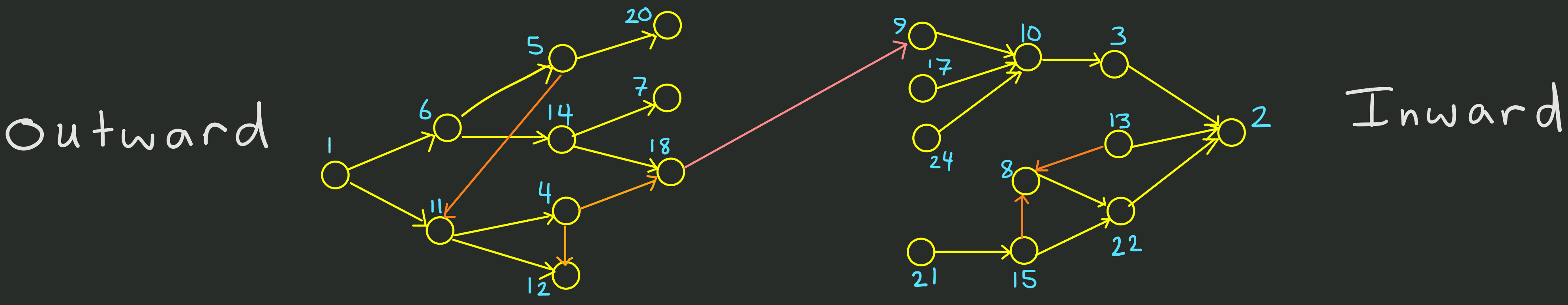
Observation: $\boxed{\mathrm{dist}(u,v) = \min\{ j+k : N_j^+(u) \cap N_k^-(v) \neq \phi \}}$

Heuristics:
1) oBFS looks like BFS on complete r-ary tree
2) iBFS looks like BFS on a Poisson(r) Galton-Watson tree.
3) out-nbhds much more regular than in-nbhds so in-nbhds cause any fluctuations in distances.

$\boxed{\text{For node } u, \text{ let } k(u) = \min\{ k : |N_k^-(u)| \geq \log^4 n \}}$

Prop: $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

Breadth-first search



Observation: $\boxed{\mathrm{dist}(u,v) = \min\{ j+k : N_j^+(u) \cap N_k^-(v) \neq \phi \}}$

Heuristics:
1) oBFS looks like BFS on complete r-ary tree
2) iBFS looks like BFS on a Poisson(r) Galton-Watson tree.
3) out-nbhds much more regular than in-nbhds so in-nbhds cause any fluctuations in distances.

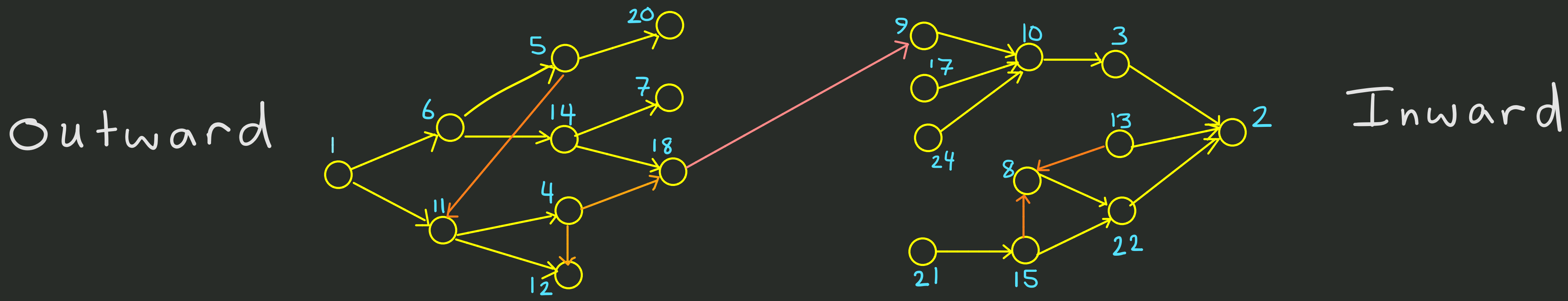
$\boxed{\text{For node } u, \text{ let } k(u) = \min\{ k : |N_k^-(u)| \geq \log^4 n \}}$

Prop: $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

Proof idea:

If $|N_j^+(v)| \simeq r^j$ then $\mathbb{P}(N_j^+(v) \cap N_{k(u)}^-(u) \neq \phi) \simeq \left(1 - \frac{\log^4 n}{n}\right)^{r^j} \begin{cases} \simeq 1 \text{ when } r^j \ll \frac{n}{\log^4 n} \\ \simeq 0 \text{ when } r^j \gg \frac{n}{\log^4 n} \end{cases}$

Breadth-first search



Observation: $\boxed{\mathrm{dist}(u,v) = \min\{ j+k : N_j^+(u) \cap N_k^-(v) \neq \phi \}}$

Heuristics:
1) oBFS looks like BFS on complete r-ary tree
2) iBFS looks like BFS on a Poisson(r) Galton-Watson tree.
3) out-nbhds much more regular than in-nbhds so in-nbhds cause any fluctuations in distances.

$$\boxed{\text{For node } u, \text{ let } k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}}$$

Prop: $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

Proof idea:

If $|N_j^+(v)| \simeq r^j$ then $\mathbb{P}(N_j^+(v) \cap N_{k(u)}^-(u) \neq \phi) \simeq \left(1 - \frac{\log^4 n}{n}\right)^{r^j} \begin{cases} \simeq 1 & \text{when } r^j \ll \frac{n}{\log^4 n} \\ \simeq 0 & \text{when } r^j \gg \frac{n}{\log^4 n} \end{cases}$

So first have $N_j^+(v) \cap N_{k(u)}^-(u) \neq \phi$ when $r^j \simeq \frac{n}{\log^4 n}$; $j = \log_r n - \log_r(\log^4 n)$

At first I was afraid...

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$

$\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\underline{\text{Prop}} : \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

At first I was afraid...

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\underline{\text{Prop}}: \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

Idea (Athreya & Ney; Kingman; McDiarmid; Riordan & Wormald; Berestycki, Gantert, Mörters & Sidorova; ...)

Long-term growth is due to surviving lineages.

Let $T \sim$ Poisson(r) GW, root ρ

$$\boxed{\mathbb{P}(T = \infty) = s \qquad 1-s = e^{-rs}}$$

Say $v \in T$ survives if v has ∞ many descendants.

At first I was afraid...

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\underline{\text{Prop}}: \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

<u>Idea</u> (Athreya & Ney; Kingman; McDiarmid; Riordan & Wormald; Berestycki, Gantert, Mörters & Sidorova; ...)

Long-term growth is due to <u>surviving</u> lineages.

Let $T \sim$ Poisson(r) GW, root ρ $\qquad$ $\boxed{\mathbb{P}(T=\infty) = s \qquad 1-s = e^{-rs}}$

Say $v \in T$ survives if v has ∞ many descendants.

NB: $\#\{\text{survivors in gen } g\}$ is <u>increasing</u> in g

At first I was afraid...

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\underline{\text{Prop}}: \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

Idea (Athreya & Ney; Kingman; McDiarmid; Riordan & Wormald; Berestycki, Gantert, Mörters & Sidorova; ...)

Long-term growth is due to surviving lineages.

Let $T \sim$ Poisson(r) GW, root ρ

$$\boxed{\mathbb{P}(T=\infty) = s \qquad 1-s = e^{-rs}}$$

Say $v \in T$ survives if v has ∞ many descendants.

NB: $\#\{$survivors in gen $g\}$ is increasing in g

$\mathbb{P}(\rho$ has a surviving children$) = \mathbb{P}(\text{Poisson}(rs) = a)$

$\mathbb{P}(\rho$ has a surviving children $| \rho$ survives$) = s^{-1} \cdot \mathbb{P}(\text{Poisson}(rs) = a)$

At first I was afraid...

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\text{Prop: } \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

Idea (Athreya & Ney; Kingman; McDiarmid; Riordan & Wormald; Berestycki, Gantert, Mörters & Sidorova; ...)

Long-term growth is due to surviving lineages.

Let $T \sim$ Poisson(r) GW, root ρ

$$\boxed{\mathbb{P}(T = \infty) = s \qquad 1-s = e^{-rs}}$$

Say $v \in T$ survives if v has ∞ many descendants.

NB: $\#\{\text{survivors in gen } g\}$ is increasing in g

$\mathbb{P}(\rho \text{ has } a \text{ surviving children}) = \mathbb{P}(\text{Poisson}(rs) = a)$

$\mathbb{P}(\rho \text{ has } a \text{ surviving children} \mid \rho \text{ survives}) = s^{-1} \cdot \mathbb{P}(\text{Poisson}(rs) = a)$

$$\begin{aligned} \mathbb{P}(\rho \text{ has } 1 \text{ surviving child} \mid \rho \text{ survives}) &= s^{-1}\, \mathbb{P}(\text{Poisson}(rs) = 1) \\ &= s^{-1} \cdot (rs e^{-rs}) = r e^{-rs} \end{aligned}$$

At first I was afraid...

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\underline{\text{Prop}} : \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

<u>Idea</u> (Athreya & Ney; Kingman; McDiarmid; Riordan & Wormald; Berestycki, Gantert, Mörters & Sidorova; ...)

Long-term growth is due to <u>surviving</u> lineages.

Let $T \sim$ Poisson(r) GW, root ρ

$$\boxed{\mathbb{P}(T=\infty) = s \qquad 1-s = e^{-rs}}$$

Say $v \in T$ survives if v has ∞ many descendants.

NB: $\#\{$survivors in gen $g\}$ is in<u>creasing</u> in g

$\mathbb{P}(\rho$ has a surviving children$) = \mathbb{P}(\text{Poisson}(rs) = a)$

$\mathbb{P}(\rho$ has a surviving children $| \rho$ survives$) = s^{-1} \cdot \mathbb{P}(\text{Poisson}(rs) = a)$

$\mathbb{P}(\rho$ has 1 surviving child $| \rho$ survives$) = s^{-1} \mathbb{P}(\text{Poisson}(rs) = 1)$

$= s^{-1} \cdot (rse^{-rs}) = re^{-rs}$

So $\mathbb{P}($exactly 1 survivor in gen $g) = \mathbb{P}(1$ survivor in gen $1)^g = (re^{-rs})^g$

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1))\log_r n$

$$\boxed{\underline{\text{Prop}} : \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

$$\mathbb{P}(\text{exactly } 1 \text{ survivor in gen } g) = (r e^{-rs})^g \cong \mathbb{P}(\text{exactly } 1 \text{ in}\underline{\text{dividual}} \text{ in gen. } g)$$

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\text{Prop: } \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

$\mathbb{P}(\text{exactly 1 survivor in gen } g) = (r e^{-rs})^g \overset{\smile}{=} \mathbb{P}(\text{exactly 1 individual in gen. } g)$

Back to in-neighbourhoods of u.

Idea: (Riordan-Wormald) For $k-t$ large, least $\underbrace{\text{costly}}_{\text{unlikely!}}$ way to have $0 < N_k^-(u) < r^t$ is to have $N_{k-t}^-(u) = 1$.

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

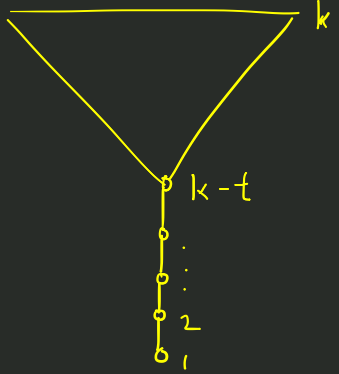
$$\boxed{\underline{\text{Prop}}: \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

$\mathbb{P}(\text{exactly } 1 \text{ survivor in gen } g) = (r e^{-rs})^g \cong \mathbb{P}(\text{exactly } 1 \text{ in}\underline{\text{dividual}} \text{ in gen. } g)$

Back to in-neighbourhoods of u.

Idea: (Riordan-Wormald) For $k-t$ large, least $\underbrace{\text{costly}}_{\text{unlikely!}}$ way to have $0 < N_k^-(u) < r^t$ is to have $N_{k-t}^-(u) = 1$.



$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\text{Prop: } \max_u k(u) = (1+o(1))\, c_r \log n}$$

First try to understand a single vertex u.

$\mathbb{P}$(exactly 1 survivor in gen g) $= (r e^{-rs})^g \simeq \mathbb{P}$(exactly 1 individual in gen. g)

Back to in-neighbourhoods of u.

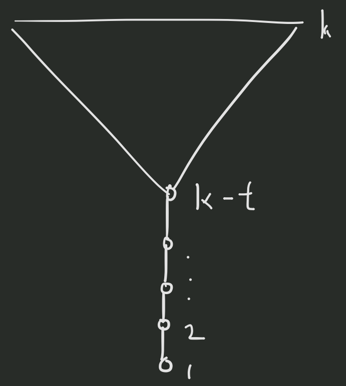
Idea: (Riordan-Wormald) For $k-t$ large, least $\underbrace{\text{costly}}_{\text{unlikely!}}$ way to have $0 < N_k^-(u) < r^t$ is to have $N_{k-t}^-(u) = 1$.

Cost: $\mathbb{P}(0 < |N_k^-(u)| < r^t) \simeq \mathbb{P}(|N_{k-t}^-(u)| = 1,\ 0 < N_k^-(u) < r^t)$

$$\simeq \mathbb{P}(|N_{k-t}^-(u)| = 1,\ N_k^-(u) > 0)$$

$$\simeq \mathbb{P}(|N_{k-t}^-(u)| = 1) \cdot \mathbb{P}(\mathrm{Po}(r)\ \mathrm{GW}\ \text{survives})$$

$$\simeq (r e^{-rs})^{k-t} \cdot s$$



$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\underline{\text{Prop}} : \max_u k(u) = (1+o(1))\, c_r \log n}$$

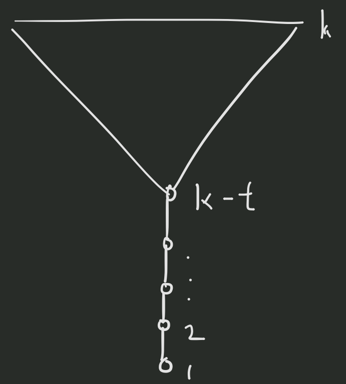
First try to understand a single vertex u.

$\mathbb{P}(\text{exactly 1 survivor in gen } g) = (r e^{-rs})^g \cong \mathbb{P}(\text{exactly 1 } \underline{\text{individual}} \text{ in gen. } g)$

Back to in-neighbourhoods of u.

<u>Idea</u>: (Riordan-Wormald) For $k-t$ large, least $\underbrace{\text{costly}}_{\text{unlikely!}}$ way to have $0 < N_k^-(u) < r^t$ is to have $N_{k-t}^-(u) = 1$.

<u>Cost</u>:
$$
\begin{aligned}
\mathbb{P}(0 < |N_k^-(u)| < r^t) &\cong \mathbb{P}(|N_{k-t}^-(u)| = 1,\ 0 < N_k^-(u) < r^t) \\
&\cong \mathbb{P}(|N_{k-t}^-(u)| = 1,\ N_k^-(u) > 0) \\
&\cong \mathbb{P}(|N_{k-t}^-(u)| = 1) \cdot \mathbb{P}(\mathrm{Po}(r) \text{ GW survives}) \\
&\simeq (r e^{-rs})^{k-t} \cdot s
\end{aligned}
$$



<u>Prop</u> For $k-t$ large and k, t small $\leftarrow (k \cdot 2^t = o(n^{1/2}))$ enough that GW approximation is good,

$$\mathbb{P}(0 < |N_k^-(u)| < r^t) = \Theta(1) \cdot (r e^{-rs})^{k-t}$$

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\text{Prop: } \max_u k(u) = (1+o(1))\, c_r \log n}$$

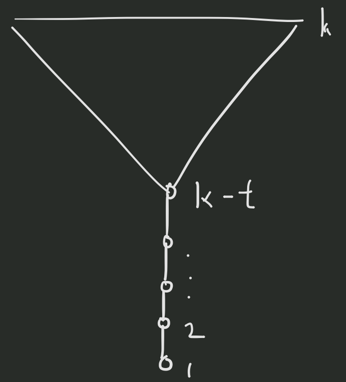
First try to understand a single vertex u.

$\mathbb{P}(\text{exactly 1 survivor in gen } g) = (r e^{-rs})^g \cong \mathbb{P}(\text{exactly 1 individual in gen. } g)$

Back to in-neighbourhoods of u.

Idea: (Riordan-Wormald) For $k-t$ large, least costly way to have $0 < N_k^-(u) < r^t$
unlikely! is to have $N_{k-t}^-(u) = 1$.

Cost:
$$\begin{aligned}\mathbb{P}(0<|N_k^-(u)|<r^t) &\cong \mathbb{P}(|N_{k-t}^-(u)=1,\ 0<N_k^-(u)<r^t) \\ &\cong \mathbb{P}(|N_{k-t}^-(u)|=1,\ N_k^-(u)>0) \\ &\cong \mathbb{P}(|N_{k-t}^-(u)|=1) \cdot \mathbb{P}(\mathrm{Po}(r) \text{ GW survives}) \\ &\simeq (r e^{-rs})^{k-t} \cdot s\end{aligned}$$



Prop For $k-t$ large and k, t small $\leftarrow$ $(k \cdot 2^t = o(n^{1/2}))$
enough that GW approximation is good,

$$\mathbb{P}(0<|N_k^-(u)|<r^t) = \Theta(1) \cdot (r e^{-rs})^{k-t}$$

Here: $r^t = \log^4 n$. Then $(r e^{-rs})^{k-t} = \frac{1}{n}$ when $k = c_r \log_r n + O(\log\log n)$

Bounding the diameter: wrap-up.

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

$$\boxed{\text{Prop: } \max_u k(u) = (1+o(1))\, c_r \log n}$$

$$\mathbb{P}(0 < |N_k^-(u)| < r^t) = \Theta(1) \cdot (r e^{-rs})^{k-t}$$

Here: $r^t = \log^4 n$. Then $(r e^{-rs})^{k-t} = \frac{1}{n}$ when $k = c_r \log_r n + O(\log\log n)$

Bounding the diameter: wrap-up.

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1))\log_r n$

$$\boxed{\underline{\text{Prop}}: \max_u k(u) = (1+o(1))\, c_r \log n}$$

$$\mathbb{P}(0<|N_k^-(u)|<r^t) = \Theta(1)\cdot (re^{-rs})^{k-t}$$

<u>Here</u>: $r^t = \log^4 n$. Then $(re^{-rs})^{k-t} = \frac{1}{n}$ when $k = c_r \log_r n + O(\log\log n)$

So if $k > c_r \log_r n + C\log\log n$ then $\mathbb{E}[\#\{u : k(u) \geq k\}] \longrightarrow 0$

$k < c_r \log_r n - C\log\log n$ then $\mathbb{E}[\#\{u : k(u) \geq k\}] \longrightarrow \infty$

Bounding the diameter: wrap-up.

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\mathrm{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1))\log_r n$

$$\boxed{\underline{\text{Prop}}: \max_u k(u) = (1+o(1))\, c_r \log n}$$

$$\mathbb{P}(0 < |N_k^-(u)| < r^t) = \Theta(1)\cdot (re^{-rs})^{k-t}$$

Here: $r^t = \log^4 n$. Then $(re^{-rs})^{k-t} = \frac{1}{n}$ when $k = c_r \log_r n + O(\log\log n)$

so if $k > c_r \log_r n + C\log\log n$ then $\mathbb{E}[\#\{u : k(u) \geq k\}] \longrightarrow 0$

$k < c_r \log_r n - C\log\log n$ then $\mathbb{E}[\#\{u : k(u) \geq k\}] \longrightarrow \infty$

First moment method $\Rightarrow$ $k(u) \leq c_r \log_r n + O(\log\log n)$ whp

Second moment method $\Rightarrow$ $k(n) \geq c_r \log_r n - O(\log\log n)$ whp

$\downarrow$

Count $\mathbb{E}[\#$ pairs u,v s.t. $k(u) \geq k,\ k(v) \geq k]$

Bounding Π_{max} and Π_{min}

View $N^-_{\leq k}(u)$ as a maze.

(Treasure at u.)

Maze is h-hard if every path from $N^-_k(u)$ to u has $\geq h$ traps.

Bounding Π_{max} and Π_{min}

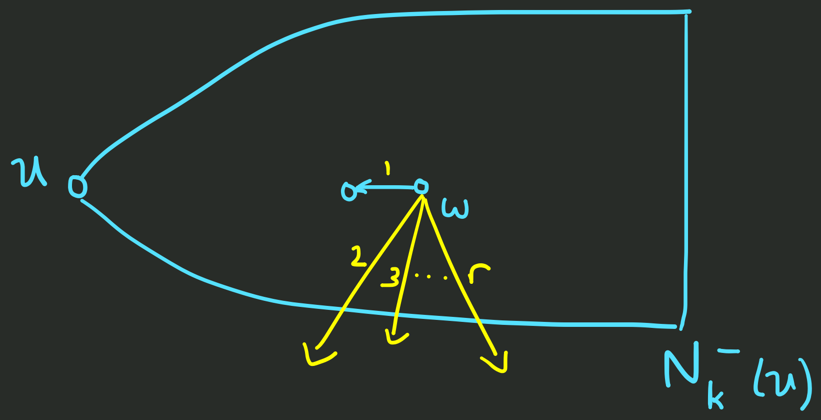
View $N^-_{\leq k}(u)$ as a maze.

(Treasure at u.)

Maze is h-hard if every path from $N^-_k(u)$ to u has $\geq h$ traps.

Trap:

A vertex w with $|N^+(w) \cap N^-_{\leq k}(u)| = 1$.



Bounding $\Pi_{\max}$ and $\Pi_{\min}$

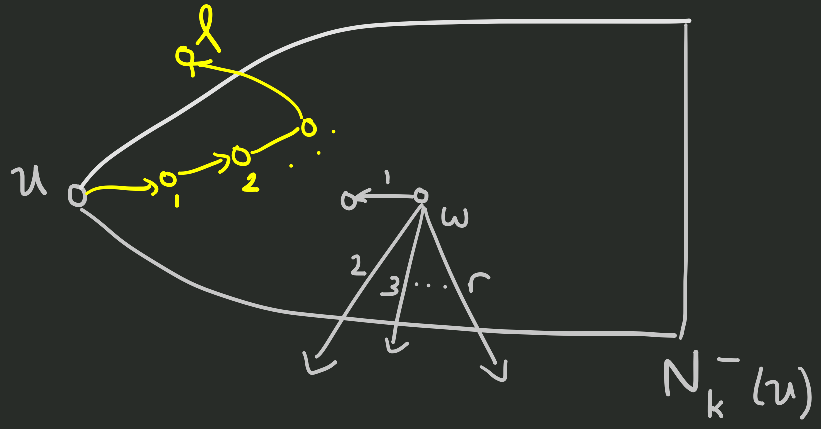
View $N^-_{\leq k}(u)$ as a maze.

(Treasure at u.)

Maze is h-hard if every path from $N^-_k(u)$ to u has $\geq h$ traps.

Trap:

A vertex w with $|N^+(w) \cap N^-_{\leq k}(u)| = 1$.



Prop: If $N^-_{\leq k}(u)$ is h-hard and $\exists$ at least one path of length ℓ from u to $N^-_{\leq k}(u)^c$ then

$\pi(u) \leq r^{\ell-h}$

Bounding Π_{max} and Π_{min}

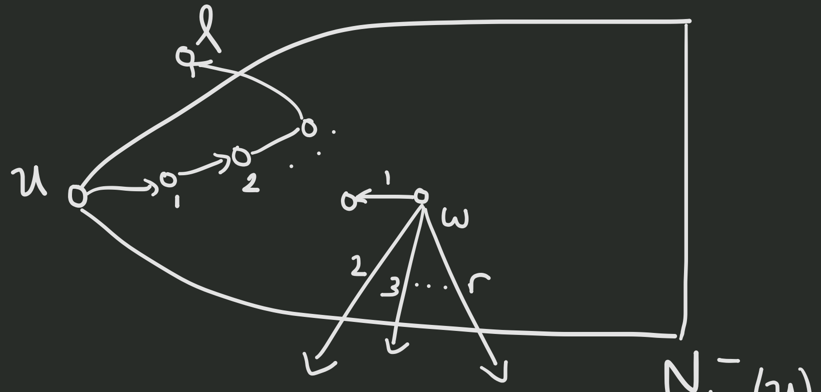
View $N^-_{\leq k}(u)$ as a maze.

(Treasure at u.)

Maze is h-hard if every path from $N^-_k(u)$ to u has $\geq h$ traps.

Trap:

A vertex w with $|N^+(w) \cap N^-_{\leq k}(u)| = 1$.



Prop: If $N^-_{\leq k}(u)$ is h-hard and $\exists$ at least one path of length ℓ from u to $N^-_{\leq k}(u)^c$ then

$$\pi(u) \leq r^{\ell-h}$$

Proof: In this case

$$\mathbb{E}_u(\tau_u^+) \geq \mathbb{P}_u(X_\ell \notin N^-_{\leq k}(u)) \cdot \inf_{v \in N^-_k(u)} \mathbb{E}_w(\tau_u)$$

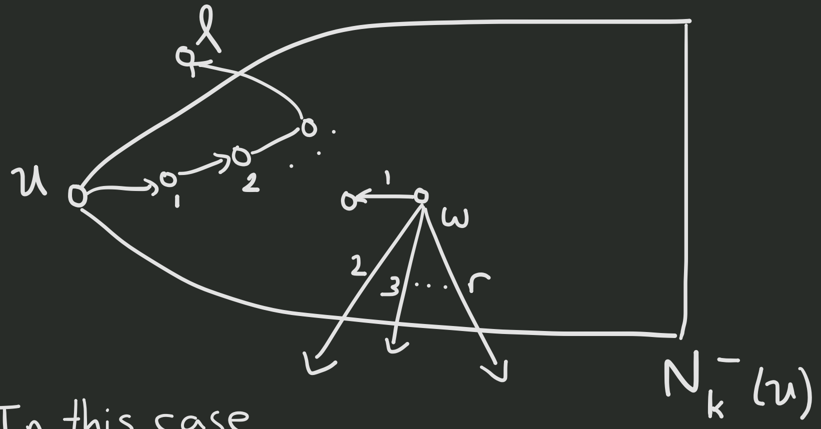
$$\geq r^{-\ell} \cdot r^h$$

∎

Bounding π_{max} and π_{min}

View $N^-_{\le k}(u)$ as a maze.
(Treasure at u.)

Maze is h-hard if every path from $N^-_k(u)$ to u has $\geq h$ traps.



Trap:
A vertex w with $|N^+(w) \cap N^-_{\le k}(u)| = 1$.

Prop: If $N^-_{\le k}(u)$ is h-hard and
$\exists$ at least one path of length ℓ
from u to $N^-_{\le k}(u)^c$ then
$\pi(u) \le r^{\ell - h}$

Proof: In this case

$$\mathbb{E}_u(\tau_u^+) \geq \mathbb{P}_u(X_\ell \notin N^-_{\le k}(u)) \cdot \inf_{v \in N_k^-(u)} \mathbb{E}_w(\tau_u) \geq r^{-\ell} \cdot r^h$$

☒

Fact: Whp, for all u

For $k = (1-\varepsilon)\log_r n$, $N^-_{\le k}(u)$ essentially a tree so is whp h-hard ($h = (1-2\varepsilon)\log_r n$)

Also whp $\exists$ a path of length $\le \log\log n$ to $N^-_{\le k}(u)^c$

Bounding $\pi_{\max}$ and $\pi_{\min}$

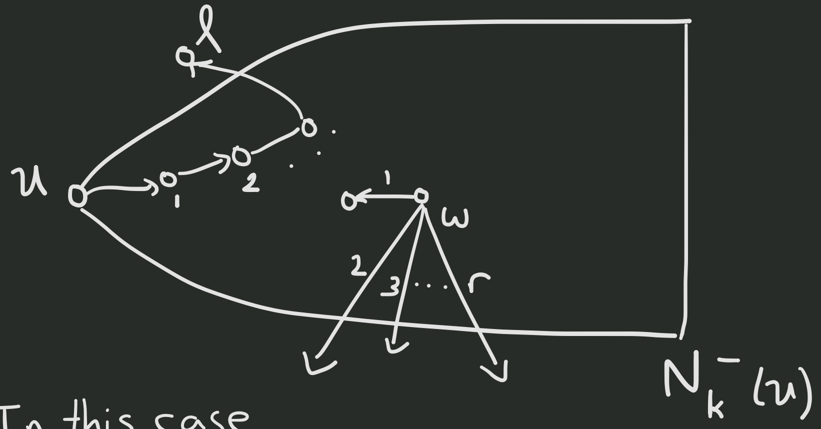
View $N^-_{\le k}(u)$ as a maze.

(Treasure at u.)

Maze is h-hard if every path from $N^-_k(u)$ to u has $\ge h$ traps.

Trap:

A vertex w with $|N^+(w) \cap N^-_{\le k}(u)| = 1$.



Prop: If $N^-_{\le k}(u)$ is h-hard and $\exists$ at least one path of length ℓ from u to $N^-_{\le k}(u)^c$ then

$$\pi(u) \le r^{\ell - h}$$

Proof: In this case

$$\mathbb{E}_u(\tau_u^+) \ge \mathbb{P}_u(X_\ell \notin N^-_{\le k}(u)) \cdot \inf_{v \in N_k^-(u)} \mathbb{E}_w(\tau_u) \ge r^{-\ell} \cdot r^h$$

∎

Fact: Whp, for all u

For $k = (1-\varepsilon)\log_r n$, $N^-_{\le k}(u)$ essentially a tree so is whp h-hard ($h = (1-2\varepsilon)\log_r n$)

Also whp $\exists$ a path of length $\le \log\log n$ to $N^-_{\le k}(u)^c$

So $\pi_{\max} = \max_u \pi(u) \le r^{\ell - k} < n^{-(1-3\varepsilon)}$ whp.

Bounding π_{max} and π_{min}

Lower bound: If D is strongly connected, r-reg, diameter d then

$$\pi_{min} \geq \frac{1}{1+d\cdot r^d}$$

Bounding π_{max} and π_{min}

Lower bound: If D is strongly connected, r-reg, diameter d then

$$\pi_{min} \geq \frac{1}{1+d\cdot r^d}$$

Proof: For all v, $\mathbb{P}_v(\tau_u \leq d) \geq r^{-d}$

So $\mathbb{E}_u(\tau_u^+) \leq 1 + d\cdot r^d$

So $\pi(u) \geq (1+d\cdot r^d)^{-1}$ $\blacksquare$

Bounding $\pi_{\max}$ and $\pi_{\min}$

Lower bound: If D is strongly connected, r-reg, diameter d then

$$\pi_{\min} \geq \frac{1}{1+d\cdot r^d}$$

Proof: For all v, $\mathbb{P}_v(\tau_u \leq d) \geq r^{-d}$

So $\mathbb{E}_u(\tau_u^+) \leq 1 + d\cdot r^d$

So $\pi(u) \geq (1+d\cdot r^d)^{-1}$ ∎

LOWER BOUND ON $\pi_{\min}$ FOLLOWS SINCE $d = (1+o(1))(1+c_r)\log_r n$

Bounding $\pi_{\max}$ and $\pi_{\min}$

Lower bound: If D is strongly connected, r-reg, diameter d then

$$\pi_{\min} \geq \frac{1}{1 + d \cdot r^d}$$

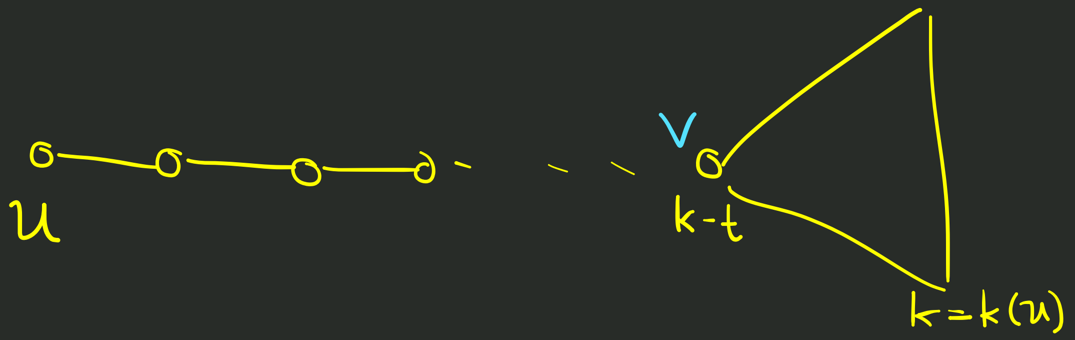
Proof: For all v, $\mathbb{P}_v(\tau_u \leq d) \geq r^{-d}$

So $\mathbb{E}_u(\tau_u^+) \leq 1 + d \cdot r^d$

So $\pi(u) \geq (1 + d \cdot r^d)^{-1}$

LOWER BOUND ON $\pi_{\min}$ FOLLOWS SINCE $d = (1+o(1))(1+c_r)\log_r n$

Upper bound: Consider vertex maximizing $k(u)$.



Bounding π_{max} and π_{min}

Lower bound: If D is strongly connected, r-reg, diameter d then

$$\pi_{min} \geq \frac{1}{1+d\cdot r^d}$$

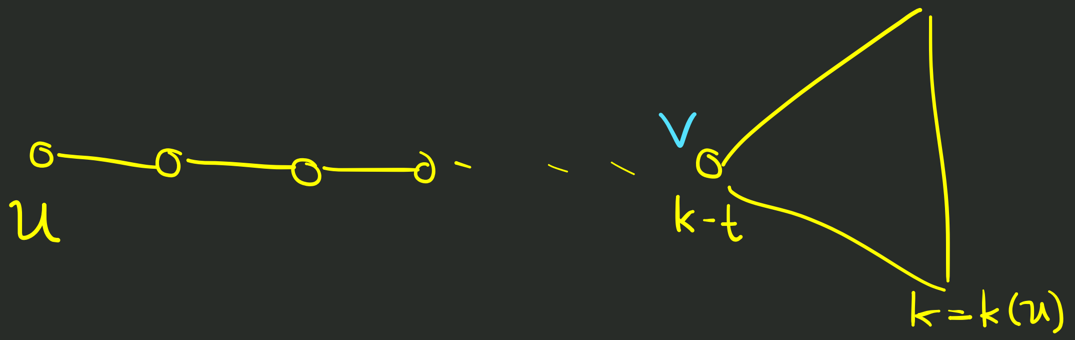
Proof: For all v, $\mathbb{P}_v(\tau_u \leq d) \geq r^{-d}$

$$\text{So } \mathbb{E}_u(\tau_u^+) \leq 1 + d\cdot r^d$$

$$\text{So } \pi(u) \geq (1+d\cdot r^d)^{-1}$$

LOWER BOUND ON π_{min} FOLLOWS SINCE $d = (1+o(1))(1+c_r)\log_r n$

Upper bound: Consider vertex maximizing $k(u)$.



Must visit v many $[r^{k-t}]$ times before reaching u

Bounding $\pi_{\max}$ and $\pi_{\min}$

Lower bound: If D is strongly connected, r-reg, diameter d then

$$\pi_{\min} \geq \frac{1}{1+d\cdot r^d}$$

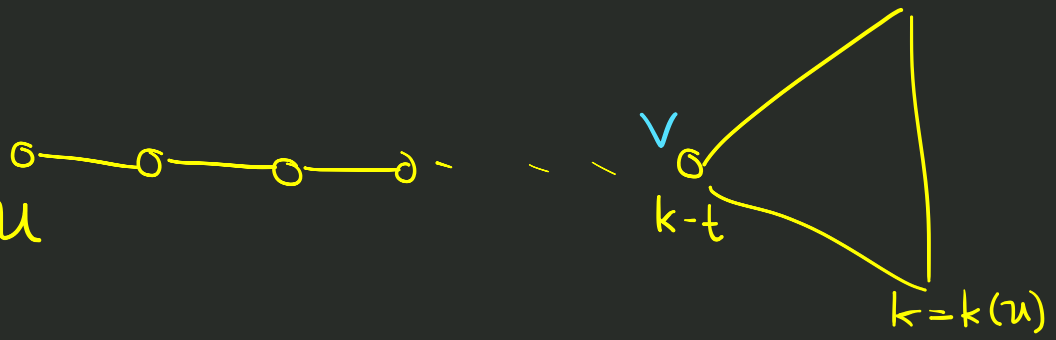
Proof: For all v, $\mathbb{P}_v(\tau_u \leq d) \geq r^{-d}$

So $\mathbb{E}_u(\tau_u^+) \leq 1 + d\cdot r^d$

So $\pi(u) \geq (1+d\cdot r^d)^{-1}$

LOWER BOUND ON $\pi_{\min}$ FOLLOWS SINCE $d = (1+o(1))(1+c_r)\log_r n$

Upper bound: Consider vertex maximizing $k(u)$.



Must visit v many $[r^{k-t}]$ times before reaching u

So $\mathbb{E}_u[\tau_u^+] \gtrsim r^{k-t}\,\mathbb{E}_v[\tau_v^+] \gtrsim r^{k-t}\cdot 1/n$

$= n^{-(1+c_r+o_p(1))}$.

Thank you!

$k(u) = \min\{k : |N_k^-(u)| \geq \log^4 n\}$ $\qquad$ $\operatorname{diam}(D(n,r)) = \max_u k(u) + (1+o_p(1)) \log_r n$

Prop: $\max_u k(u) = (1+o(1))\, c_r \log n$

First try to understand a single vertex u.

Idea (Athreya & Ney; Kingman; McDiarmid; Riordan & Wormald):

Study BP of survivors.

Let $T \sim$ Poisson(r) GW, $\mathbb{P}(T=\infty)=s$. $\qquad 1-s=e^{-rs}$.

Say $v \in T$ survives if v has ∞ many descendants.

Let ρ be root of T, $Z =$ # surviving children of ρ

$\mathbb{P}(\rho \text{ has } a \text{ surviving children}) = \mathbb{P}(\text{Poisson}(rs) = a)$

$\mathbb{P}(\rho \text{ has } a \text{ surviving children} \mid \rho \text{ survives}) = s^{-1} \cdot \mathbb{P}(\text{Poisson}(rs)=a)$

$$\begin{aligned}\mathbb{P}(\rho \text{ has } 1 \text{ surviving child} \mid \rho \text{ survives}) &= s^{-1}\, \mathbb{P}(\text{Poisson}(rs)=1)\\ &= s^{-1} \cdot (rs\, e^{-rs}) = r\, e^{-rs}\\ &= r(1-s)\end{aligned}$$

Idea: For $k-t$ large, least $\underbrace{\text{costly}}_{\text{unlikely!}}$ way to have $0 < N_k^-(u) < r^t$

is to have $N_{k-t}^-(u) = 1$.

Cost: $$\begin{aligned}\mathbb{P}(|N_k^-(u)| \leq 2^t) &\cong \mathbb{P}(|N_{k-t}^-(u)=1,\ 0 < N_k^-(u) < r^t)\\ &\cong \mathbb{P}(|N_{k-t}^-(u)=1,\ N_k^-(u) > 0)\\ &\cong \mathbb{P}(|N_{k-t}^-(u)|=1) \cdot \mathbb{P}(\text{Po}(r) \text{ GW survives})\end{aligned}$$

survivors in gen k is increasing in k