The graph structure of a deterministic automaton chosen at random

Xing Shi Cai   Luc Devroye

School of Computer Science
McGill University

RS&A 2015

k-out digraph

- A *k*-out digraph is a digraph in which each vertex has *k* labeled out-arcs.

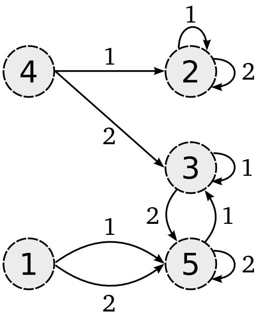


Figure 1 : A 2-out digraph with 5 vertices.

k-out digraph

- A k-out digraph is a digraph in which each vertex has k labeled out-arcs.
- Let $\mathcal{D}_{n,k}$ be a uniform random k-out digraph of n vertices.

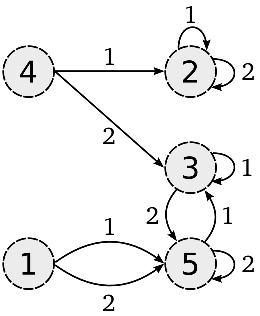


Figure 1 : A 2-out digraph with 5 vertices.

k-out digraph

- A *k-out digraph* is a digraph in which each vertex has k labeled out-arcs.
- Let $\mathcal{D}_{n,k}$ be a uniform random k-out digraph of n vertices.
- Each arc chooses its endpoint independently and uniformly at random.

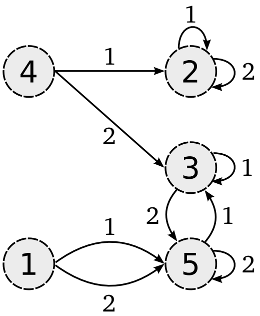


Figure 1 : A 2-out digraph with 5 vertices.

k-out digraph

- A k-out digraph is a digraph in which each vertex has k labeled out-arcs.
- Let $\mathcal{D}_{n,k}$ be a uniform random k-out digraph of n vertices.
- Each arc chooses its endpoint independently and uniformly at random.
- Assume $k \geq 2$.

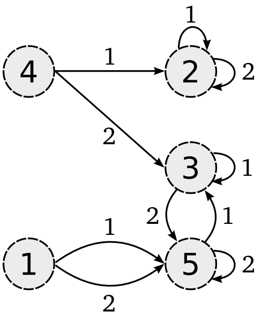


Figure 1 : A 2-out digraph with 5 vertices.

The largest SCC

In 1973 Grusho [4] proved that

The largest SCC

In 1973 Grusho [4] proved that

Theorem 1

1. *Whp there is a* SCC $\mathcal{G}_n$ *in* $\mathcal{D}_{n,k}$ *that is reachable from all vertices.*

The largest SCC

In 1973 Grusho [4] proved that

Theorem 1

1. *Whp there is a* SCC $\mathcal{G}_n$ *in* $\mathcal{D}_{n,k}$ *that is reachable from all vertices.*
2. *And*

$$\frac{|\mathcal{G}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z},$$

where $\mathcal{Z}$ *is the standard normal distribution.*

The largest SCC

In 1973 Grusho [4] proved that

Theorem 1

1. *Whp there is a* SCC $\mathcal{G}_n$ *in* $\mathcal{D}_{n,k}$ *that is reachable from all vertices.*
2. *And*

$$\frac{|\mathcal{G}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z},$$

where $\mathcal{Z}$ *is the standard normal distribution.*

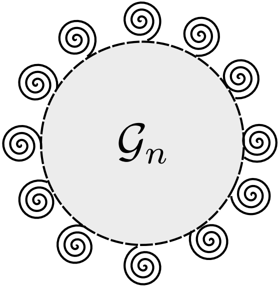


Figure 2 : Since $\nu_k > 1/2$, $\mathcal{G}_n$ is the largest SCC. We call $\mathcal{G}_n$ the giant.

Recent research on $\mathcal{D}_{n,k}$

- In 2012, Carayol and Nicaud [3] reproved some of Grusho's results.

Recent research on $\mathcal{D}_{n,k}$

- In 2012, Carayol and Nicaud [3] reproved some of Grusho's results.
- In 2015, Addario-Berry, Balle, and Perarnau [1] proved that the diameter and the typical distance of $\mathcal{D}_{n,k}$, rescaled by $\log n$, converge in probability to constants.

Recent research on $\mathcal{D}_{n,k}$

- In 2012, Carayol and Nicaud [3] reproved some of Grusho's results.
- In 2015, Addario-Berry, Balle, and Perarnau [1] proved that the diameter and the typical distance of $\mathcal{D}_{n,k}$, rescaled by $\log n$, converge in probability to constants.
- Also in 2014, Angluin and Chen [2] studied the mixing time of simple random walk on $\mathcal{D}_{n,k}$.

Our Result 1 — Cycles outside the giant

Let $\mathcal{G}_n^c \equiv [n] \setminus \mathcal{G}_n$. Let $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$ be the sub-digraph induced by $\mathcal{G}_n^c$.

Our Result 1 — Cycles outside the giant

Let $\mathcal{G}_n^c \equiv [n] \setminus \mathcal{G}_n$. Let $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$ be the sub-digraph induced by $\mathcal{G}_n^c$.

Theorem 2

1. *Let L_n be the length of the longest cycle in $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$. Then $L_n = O_p(1)$.*

Our Result 1 — Cycles outside the giant

Let $\mathcal{G}_n^c \equiv [n] \setminus \mathcal{G}_n$. Let $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$ be the sub-digraph induced by $\mathcal{G}_n^c$.

Theorem 2

1. *Let L_n be the length of the longest cycle in $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$. Then $L_n = O_p(1)$.*
2. *Let C_n be the number of cycles in $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$. Then*

$$C_n \xrightarrow{d} \mathrm{Poi}\left(\log \frac{1}{1 - k\mu_k}\right).$$

The intuition of Theorem 2

- When two cycles share vertices, they contain fewer vertices than arcs.

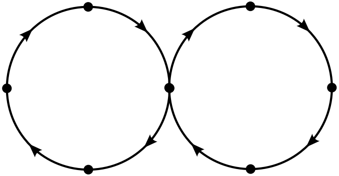


Figure 3 : Two cycles sharing one vertex.

The intuition of Theorem 2

- When two cycles share vertices, they contain fewer vertices than arcs.
- The expected number of this type of sub-digraphs goes to zero.

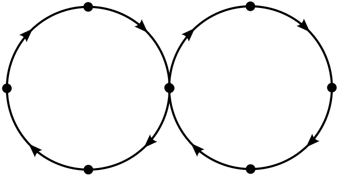


Figure 3 : Two cycles sharing one vertex.

The intuition of Theorem 2

- When two cycles share vertices, they contain fewer vertices than arcs.
- The expected number of this type of sub-digraphs goes to zero.
- Let $\mathcal{C}$ be the set of all possible cycles. Then $C_n \equiv \sum_{\alpha \in \mathcal{C}} \mathbb{1}_{[\alpha]}$.

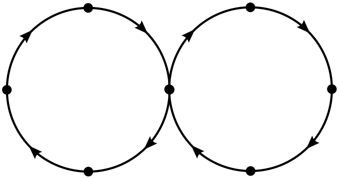


Figure 3 : Two cycles sharing one vertex.

The intuition of Theorem 2

- When two cycles share vertices, they contain fewer vertices than arcs.
- The expected number of this type of sub-digraphs goes to zero.
- Let $\mathcal{C}$ be the set of all possible cycles. Then $C_n \equiv \sum_{\alpha \in \mathcal{C}} \mathbb{1}_{[\alpha]}$.
- The dependence between these indicators is very small.

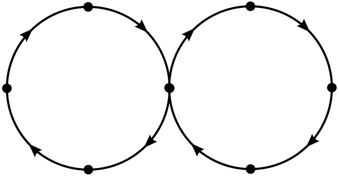


Figure 3 : Two cycles sharing one vertex.

The three layers of $\mathcal{D}_{n,k}$

- $\mathcal{D}_{n,k}$ can be divided into three layers:
  - Center — $\mathcal{G}_n$

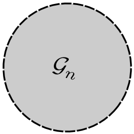


Figure 4 : Three layers of $\mathcal{D}_{n,k}$

The three layers of $\mathcal{D}_{n,k}$

- $\mathcal{D}_{n,k}$ can be divided into three layers:
  - Center — $\mathcal{G}_n$
  - Middle — $\mathcal{O}_n \setminus \mathcal{G}_n$

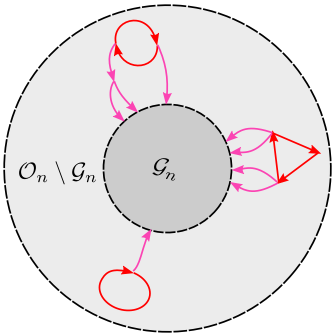


Figure 4 : Three layers of $\mathcal{D}_{n,k}$

The three layers of $\mathcal{D}_{n,k}$

- $\mathcal{D}_{n,k}$ can be divided into three layers:
  - Center — $\mathcal{G}_n$
  - Middle — $\mathcal{O}_n \setminus \mathcal{G}_n$
  - Outer — $\mathcal{O}_n^c$

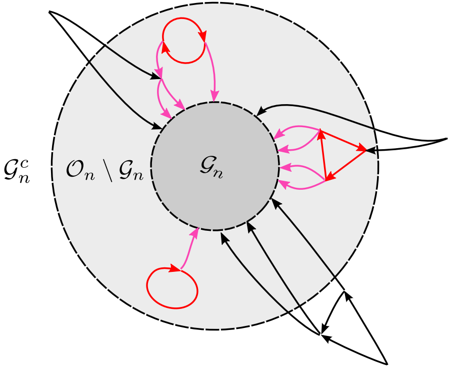


Figure 4 : Three layers of $\mathcal{D}_{n,k}$

The three layers of $\mathcal{D}_{n,k}$

- $\mathcal{D}_{n,k}$ can be divided into three layers:
  - Center — $\mathcal{G}_n$
  - Middle — $\mathcal{O}_n \setminus \mathcal{G}_n$
  - Outer — $\mathcal{O}_n^c$
- $\mathcal{O}_n$ induces the maximum sub-digraph in which in-degree ≥ 1.

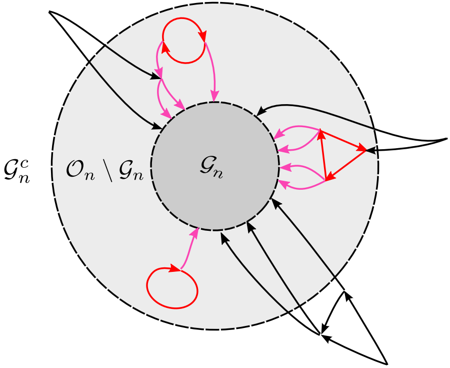


Figure 4 : Three layers of $\mathcal{D}_{n,k}$

The three layers of $\mathcal{D}_{n,k}$

- $\mathcal{D}_{n,k}$ can be divided into three layers:
  - Center — $\mathcal{G}_n$
  - Middle — $\mathcal{O}_n \setminus \mathcal{G}_n$
  - Outer — $\mathcal{O}_n^c$
- $\mathcal{O}_n$ induces the maximum sub-digraph in which in-degree ≥ 1.
- We call $\mathcal{O}_n$ the one-in-core.

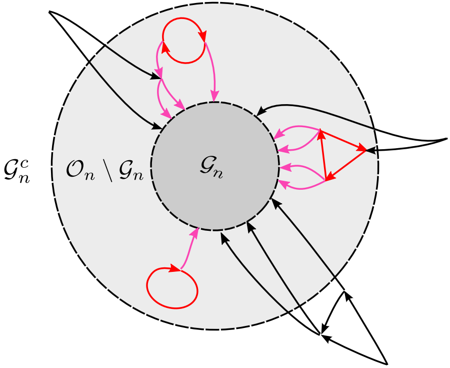


Figure 4 : Three layers of $\mathcal{D}_{n,k}$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.
- In fact $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.
- In fact $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$.
- Thus Theorem 3 implies the central limit law of $|\mathcal{G}_n|$.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.
- In fact $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$.
- Thus Theorem 3 implies the central limit law of $|\mathcal{G}_n|$.
- $\mathcal{O}_n = \mathcal{V}$ iff.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.
- In fact $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$.
- Thus Theorem 3 implies the central limit law of $|\mathcal{G}_n|$.
- $\mathcal{O}_n = \mathcal{V}$ iff.
  (a) $\mathcal{D}_{n,k}[\mathcal{V}]$ is a k-surjection.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.
- In fact $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$.
- Thus Theorem 3 implies the central limit law of $|\mathcal{G}_n|$.
- $\mathcal{O}_n = \mathcal{V}$ iff.
  (a) $\mathcal{D}_{n,k}[\mathcal{V}]$ is a k-surjection.
  (b) $\mathcal{D}_{n,k}[\mathcal{V}^c]$ is acyclic.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Our result 2 — A new proof of Theorem 1

- By Theorem 2, whp $|\mathcal{O}_n| - |\mathcal{G}_n|$ is small.
- In fact $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$.
- Thus Theorem 3 implies the central limit law of $|\mathcal{G}_n|$.
- $\mathcal{O}_n = \mathcal{V}$ iff.
  (a) $\mathcal{D}_{n,k}[\mathcal{V}]$ is a k-surjection.
  (b) $\mathcal{D}_{n,k}[\mathcal{V}^c]$ is acyclic.
- $\mathbb{P}\left\{\mathcal{O}_n = \mathcal{V}\right\} = \mathbb{P}\left\{(a)\right\}\mathbb{P}\left\{(b)\right\}$.

Theorem 3

As $n \to \infty$,

$$\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

Spectra outside the giant

- $\mathcal{S}_v$ (the spectrum of v) is the set of vertices the are reachable from v.

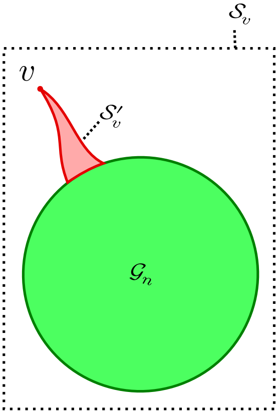


Figure 5 : The spectrum of a vertex.

Spectra outside the giant

- $\mathcal{S}_v$ (the spectrum of v) is the set of vertices the are reachable from v.
- Let $\mathcal{S}'_v \equiv \mathcal{S}_v \cap \mathcal{G}_n^c$.

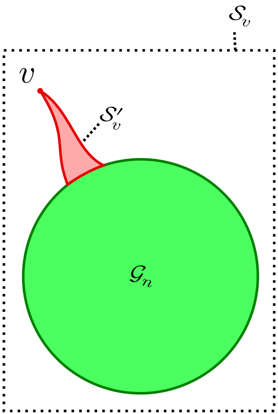


Figure 5 : The spectrum of a vertex.

Spectra outside the giant

- $\mathcal{S}_v$ (the spectrum of v) is the set of vertices the are reachable from v.
- Let $\mathcal{S}'_v \equiv \mathcal{S}_v \cap \mathcal{G}_n^c$.
- Grusho showed that

$$\frac{|\mathcal{S}_1| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

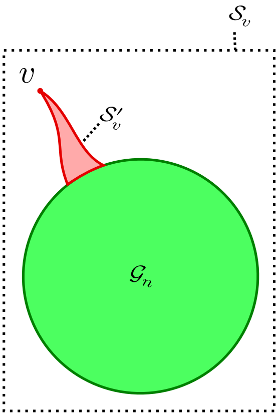


Figure 5 : The spectrum of a vertex.

Spectra outside the giant

- $\mathcal{S}_v$ (the spectrum of v) is the set of vertices the are reachable from v.
- Let $\mathcal{S}'_v \equiv \mathcal{S}_v \cap \mathcal{G}_n^c$.
- Grusho showed that

$$\frac{|\mathcal{S}_1| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{Z}.$$

- And $|\mathcal{S}'_1| = O_p(1)$.

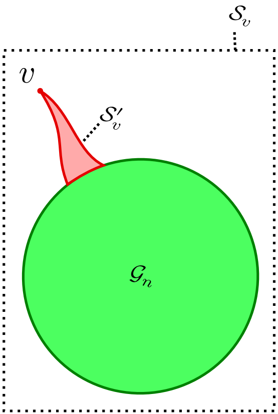


Figure 5 : The spectrum of a vertex.

Our result 3 — Spectra outside the giant

Theorem 4

1. *Whp every S'_v induces either a tree or* ...

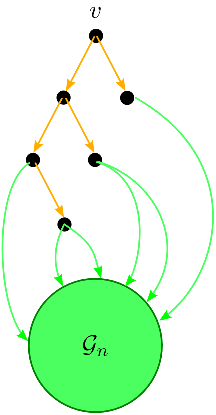


Figure 6 : The tree-like structure of spectra.

Our result 3 — Spectra outside the giant

Theorem 4

1. *Whp every S'_v induces either a tree or a tree plus an extra arc.*

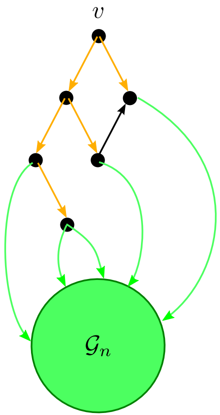


Figure 6 : The tree-like structure of spectra.

Our result 3 — Spectra outside the giant

Theorem 4

1. *Whp every S'_v induces either a tree or a tree plus an extra arc.*
2. *Let* $S_n \equiv \max_{v \in \mathcal{G}_n^c} |\mathcal{S}'_v|$. *Then*

$$\frac{S_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/\lambda_k)}.$$

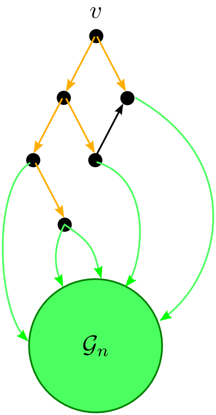


Figure 6 : The tree-like structure of spectra.

Our result 3 — Spectra outside the giant

Theorem 4

1. *Whp every S'_v induces either a tree or a tree plus an extra arc.*
2. *Let* $S_n \equiv \max_{v \in \mathcal{G}_n^c} |\mathcal{S}'_v|$. *Then*

$$\frac{S_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/\lambda_k)}.$$

3. *Let* $W_n \equiv \max_{v \in \mathcal{G}_n^c} \operatorname{dist}(v, \mathcal{G}_n)$. *Then*

$$\frac{W_n}{\log_k \log n} \xrightarrow{p} 1.$$

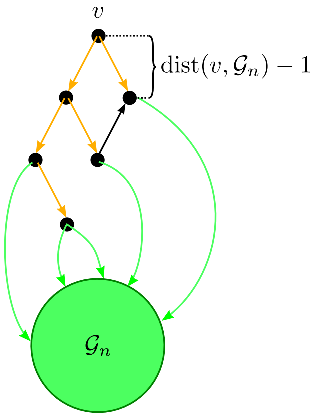


Figure 6 : The tree-like structure of spectra.

Our result 3 — Spectra outside the giant (continued)

Theorem 4

4. *Let* $D_n \equiv \max_{v \in \mathcal{G}_n^c} \max_{u \in \mathcal{S}'_v} \mathrm{dist}(v, u)$. *Then*

$$\frac{D_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/k\mu_k)}.$$

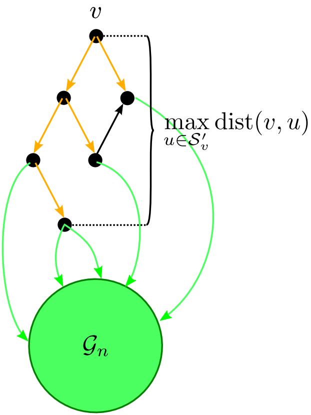


Figure 7 : The tree-like structure of spectra.

Our result 3 — Spectra outside the giant (continued)

Theorem 4

4. *Let* $D_n \equiv \max_{v \in \mathcal{G}_n^c} \max_{u \in \mathcal{S}'_v} \mathrm{dist}(v,u)$. *Then*

$$\frac{D_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/k\mu_k)}.$$

5. *Let* M_n *be the length of the longest path in* $\mathcal{D}_{n,k}[\mathcal{G}_n^c]$. *Then*

$$\frac{M_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/k\mu_k)}.$$

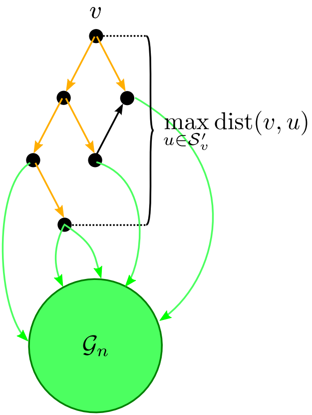


Figure 7 : The tree-like structure of spectra.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.

Figure 8 : Try until success.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.
- If it is large enough, we succeed.

Figure 8 : Try until success.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.
- If it is large enough, we succeed.
- Otherwise, we try again.

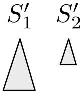


Figure 8 : Try until success.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.
- If it is large enough, we succeed.
- Otherwise, we try again.
- And again.

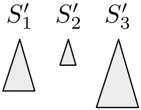


Figure 8 : Try until success.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.
- If it is large enough, we succeed.
- Otherwise, we try again.
- And again.
- Until we find a large one.

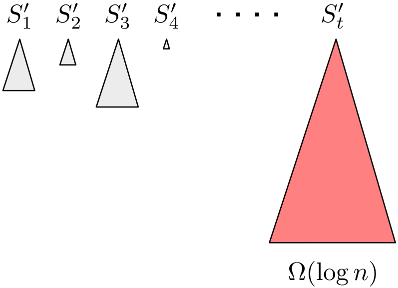


Figure 8 : Try until success.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.
- If it is large enough, we succeed.
- Otherwise, we try again.
- And again.
- Until we find a large one.
- The probability to succeed in one trial is small.

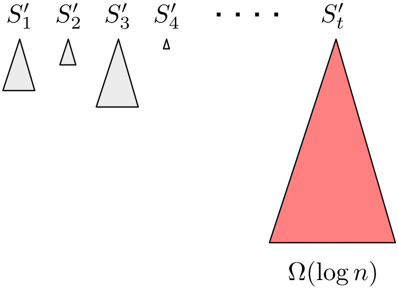


Figure 8 : Try until success.

Proof technique — Try until success

To find a $\mathcal{S}'_v = \Omega(\log n)$:

- Explore the spectrum of one vertex.
- If it is large enough, we succeed.
- Otherwise, we try again.
- And again.
- Until we find a large one.
- The probability to succeed in one trial is small.
- The probability to eventually succeed goes to one.

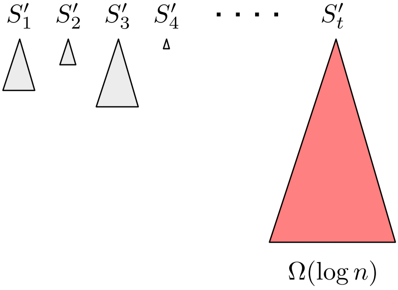


Figure 8 : Try until success.

Our result 4 – Strong connectivity

1. Recall that $|\mathcal{G}_n| \approx \nu_k n$.

Our result 4 – Strong connectivity

1. Recall that $|\mathcal{G}_n| \approx \nu_k n$.
2. As $\nu_k \geq 1 - e^{-k+O(1)}$, $\nu_k \to 1$ as $k \to \infty$.

Our result 4 – Strong connectivity

1. Recall that $|\mathcal{G}_n| \approx \nu_k n$.
2. As $\nu_k \geq 1 - e^{-k+O(1)}$, $\nu_k \to 1$ as $k \to \infty$.
3. When does the whole digraph get strongly connected?

Our result 4 – Strong connectivity

1. Recall that $|\mathcal{G}_n| \approx \nu_k n$.
2. As $\nu_k \geq 1 - e^{-k+O(1)}$, $\nu_k \to 1$ as $k \to \infty$.
3. When does the whole digraph get strongly connected?

Our result 4 – Strong connectivity

1. Recall that $|\mathcal{G}_n| \approx \nu_k n$.
2. As $\nu_k \geq 1 - e^{-k+O(1)}$, $\nu_k \to 1$ as $k \to \infty$.
3. When does the whole digraph get strongly connected?

Theorem 5 (Phase transition in strong connectivity)

1. *If* $k - \log n \to -\infty$, *then whp* $\mathcal{D}_{n,k}$ *is not strongly connected.*
2. *If* $k - \log n \to \infty$, *then whp* $\mathcal{D}_{n,k}$ *is strongly connected.*

Some open questions

1. Vertex connectivity of $\mathcal{G}_n$.

Some open questions

1. Vertex connectivity of $\mathcal{G}_n$.
2. Extends central limit laws to simple digraphs.

Some open questions

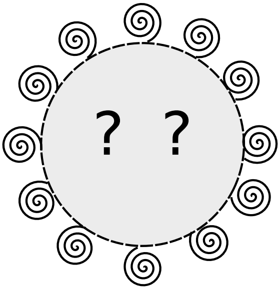
1. Vertex connectivity of $\mathcal{G}_n$.
2. Extends central limit laws to simple digraphs.
3. Generalize the model to non-regular out-degree sequence.

Some open questions

1. Vertex connectivity of $\mathcal{G}_n$.
2. Extends central limit laws to simple digraphs.
3. Generalize the model to non-regular out-degree sequence.
4. Generating a uniform random surjection from $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$ for arbitrary $m = m(n)$.

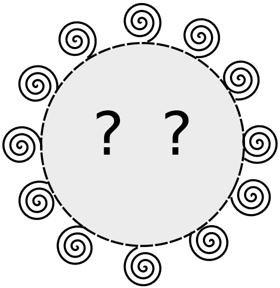
Acknowledgment and Questions

1. The authors thank Laura Eslava, Hamed Hatami, Guillem Perarnau, Bruce Reed, Henning Sulzbach and Yelena Yuditsky for valuable comments on this work,



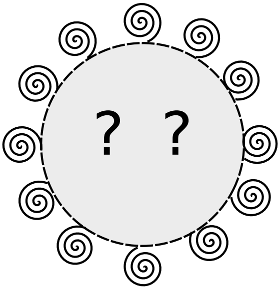
Acknowledgment and Questions

1. The authors thank Laura Eslava, Hamed Hatami, Guillem Perarnau, Bruce Reed, Henning Sulzbach and Yelena Yuditsky for valuable comments on this work,
2. and Denis Thérien for pointing out the importance of the model,



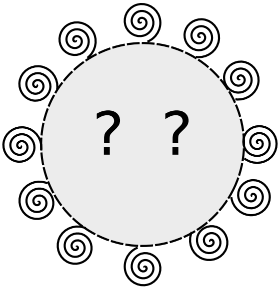
Acknowledgment and Questions

1. The authors thank Laura Eslava, Hamed Hatami, Guillem Perarnau, Bruce Reed, Henning Sulzbach and Yelena Yuditsky for valuable comments on this work,
2. and Denis Thérien for pointing out the importance of the model,
3. and Ross Kang for suggesting me to give this talk.



Acknowledgment and Questions

1. The authors thank Laura Eslava, Hamed Hatami, Guillem Perarnau, Bruce Reed, Henning Sulzbach and Yelena Yuditsky for valuable comments on this work,
2. and Denis Thérien for pointing out the importance of the model,
3. and Ross Kang for suggesting me to give this talk.
4. Thanks for listening!



References

[1] L. Addario-Berry, B. Balle, and G. Perarnau. Diameter and Stationary Distribution of Random r-out Digraphs. *ArXiv e-prints*, 2015.

[2] D. Angluin and D. Chen. Random walks on random uni-regular graphs. 2015.

[3] A. Carayol and C. Nicaud. Distribution of the number of accessible states in a random deterministic automaton. In *29th International Symposium on Theoretical Aspects of Computer Science*, volume 14 of *Leibniz International Proceedings in Informatics*, 194–205. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2012.

[4] A. A. Grusho. Limit distributions of certain characteristics of random automaton graphs. *Mathematical Notes of the Academy of Sciences of the USSR*, 14(1):633–637, 1973.