Scaling limit of a critical random directed graph

Robin Stephenson
University of Oxford

Joint work with Christina Goldschmidt.

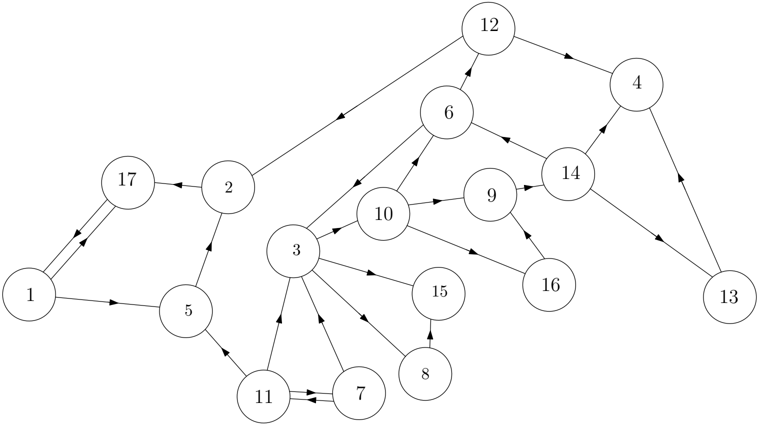
Introduction and main result

Random directed graph

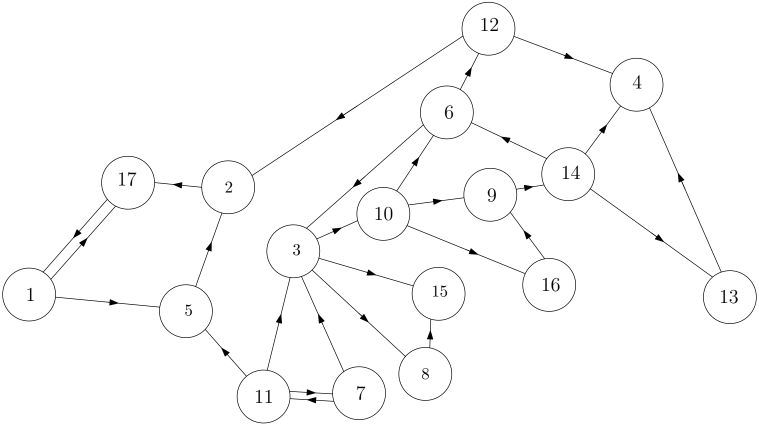
For $n \in \mathbb{N}$ and $p \in [0,1]$, let $\vec{G}(n,p)$ be the random directed defined by :

- Vertices $= \{1, \ldots, n\}$
- Take each of the $n(n-1)$ possible directed edges independently with probability p.

Random directed graph

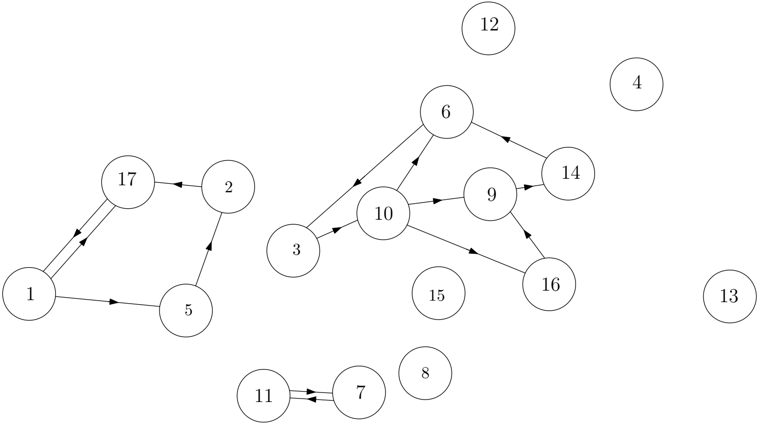


Random directed graph

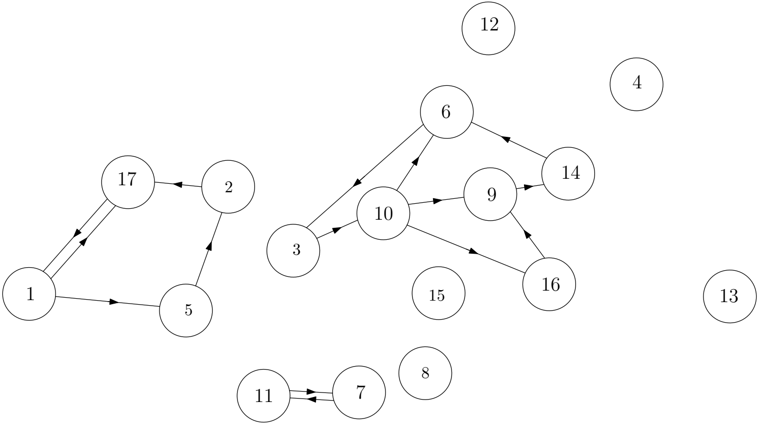


We are interested in the *strongly connected components* : maximal subgraphs where we can go from any vertex to any other in both directions.

Strongly connected components



Strongly connected components



Notice that not all edges are part of a single strongly connected component. Very different from undirected graphs !

Phase transition and critical window

It is known that $\vec{G}(n,p)$ has the same phase transition as $G(n,p)$ for the size of connected components : giant component when $p = c/n$ with $c > 1$ etc.

Phase transition and critical window

It is known that $\vec{G}(n,p)$ has the same phase transition as $G(n,p)$ for the size of connected components : giant component when $p = c/n$ with $c > 1$ etc.
It even has the same *critical window* :

Phase transition and critical window

It is known that $\vec{G}(n,p)$ has the same phase transition as $G(n,p)$ for the size of connected components : giant component when $p = c/n$ with $c > 1$ etc.

It even has the same *critical window* :

Theorem (Łuczak and Seierstad '09)

Assume $p = \frac{1}{n} + \frac{\lambda_n}{n^{4/3}}$.

(i) *If* $\lambda_n \to \infty$ *then the largest strongly connected component of* $\vec{G}(n,p)$ *has size* $\sim 4\lambda_n^2 n^{1/3}$ *and the second largest has size* $O(\gamma_n^{-1} n^{1/3})$.

(ii) *If* $\lambda_n \to -\infty$ *then the largest strongly connected component of* $\vec{G}(n,p)$ *has size* $O(|\lambda_n^{-1}| n^{1/3})$.

Phase transition and critical window

It is known that $\vec{G}(n,p)$ has the same phase transition as $G(n,p)$ for the size of connected components : giant component when $p = c/n$ with $c > 1$ etc.

It even has the same *critical window* :

Theorem (Łuczak and Seierstad '09)

Assume $p = \frac{1}{n} + \frac{\lambda_n}{n^{4/3}}$.

(i) *If* $\lambda_n \to \infty$ *then the largest strongly connected component of* $\vec{G}(n,p)$ *has size* $\sim 4\lambda_n^2 n^{1/3}$ *and the second largest has size* $O(\gamma_n^{-1} n^{1/3})$.

(ii) *If* $\lambda_n \to -\infty$ *then the largest strongly connected component of* $\vec{G}(n,p)$ *has size* $O(|\lambda_n^{-1}| n^{1/3})$.

We investigate what happens within the critical window : $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$.

A good reference point : the scaling limit of the Erdős–Rényi graph

Let $G(n,p)$ be the undirected Erdős–Rényi graph. We call :

- $A_1(n), A_2(n), \ldots$ the connected components of $G(n,p)$.
- $Z_1^n \geqslant Z_2^n \geqslant \ldots$ their sizes.

A good reference point : the scaling limit of the Erdős–Rényi graph

Let $G(n,p)$ be the undirected Erdős–Rényi graph. We call :

- $A_1(n), A_2(n), \ldots$ the connected components of $G(n,p)$.
- $Z_1^n \geqslant Z_2^n \geqslant \ldots$ their sizes.

Theorem

- *(Aldous '97)*

$$\left(\frac{Z_i^n}{n^{2/3}}, i \in \mathbb{N}\right) \xrightarrow[\ell^2]{(\mathrm{d})} (\sigma_i, i \in \mathbb{N})$$

A good reference point : the scaling limit of the Erdős–Rényi graph

Let $G(n,p)$ be the undirected Erdős–Rényi graph. We call :

- $A_1(n), A_2(n), \ldots$ the connected components of $G(n,p)$.
- $Z_1^n \geqslant Z_2^n \geqslant \ldots$ their sizes.

Theorem

- *(Aldous '97)*

$$\left(\frac{Z_i^n}{n^{2/3}}, i \in \mathbb{N}\right) \xrightarrow[\ell^2]{(\mathrm{d})} (\sigma_i, i \in \mathbb{N})$$

- *(Addario-Berry, Broutin and Goldschmidt '12)*

$$\left(\frac{A_i(n)}{n^{1/3}}, i \in \mathbb{N}\right) \xrightarrow[\ell^4\text{-}GH]{(\mathrm{d})} (\mathcal{A}_i, i \in \mathbb{N}),$$

Graphs as metric spaces

- This views the $A_i(n)$ as metric spaces by giving each edge a length of 1, and then rescaling everything by $n^{1/3}$.
- They then converge for the *Gromov-Hausdorff topology*

Graphs as metric spaces

- This views the $A_i(n)$ as metric spaces by giving each edge a length of 1, and then rescaling everything by $n^{1/3}$.
- They then converge for the *Gromov-Hausdorff topology*
- Problem : this isn't an ideal setting for directed graphs.

The correct setting : multigraphs with edge lengths

- Let $C_1(n), C_2(n), \ldots$ be the strongly connected components of $\vec{G}(n,p)$, ordered by decreasing sizes.

The correct setting : multigraphs with edge lengths

- Let $C_1(n), C_2(n), \ldots$ be the strongly connected components of $\vec{G}(n,p)$, ordered by decreasing sizes.
- Assign to each edge a length of 1.

The correct setting : multigraphs with edge lengths

- Let $C_1(n), C_2(n), \ldots$ be the strongly connected components of $\vec{G}(n,p)$, ordered by decreasing sizes.
- Assign to each edge a length of 1.
- Remove any vertex with indegree and outdegree 1, merging its adjacent edges.
- One exception : if a component is just a cycle, keep a vertex.

The correct setting : multigraphs with edge lengths

- Let $C_1(n), C_2(n), \ldots$ be the strongly connected components of $\vec{G}(n,p)$, ordered by decreasing sizes.
- Assign to each edge a length of 1.
- Remove any vertex with indegree and outdegree 1, merging its adjacent edges.
- One exception : if a component is just a cycle, keep a vertex.

This makes the $C_i(n)$ into *directed multigraphs with edge lengths.*

A metric for directed multigraphs with edge lengths

Let $\vec{\mathcal{G}}$ be the set of (equivalence classes of) directed multigraphs with edge lengths. For X and Y in $\vec{\mathcal{G}}$ we let

$$d_{\vec{\mathcal{G}}}(X,Y) = \begin{cases} \infty & \text{if the underlying graphs are different} \\ \inf_{\text{isomorphisms}} \sum_{e \in \{\text{edges}\}} |\ell_X(e) - \ell_Y(e)| & \text{otherwise} \end{cases}$$

A metric for directed multigraphs with edge lengths

Let $\vec{\mathcal{G}}$ be the set of (equivalence classes of) directed multigraphs with edge lengths. For X and Y in $\vec{\mathcal{G}}$ we let

$$d_{\vec{\mathcal{G}}}(X,Y) = \begin{cases} \infty & \text{if the underlying graphs are different} \\ \inf\limits_{\text{isomorphisms}} & \sum\limits_{e \in \{\text{edges}\}} |\ell_X(e) - \ell_Y(e)| \text{ otherwise} \end{cases}$$

For sequences, we use the ℓ^1 version : for $\mathbf{A} = (A_1, A_2, \dots,)$ and $\mathbf{B} = (B_1, B_2, \dots,)$.

$$d(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{\infty} d_{\vec{\mathcal{G}}}(A_i, B_i),$$

Convergence theorem

Theorem (Goldschmidt-S. '19)

There exists a sequence $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ *of random strongly connected directed multigraphs with edge lengths such that, for each* $i \geq 1$, $\mathcal{C}_i$ *is either 3-regular or a loop, and such that*

$$\left(\frac{C_i(n)}{n^{1/3}}, i \in \mathbb{N} \right) \xrightarrow{\text{(d)}} (\mathcal{C}_i, i \in \mathbb{N})$$

Convergence theorem

Theorem (Goldschmidt-S. '19)

There exists a sequence $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ of random strongly connected directed multigraphs with edge lengths such that, for each $i \geq 1$, $\mathcal{C}_i$ is either 3-regular or a loop, and such that

$$\left(\frac{C_i(n)}{n^{1/3}}, i \in \mathbb{N}\right) \xrightarrow{(\mathrm{d})} (\mathcal{C}_i, i \in \mathbb{N})$$

Remarks :

- The number of degree 2 vertices of $C_i(n)$ is of order $n^{1/3}$.
- The number of degree 3 vertices of $C_i(n)$ is of order 1.
- No vertices of degree $\geqslant 4$ with probability tending to 1.

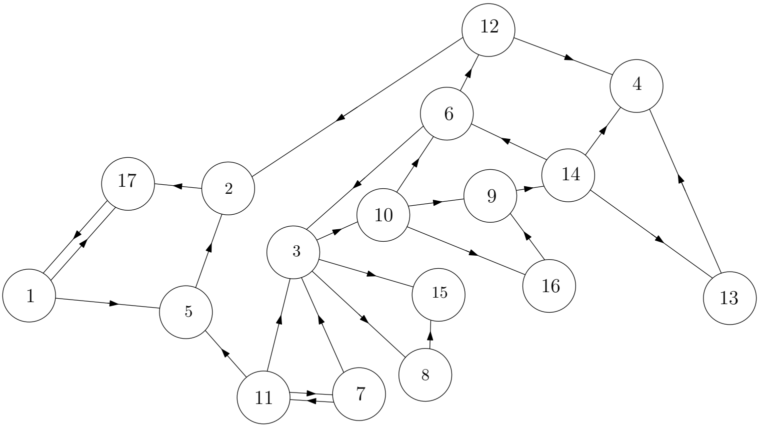
Using an exploration process

Exploration and a spanning forest

We build a *planar spanning forest* $\mathcal{F}_{\vec{G}(n,p)}$ of $\vec{G}(n,p)$ by using a variant of *depth-first search*.

- Start by classifying 1 as "seen".
- At each step, *explore* the leftmost seen vertex : add all of its yet unseen outneighbours to the forest from left to right with increasing labels, along with their linking edge, and count them as seen.
- If there are no available seen vertices, we take the unseen vertex with smallest label, and put it in a new tree component on the right.

Reminder and practice



Edge classification

There are three kinds of edges :

- Edges of $\mathcal{F}_{\vec{G}(n,p)}$.
- "Surplus" edges. These are edges which are not in the forest because their target was already seen when we explored the origin.
- "Back" edges. These go backwards for the planar structure on the forest.

The interaction between back and forward edges is what creates strongly connected components.

Strategy

To understand the scaling limit, all we need to do is understand these three parts, and how they interact.

Scaling limit of the trees

Comparison with Erdős–Rényi

Fact : $\mathcal{F}_{\vec{G}(n,p)}$ has the same distribution as $\mathcal{F}_{G(n,p)}$, the forest obtained by applying the same procedure to $G(n,p)$.

Comparison with Erdős–Rényi

Fact : $\mathcal{F}_{\vec{G}(n,p)}$ has the same distribution as $\mathcal{F}_{G(n,p)}$, the forest obtained by applying the same procedure to $G(n,p)$.
Consequence : let

- $T_1^n, T_2^n, \ldots$ the tree components of $\mathcal{F}_{\vec{G}(n,p)}$.
- $Z_1^n \geqslant Z_2^n \geqslant \ldots$ their sizes.

We have the convergences :

$$n^{-2/3}(Z_i^n, i \in \mathbb{N}) \xrightarrow[\ell^2]{(\mathrm{d})} (\sigma_i, i \in \mathbb{N})$$

Comparison with Erdős–Rényi

Fact : $\mathcal{F}_{\vec{G}(n,p)}$ has the same distribution as $\mathcal{F}_{G(n,p)}$, the forest obtained by applying the same procedure to $G(n,p)$.
Consequence : let

- $T_1^n, T_2^n, \ldots$ the tree components of $\mathcal{F}_{\vec{G}(n,p)}$.
- $Z_1^n \geqslant Z_2^n \geqslant \ldots$ their sizes.

We have the convergences :

$$n^{-2/3}(Z_i^n, i \in \mathbb{N}) \xrightarrow[\ell^2]{(\mathrm{d})} (\sigma_i, i \in \mathbb{N})$$

$$\left(\frac{T_i^n}{n^{1/3}}, i \in \mathbb{N}\right) \xrightarrow{(\mathrm{d})} (\mathcal{T}_i, i \in \mathbb{N})$$

Details (for those who know)

- $(\sigma_i, i \in \mathbb{N})$ are the excursion lengths of a drifted Brownian motion :

$$W^\lambda(t) = W(t) + \lambda t - t^2/2$$

Details (for those who know)

- $(\sigma_i, i \in \mathbb{N})$ are the excursion lengths of a drifted Brownian motion :

$$W^\lambda(t) = W(t) + \lambda t - t^2/2$$

- Conditionally on $(\sigma_i, i \in \mathbb{N})$, $(\mathcal{T}_i, i \in \mathbb{N})$ are independent biased Brownian trees. Specifically, $\mathcal{T}_i$ has the distribution of the tree encoded by the function $2\tilde{\mathbf{e}}^{(\sigma_i)}$, where

$$\mathbb{E}[g(\tilde{\mathbf{e}}^{(\sigma)})] = \frac{\mathbb{E}\left[g(\sqrt{\sigma}\mathbf{e}(\cdot/\sigma)) \exp\left(\sigma^{3/2}\int_0^1 \mathbf{e}(x)dx\right)\right]}{\mathbb{E}\left[\exp\left(\sigma^{3/2}\int_0^1 \mathbf{e}(x)dx\right)\right]}$$

and $\mathbf{e}$ is a standard brownian excursion.

Limiting behaviour of the surplus and back edges

Working on a single tree

- Note that any strongly connected component of $\vec{G}(n,p)$ is contained within one of the trees of $\mathcal{F}_{\vec{G}(n,p)}$.

Working on a single tree

- Note that any strongly connected component of $\vec{G}(n,p)$ is contained within one of the trees of $\mathcal{F}_{\vec{G}(n,p)}$.
- So we can focus on a single tree, with say m vertices, with $m \sim \sigma n^{2/3}$. Call that tree T_m.

Working on a single tree

- Note that any strongly connected component of $\vec{G}(n,p)$ is contained within one of the trees of $\mathcal{F}_{\vec{G}(n,p)}$.
- So we can focus on a single tree, with say m vertices, with $m \sim \sigma n^{2/3}$. Call that tree T_m.
- Conditionally on T_m, all the $m(m-1)/2$ back edges appear independently with probability p, and all of the $a(T_m)$ possible surplus edges also do.

Surplus edges don't matter

We can show that

$$\mathbb{P}[\text{A strongly component in } T_m \text{ features a surplus edge}] \to 0.$$

Surplus edges don't matter

We can show that

$$\mathbb{P}[\text{A strongly component in } T_m \text{ features a surplus edge}] \to 0.$$

Idea of the proof :

- The number of surplus edges is of order 1.
- The number of descendants of a surplus edges is of order 1.
- So the number of back edges starting at a descendant of a surplus edge is $O(mp) \to 0$.

Back edges - potential problem

The number of back edges in T_m follows a $\mathrm{Bin}(\frac{m(m-1)}{2}, p)$ distribution.

Back edges - potential problem

The number of back edges in T_m follows a $\mathrm{Bin}(\frac{m(m-1)}{2}, p)$ distribution.

But $p\frac{m(m-1)}{2} \sim \sigma^2/2n^{1/3} \to \infty$.

Back edges - potential problem

The number of back edges in T_m follows a $\mathrm{Bin}(\frac{m(m-1)}{2}, p)$ distribution.

But $p\frac{m(m-1)}{2} \sim \sigma^2/2n^{1/3} \to \infty$.

This is not a problem ! Because only a finite number of back edges actually are part of strongly connected components.

Back edges which matter, as a process

Do the contour exploration of T_m, recording back edges at their origins.

Back edges which matter, as a process

Do the contour exploration of T_m, recording back edges at their origins.

If the first back edge is not *ancestral*, then it will not contribute to a strongly connected component.

Back edges which matter, as a process

Do the contour exploration of T_m, recording back edges at their origins.

If the first back edge is not *ancestral*, then it will not contribute to a strongly connected component.

More generally, any back edge arriving before the first ancestral one does not contribute.

Back edges which matter, as a process

Do the contour exploration of T_m, recording back edges at their origins.

If the first back edge is not *ancestral*, then it will not contribute to a strongly connected component.

More generally, any back edge arriving before the first ancestral one does not contribute.

After the first ancestral back edge (x_1, y_1), other ancestral back edges will contribute, but also possibly those which point between y_1 and x_1.

Back edges which matter, as a process

Do the contour exploration of T_m, recording back edges at their origins.

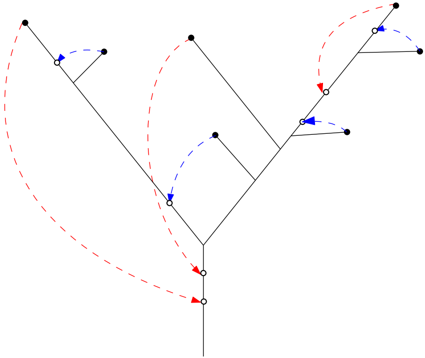
If the first back edge is not *ancestral*, then it will not contribute to a strongly connected component.

More generally, any back edge arriving before the first ancestral one does not contribute.

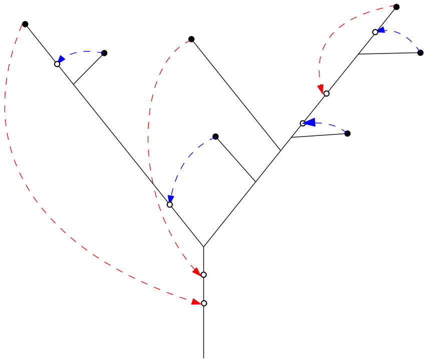
After the first ancestral back edge (x_1, y_1), other ancestral back edges will contribute, but also possibly those which point between y_1 and x_1.

And so on. We can show that the number of back edges observed in this stays bounded as $n \to \infty$.

What we end up with

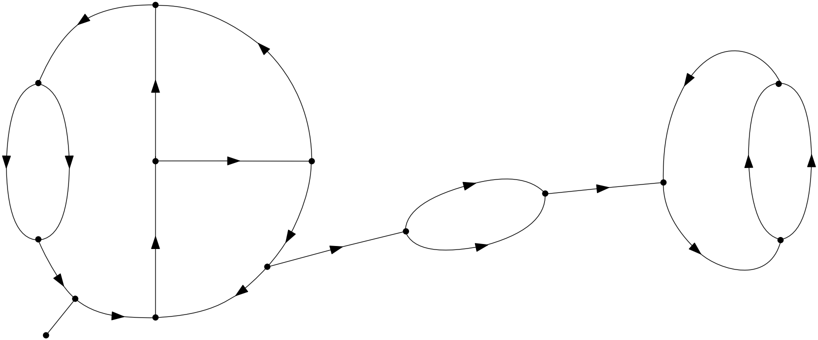


What we end up with

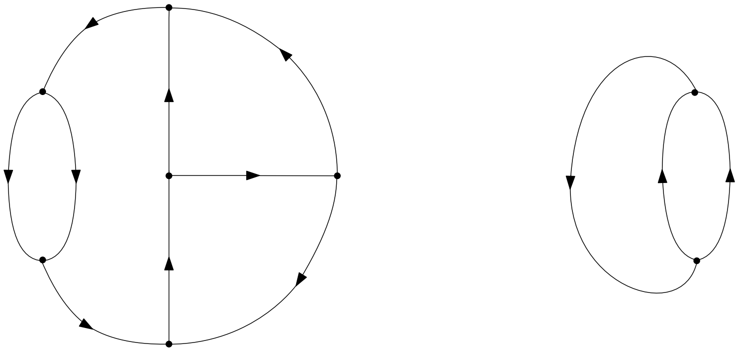


Rescale the distances by $n^{1/3}$ and this is a convergence for $d_{\vec{G}}$.

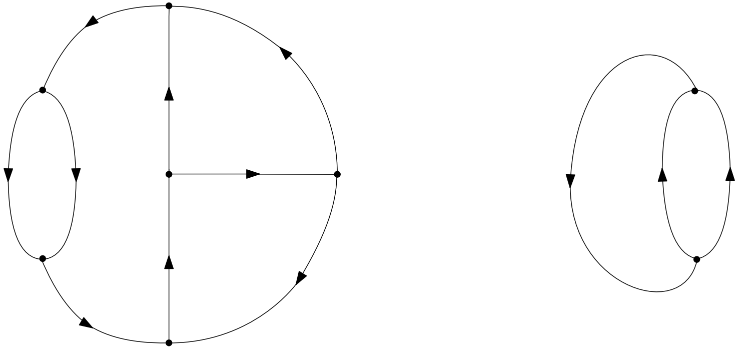
What we end up with



What we end up with



What we end up with



Do this for each tree, and we get the $\mathcal{C}_i$.

Thank you !