A Null Model of Speciation by Reproductive Isolation

Mechanisms of Reproductive Isolation Session

Flo(rence) Débarre

20 February 2017

Acknowledgements

Work done in collaboration with

François Bienvenu

Amaury Lambert

Acknowledgements

Work done in collaboration with

François Bienvenu

Amaury Lambert

Funding

ANR-14-ACHN-0003-01

Charley Harper, “Tree of Life”

But what is a species?

de Quieroz (2007)

But what is a species?

Biological

Interbreeding

Isolation

Recognition

Cohesion

Phenotypic cohesion

Ecological

Same niche or adaptive zone

Phylogenetic

Hennigian

Segment btw nodes

Monophyletic

Monophyly

Genealogical

Exclusive coalescence of alleles

Diagnosable

Smallest appropriate unit

Evolutionary

Unit of evolution

Phenetic

Phenetic cluster

Genotypic cluster

Deficits of genetic intermediates

de Quieroz (2007)

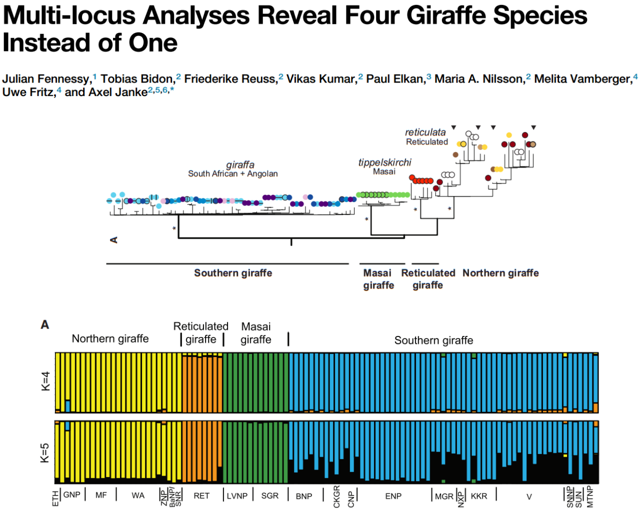
But what is a species?

"there is no unique relation which is privileged in that the species taxa it generates will answer to the needs of all biologists and will be applicable to all groups of organisms."

Kitcher (1984)

de Quieroz (2007)

Counting species

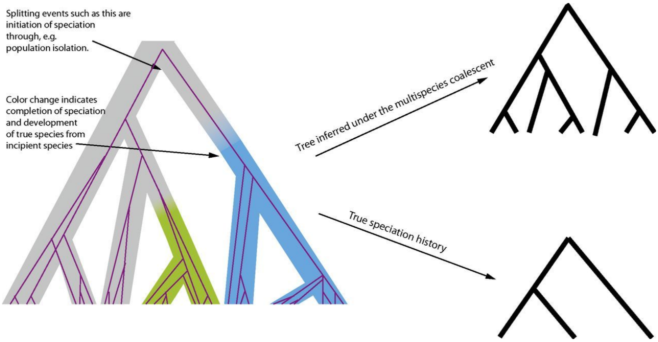


Fennessy et al. (2016)

Counting species (*continued*)

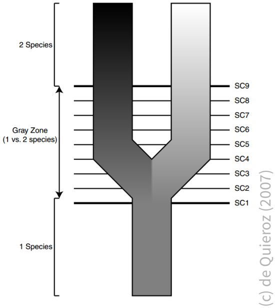
Multispecies coalescent delimits structure, not species

Jeet Sukumaran[a,1,2] and L. Lacey Knowles[a,1]



Sukumaran and Knowles (2017)

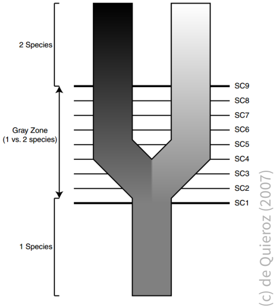
Counting species (*continued*)



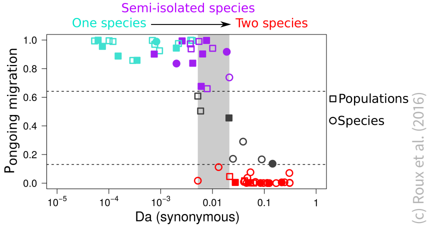
- Gre/ay zone of speciation

de Quieroz (2007)

Counting species (*continued*)

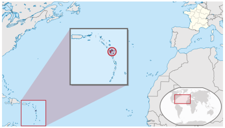


- Gre/ay zone of speciation
- Continuum of speciation



de Quieroz (2007)
Roux et al. (2016)

Other consequences of species definitions



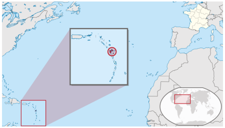
(c) Wikimedia

(c) https://flic.kr/p/dEsdrT

Helgen et al. (2008)

http://www.radiolab.org/story/stanger-paradise/

Other consequences of species definitions



(c) Wikimedia

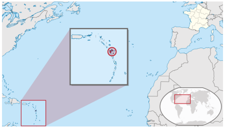
(c) https://flic.kr/p/dEsdrT

Procyon minor

Protected

Helgen et al. (2008)
`http://www.radiolab.org/story/stanger-paradise/`

Other consequences of species definitions



(c) Wikimedia

(c) https://flic.kr/p/dEsdrT

Procyon minor	*Procyon lotor*

Protected	Invasive!

Helgen et al. (2008)

http://www.radiolab.org/story/stanger-paradise/

Reproductive isolation in the Biological Species Concept

Some limits

Rothfels et al. (2015)

Reproductive isolation in the Biological Species Concept

Some limits

- Asexuals

Rothfels et al. (2015)

Reproductive isolation in the Biological Species Concept

Some limits

- Asexuals
- Horizontal gene transfer in Prokaryotes

Rothfels et al. (2015)

Reproductive isolation in the Biological Species Concept

Some limits

- Asexuals
- Horizontal gene transfer in Prokaryotes
- Hybrids

$\times$*Cystocarpium roskamianum*

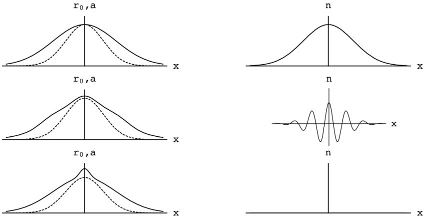
Rothfels et al. (2015)

From populations to species

Cluster formation

Models with competition/selection

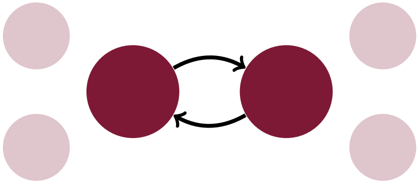
"the mathematical structure of the ecological coexistence problem itself dictates the discreteness of the species."



Gyllenberg and Meszéna (2005)

Neutral models of speciation

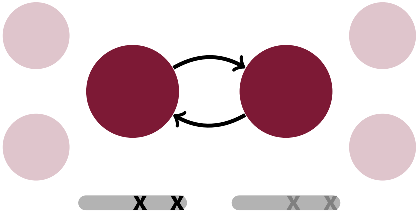
Evolution of reproductive isolation



Gavrilets et al. (2000),
Yamaguchi and Iwasa (2013, 2015)
Miró Pina and Schertzer (unpubl.)

Neutral models of speciation

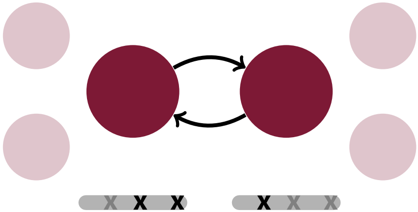
Evolution of reproductive isolation



Gavrilets et al. (2000),
Yamaguchi and Iwasa (2013, 2015)
Miró Pina and Schertzer (unpubl.)

Neutral models of speciation

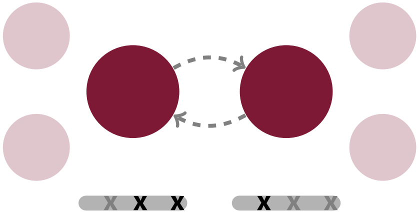
Evolution of reproductive isolation



Gavrilets et al. (2000),
Yamaguchi and Iwasa (2013, 2015)
Miró Pina and Schertzer (unpubl.)

Neutral models of speciation

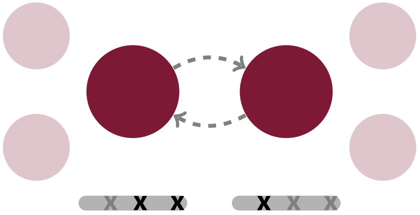
Evolution of reproductive isolation



Gavrilets et al. (2000),
Yamaguchi and Iwasa (2013, 2015)
Miró Pina and Schertzer (unpubl.)

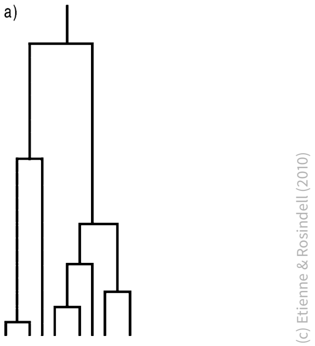
Neutral models of speciation

Evolution of reproductive isolation



Gavrilets et al. (2000),
Yamaguchi and Iwasa (2013, 2015)
Miró Pina and Schertzer (unpubl.)

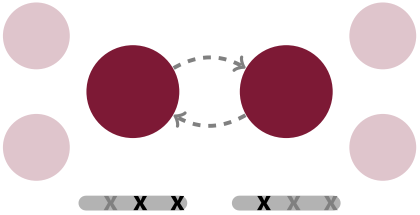
Macroevolutionary models



Rosindell et al. (2010)
Etienne and Rosindell (2010)

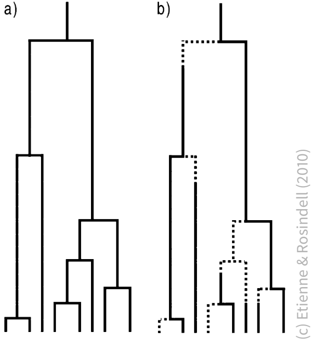
Neutral models of speciation

Evolution of reproductive isolation



Gavrilets et al. (2000),
Yamaguchi and Iwasa (2013, 2015)
Miró Pina and Schertzer (unpubl.)

Macroevolutionary models



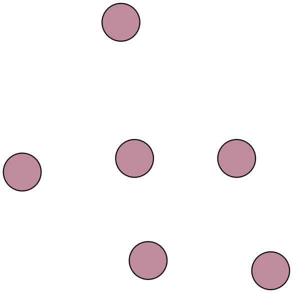
Rosindell et al. (2010)
Etienne and Rosindell (2010)

Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:

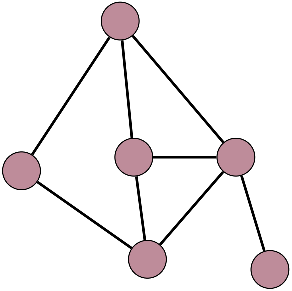
Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population



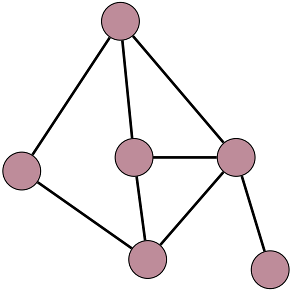
Model: "Split-Drift"

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.



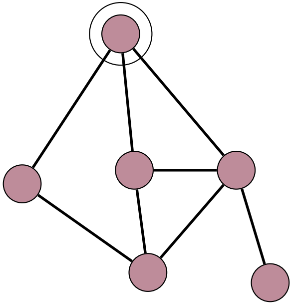
Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:



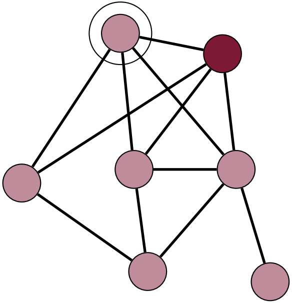
Model: "Split-Drift"

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry



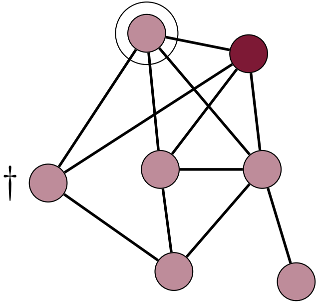
Model: "Split-Drift"

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry



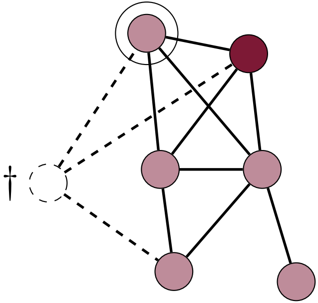
Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry



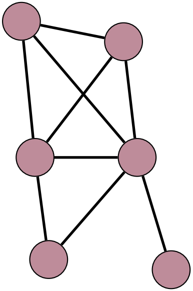
Model: "Split-Drift"

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry



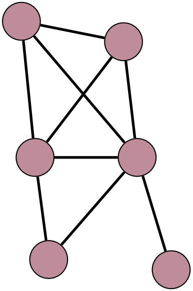
Model: "Split-Drift"

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry



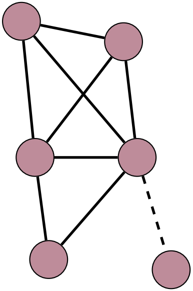
Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry
  - Edge removal (at rate r) $\rightarrow$ reproductive isolation



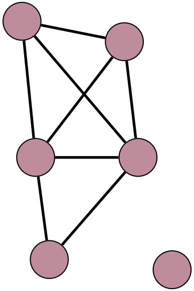
Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry
  - Edge removal (at rate r) $\rightarrow$ reproductive isolation



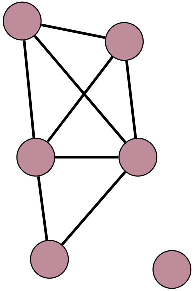
Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry
  - Edge removal (at rate r) $\rightarrow$ reproductive isolation



Model: “Split-Drift”

- n individual populations (constant), represented by a graph $\mathcal{G}$:
  - Vertex: population
  - Edge: interfecundity relationship.
- Two types of events:
  - Vertex duplication (at rate 1) $\rightarrow$ allo/peri/para-patry
  - Edge removal (at rate r) $\rightarrow$ reproductive isolation



$\rightarrow$ Markov chain $(G_n(t))_{t\in\mathbb{R}_+}$ on all graphs of size n.

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?
- Are the data that it generates compatible with what we observe in nature?

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?
- Are the data that it generates compatible with what we observe in nature?

Aims

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?
- Are the data that it generates compatible with what we observe in nature?

Aims

- Describe the expected characteristics of the graph via its stationary state $G_{n,r}$

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?
- Are the data that it generates compatible with what we observe in nature?

Aims

- Describe the expected characteristics of the graph via its stationary state $G_{n,r}$
  - Degree distribution

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?
- Are the data that it generates compatible with what we observe in nature?

Aims

- Describe the expected characteristics of the graph via its stationary state $G_{n,r}$
  - Degree distribution
  - Number of complete subgraphs (species?)

Questions and Aims

Question

- Is this simple, neutral model able to generate clusters of populations?
- Can it inform our definition of a species?
- Are the data that it generates compatible with what we observe in nature?

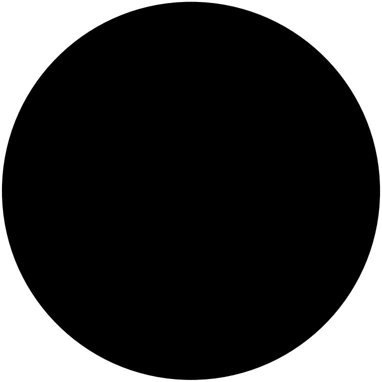
Aims

- Describe the expected characteristics of the graph via its stationary state $G_{n,r}$
  - Degree distribution
  - Number of complete subgraphs (species?)
  - Number of connected components (species?)

Simulation outcomes

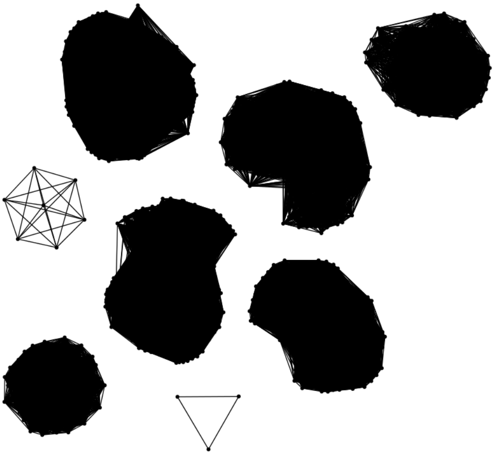
Simulation outcomes

$n = 1000,\ r = 0.1,\ \#CC = 1$



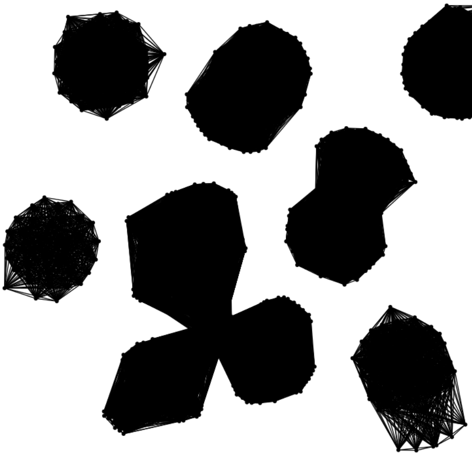
Simulation outcomes

$$n = 1000,\ r = 5,\ \#CC = 8$$



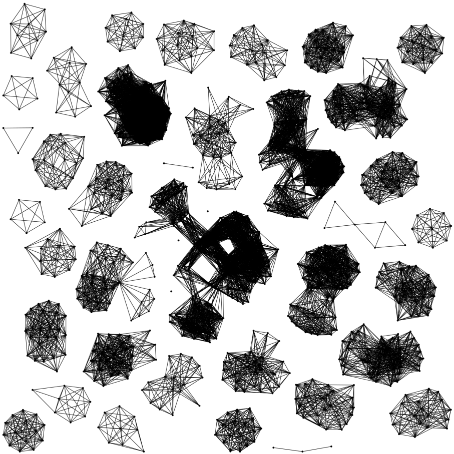
Simulation outcomes

$n = 1000$, $r = 7.5$, $\#CC = 7$



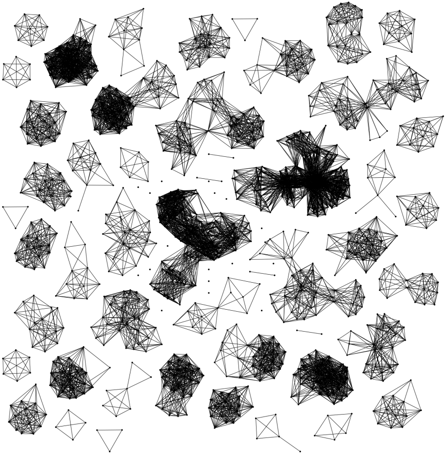
Simulation outcomes

$n = 1000$, $r = 38$, $\#CC = 40$



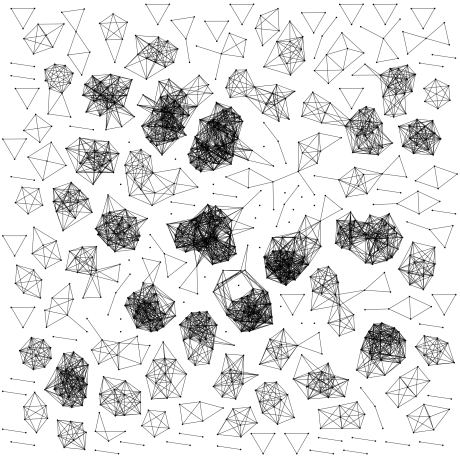
Simulation outcomes

$$n = 1000,\ r = 62,\ \#CC = 66$$



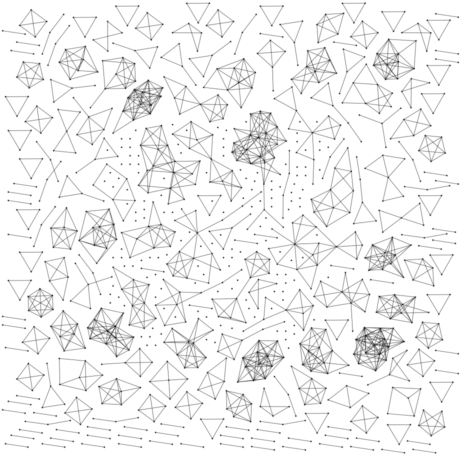
Simulation outcomes

$n = 1000,\ r = 107,\ \#CC = 141$



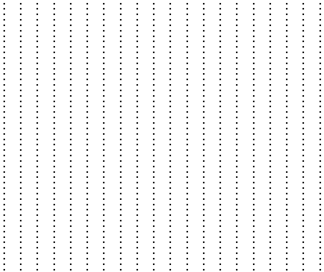
Simulation outcomes

$$n = 1000,\ r = 347,\ \#CC = 342$$



Simulation outcomes

$n = 1000,\ r = 1000,\ \#CC = 342$



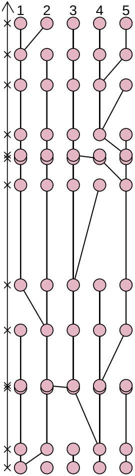
Analysis: Backward-forward approach

Vertex splitting

Edge removal

Analysis: Backward-forward approach

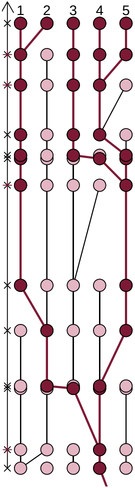
Vertex splitting



Edge removal

Analysis: Backward-forward approach

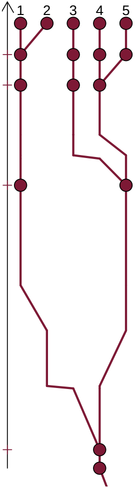
Vertex splitting



Edge removal

Analysis: Backward-forward approach

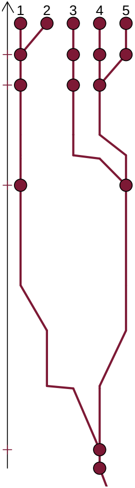
Vertex splitting



Edge removal

Analysis: Backward-forward approach

Vertex splitting

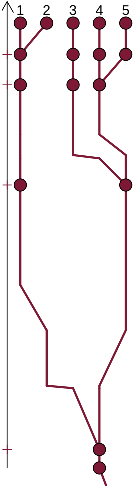


Kingman coalescent (rate 1)

Edge removal

Analysis: Backward-forward approach

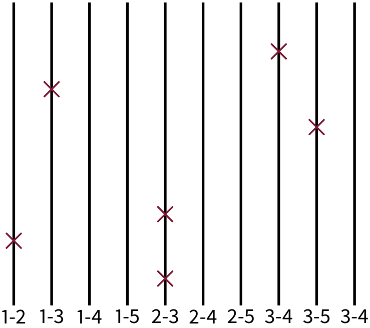
Vertex splitting



Kingman coalescent (rate 1)

and

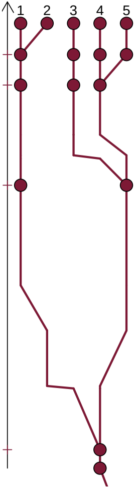
Edge removal



Pairwise Poisson processes (rate r)

Analysis: Backward-forward approach

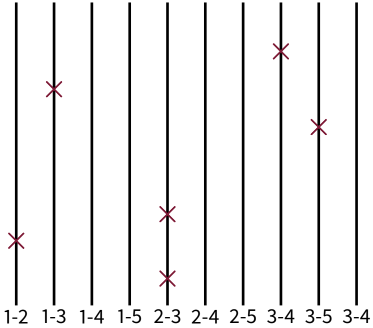
Vertex splitting



Kingman coalescent (rate 1)

and

Edge removal

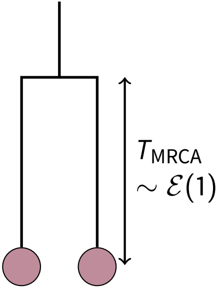


Pairwise Poisson processes (rate *r*)

Degree

Focus on two populations (2 nodes of $G_{n,r}$):

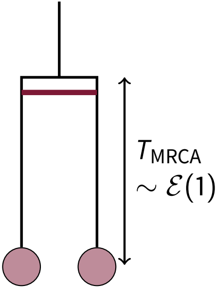
Degree



Focus on two populations (2 nodes of $G_{n,r}$):

- The time to their MRCA is exp(1) (because Kingman),

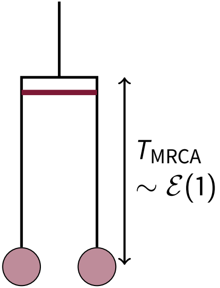
Degree



Focus on two populations (2 nodes of $G_{n,r}$):

- The time to their MRCA is exp(1) (because Kingman),
- The edge formed when the population split is removed at rate r.

Degree



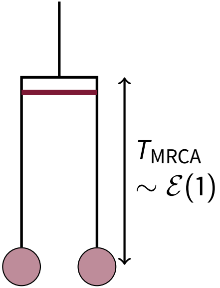
Focus on two populations (2 nodes of $G_{n,r}$):

- The time to their MRCA is exp(1) (because Kingman),
- The edge formed when the population split is removed at rate r.

Probability of interbreeding $=: p(n, r)$

$$p = \frac{1}{1+r}$$

Degree



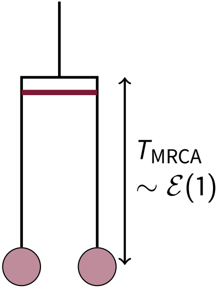
Focus on two populations (2 nodes of $G_{n,r}$):

- The time to their MRCA is exp(1) (because Kingman),
- The edge formed when the population split is removed at rate r.

Probability of interbreeding $=: p(n, r)$

$$p = \frac{1}{1 + r}$$

Degree



Focus on two populations (2 nodes of $G_{n,r}$):

- The time to their MRCA is exp(1) (because Kingman),
- The edge formed when the population split is removed at rate r.

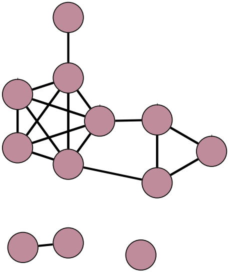
Probability of interbreeding $=: p(n, r)$

$$p = \frac{1}{1+r}$$

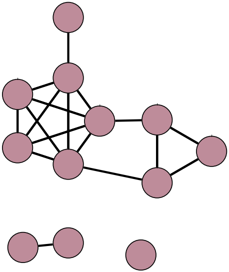
Degree of a fixed node $=: D(n, r)$

$$\mathbb{E}[D] = \frac{n-1}{1+r}, \qquad \mathrm{Var}[D] = \frac{r\,(n-1)\,(n+2r+1)}{(1+r)^2\,(3+2r)}$$

Cliques

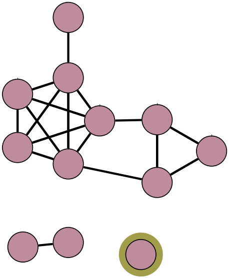


Cliques



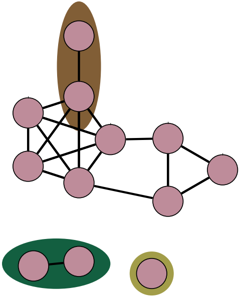
k-clique: complete subgraph of order *k*

Cliques



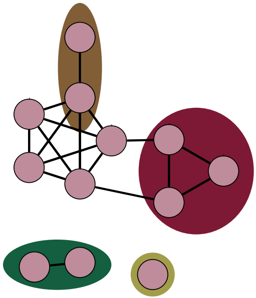
k-clique: complete subgraph of order *k*

Cliques



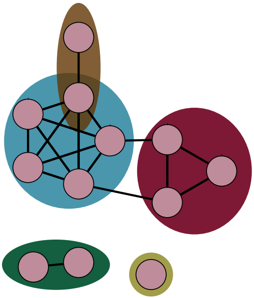
k-clique: complete subgraph of order k

Cliques



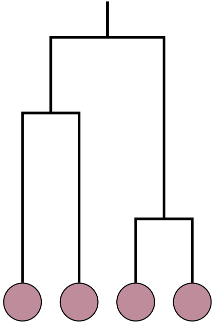
k-clique: complete subgraph of order *k*

Cliques



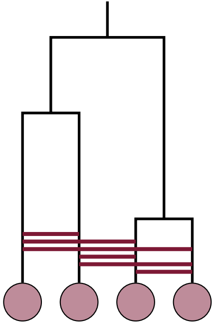
k-clique: complete subgraph of order *k*

Cliques (*continued*)



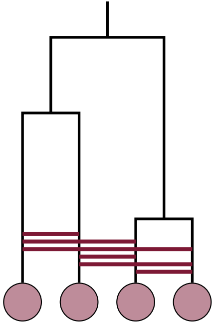
- Consider *k* populations;

Cliques (*continued*)



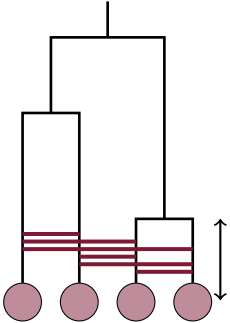
- Consider k populations;
  there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.

Cliques (*continued*)



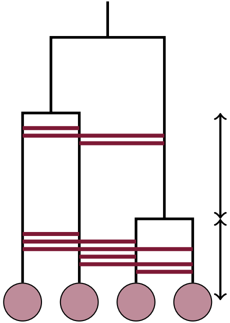
- Consider k populations;
  there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

Cliques (*continued*)



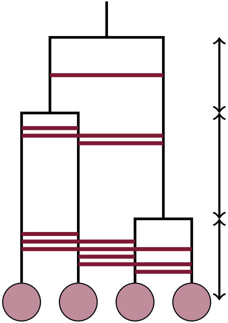
- Consider k populations; there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

Cliques (*continued*)



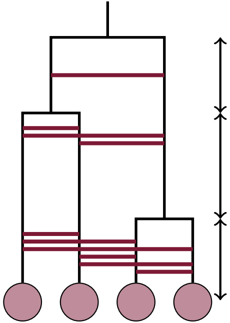
- Consider k populations; there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

Cliques (*continued*)



- Consider k populations;
  there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

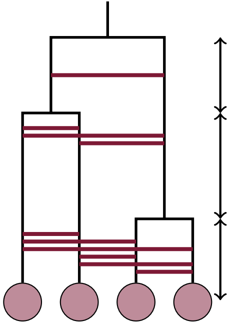
Cliques (*continued*)



- Consider k populations;
  there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

$$\prod_{j=2}^{k} \frac{\binom{j}{2}}{\binom{j}{2} + r\binom{j}{2}} = \left(\frac{1}{1+r}\right)^{k-1}.$$

Cliques (*continued*)



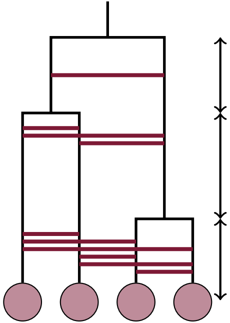
- Consider k populations; there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

$$\prod_{j=2}^{k} \frac{\binom{j}{2}}{\binom{j}{2} + r\binom{j}{2}} = \left(\frac{1}{1+r}\right)^{k-1}.$$

- \# Complete subgraphs of order $k =: X_k(n, r)$

$$\mathbb{E}(X_k) = \binom{n}{k} \left(\frac{1}{1+r}\right)^{k-1}.$$

Cliques (*continued*)



- Consider k populations; there are $\binom{j}{2} = \frac{k(k-1)}{2}$ potential interfecundity links.
- Probability that they all belong to a clique is

$$\prod_{j=2}^{k} \frac{\binom{j}{2}}{\binom{j}{2} + r\binom{j}{2}} = \left(\frac{1}{1+r}\right)^{k-1}.$$

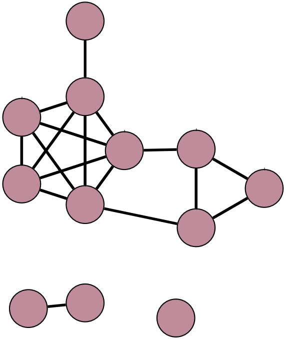
- \# Complete subgraphs of order $k =: X_k(n,r)$

$$\mathbb{E}(X_k) = \binom{n}{k}\left(\frac{1}{1+r}\right)^{k-1}.$$

- Clique number $=: \omega(n,r)$

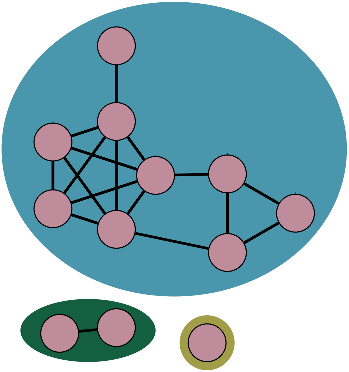
$$\mathbb{P}[\omega \geq k] = \mathbb{P}[X_k \geq 1] \leq \mathbb{E}[X_k].$$

Connected components



Subgraph in which any two nodes are connected to each other by paths, and which is connected to no additional nodes in the supergraph.

Connected components



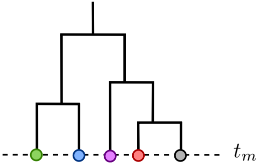
Subgraph in which any two nodes are connected to each other by paths, and which is connected to no additional nodes in the supergraph.

Connected components (*continued*)

Assume $1 \ll r_n \ll n$.

Connected components (*continued*)

Assume $1 \ll r_n \ll n$.
Let $t_k :=$ time when the coalescent tree has k lineages

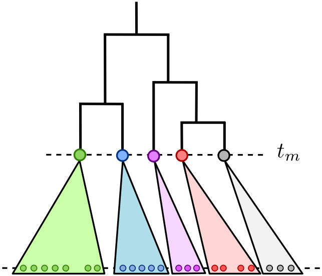


Connected components (*continued*)

Assume $1 \ll r_n \ll n$.
Let $t_k :=$ time when the coalescent tree has k lineages

Lower bound

Choose m such that the graph at time t_m is empty with high probability



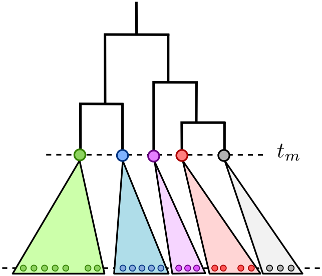
Connected components (*continued*)

Assume $1 \ll r_n \ll n$.
Let $t_k :=$ time when the coalescent tree has k lineages

Lower bound

Choose m such that the graph at time t_m is empty with high probability

Result: $m \sim \dfrac{r_n}{2}$



Connected components (*continued*)

Assume $1 \ll r_n \ll n$.
Let $t_k :=$ time when the coalescent tree has k lineages

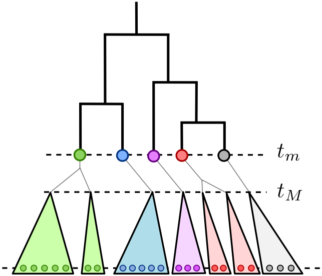
Lower bound

Choose m such that the graph at time t_m is empty with high probability

Result: $m \sim \frac{r_n}{2}$

Upper bound

Choose M such that the descending subtrees of each of the M nodes of time t_M are connected with high probability



Connected components (*continued*)

Assume $1 \ll r_n \ll n$.
Let $t_k :=$ time when the coalescent tree has k lineages

Lower bound

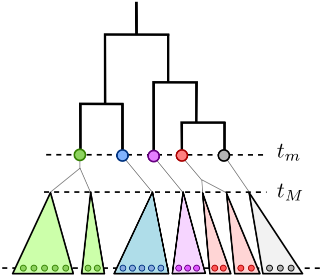
Choose m such that the graph at time t_m is empty with high probability

Result: $m \sim \dfrac{r_n}{2}$

Upper bound

Choose M such that the descending subtrees of each of the M nodes of time t_M are connected with high probability

Result: $M \sim 2 r_n \log(n)$



Connected components (*continued*)

Assume $1 \ll r_n \ll n$.
Let $t_k :=$ time when the coalescent tree has k lineages

Lower bound

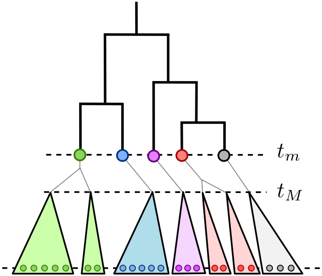
Choose m such that the graph at time t_m is empty with high probability

Result: $m \sim \frac{r_n}{2}$

Upper bound

Choose M such that the descending subtrees of each of the M nodes of time t_M are connected with high probability

Result: $M \sim 2 r_n \log(n)$



Assume that as $n \to \infty$, $r_n \to \infty$ and $r_n / n \to 0$. Then

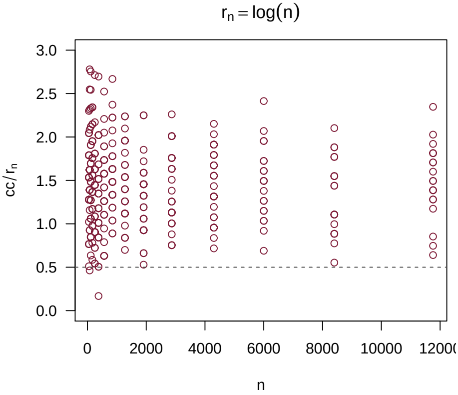
$$\lim_{n \to \infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2 r_n \log(n)\right] = 1.$$

Connected components (*continued*)

Number of species?

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2r_n \log(n)\right] = 1.$$

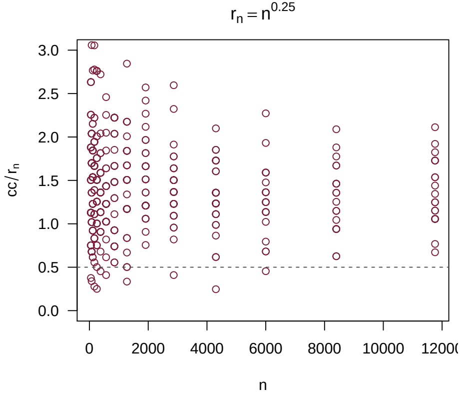
$$\#CC/r_n \text{ vs } n$$



Connected components (*continued*)
Number of species?

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2r_n \log(n)\right] = 1.$$

$\#CC/r_n$ vs n

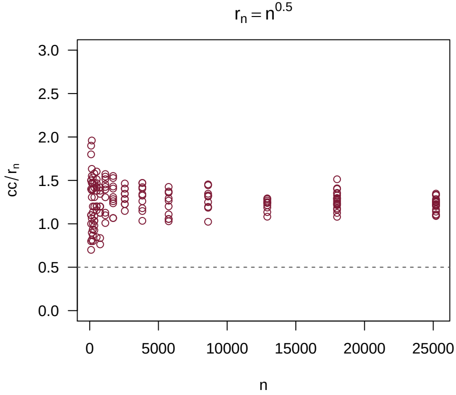


Connected components (*continued*)

Number of species?

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2r_n \log(n)\right] = 1.$$

$$\#CC/r_n \text{ vs } n$$

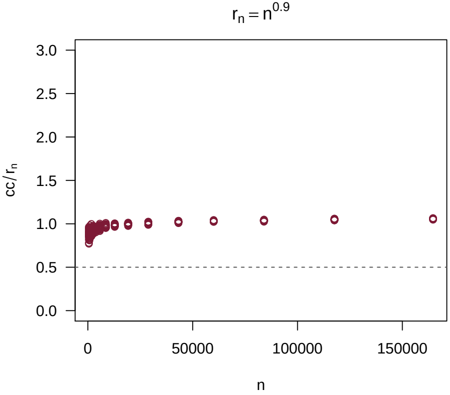


Connected components (*continued*)

Number of species?

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2r_n \log(n)\right] = 1.$$

$\#CC/r_n$ vs n

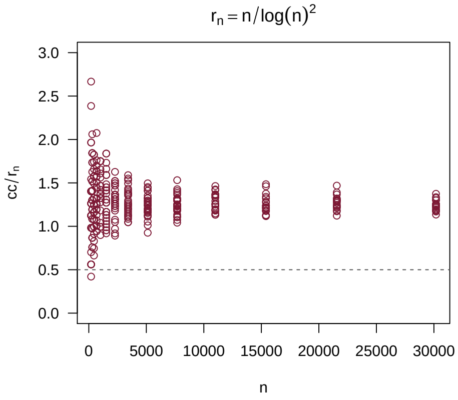


Connected components (*continued*)

Number of species?

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2r_n \log(n)\right] = 1.$$

$\#CC/r_n$ vs n

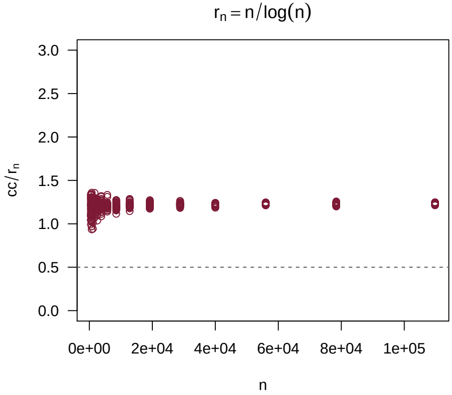


Connected components (*continued*)

Number of species?

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{r_n}{2} \leq \#CC(G_{n,r_n}) \leq 2r_n \log(n)\right] = 1.$$

$$\#CC/r_n \text{ vs } n$$

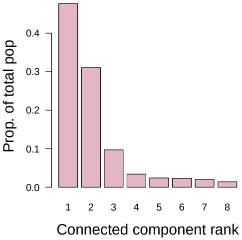


Species Abundance Distributions (SAD)

Species Abundance Distributions (SAD)

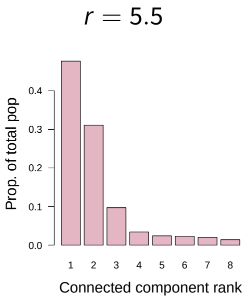
$n = 1000$

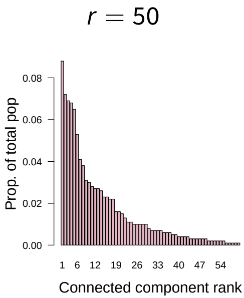
$r = 5.5$



Species Abundance Distributions (SAD)

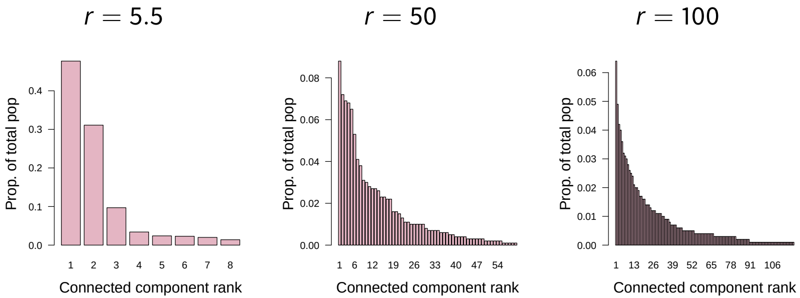
$$n = 1000$$



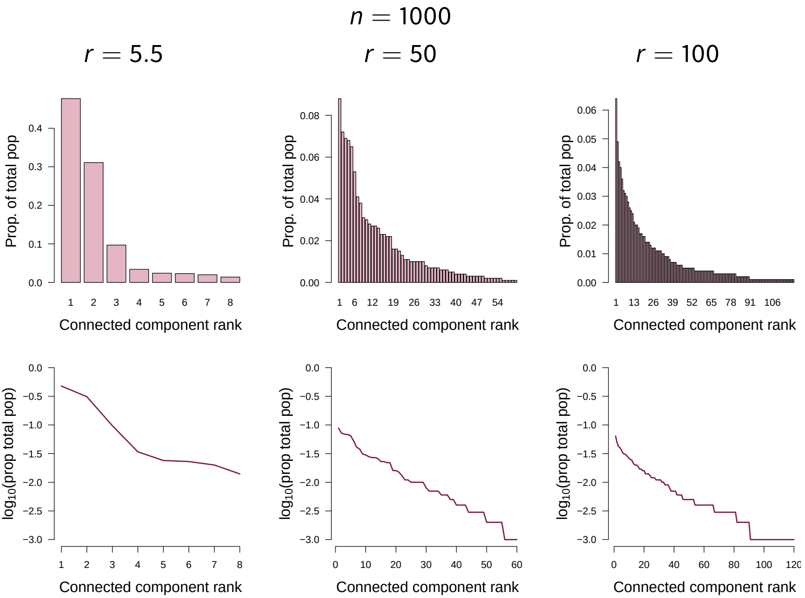


Species Abundance Distributions (SAD)

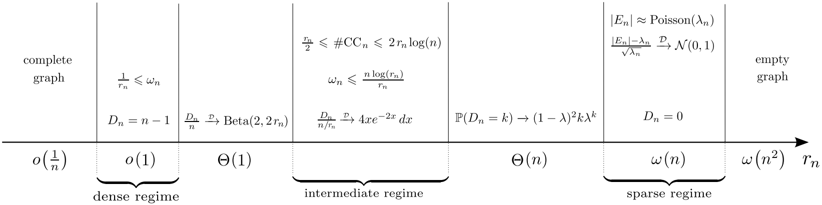
$n = 1000$

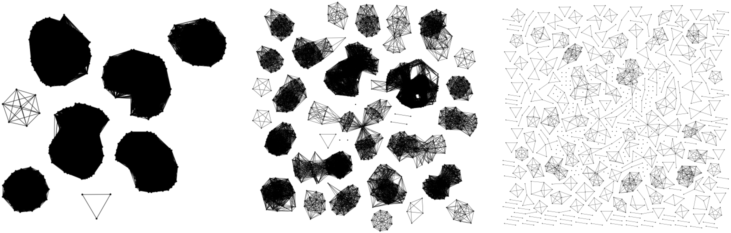


Species Abundance Distributions (SAD)



Summary of the mathematical results





$r = 5$ $\quad$ $r = 41$ $\quad$ $r = 347$

$n = 1000$

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Perspectives

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Perspectives

- Evaluation using real Species-Abundance distributions

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Perspectives

- Evaluation using real Species-Abundance distributions
- Link to macroevolutionary models

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Perspectives

- Evaluation using real Species-Abundance distributions
- Link to macroevolutionary models
  - Species tree?

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Perspectives

- Evaluation using real Species-Abundance distributions
- Link to macroevolutionary models
  - Species tree?
  - Speciation/Extinction rates?

Take-home messages

The Split-and-Drift Random Graph

- A tractable neutral model for the evolution of reproductive isolation
  - Includes a gray zone of speciation
  - Two parameters: number of populations n, rate of loss of interfecundity r
- Open Mathematical Questions
  - Convergence in distribution of $\#CC/r_n$?
  - Convergence in the graphon sense? (dense regime)
  - Distribution of abundances of connected components?

Perspectives

- Evaluation using real Species-Abundance distributions
- Link to macroevolutionary models
  - Species tree?
  - Speciation/Extinction rates?

Thanks for your attention!

Appendix

Degree Distribution

$$\mathbb{P}[D(n,r)=k] = \frac{2r(2r+1)}{(n+2r)(n-1+2r)}(k+1)\prod_{i=1}^{k}\frac{n-i}{n-i+2r-1}.$$

- As $r_n \to r$, $\dfrac{D(n,r_n)}{n}$ converges to a $Beta(2,2r)$ random variable
- If $1 \ll r_n \ll n$, $\dfrac{D(n,r_n)}{n/r_n}$ converges to a size-biased $Exp(2)$.
- If $r_n \sim \rho\, n/2$, $D(n,r_n)$ converges to a size-biased geometric random variable with parameter $1/(1+\rho)$.