

# 1 Kingman's Coalescent

**a**

The only possible transition at time  $t$  is to go from  $N_t$  to  $N_t - 1$ , so the rate as a function of the number of particles  $k$  is

$$g(k, k-1) = \begin{cases} \binom{k}{2} & \text{for } 1 < k \leq L \\ 0 & \text{otherwise.} \end{cases}$$

The generator is given by

$$\begin{aligned} (\mathcal{L}f)(k) &= \sum_{\substack{j \in S \\ j \neq k}} g(k, j)[f(j) - f(k)] \\ &= g(k, k-1)[f(k-1) - f(k)] \\ &= \binom{k}{2} [f(k-1) - f(k)], \end{aligned}$$

and the master equation

$$\begin{aligned} \frac{d}{dt}\pi_t(k) &= \sum_{j \neq k} \pi_t(j)g(j, k) - \sum_{j \neq k} \pi_t(k)g(k, j) \\ &= \pi_t(k+1)g(k+1, k) \\ &= \pi_t(k+1) \binom{k+1}{2} - \pi_t(k) \binom{k}{2}. \end{aligned}$$

The process has one absorbing state at  $k = 1$  since it needs two particles to coalesce, so once it reaches the one particle state nothing will happen. This also describes the stationary distribution given by  $(1, 0, 0, \dots)$ , and since the process converges to this stationary distribution from any starting state, the process is ergodic.

**b**

Since the expected staying time in state  $k$  is  $1/g(k, k-1)$ , the total time to go from  $k = L$  to  $k = 1$  is the sum of all the expected times, i.e.,

$$\begin{aligned} \mathbb{E}(T) &= \sum_{k=2}^L \left[ \binom{k}{2} \right]^{-1} = \sum_{k=2}^L \frac{2! (k-2)!}{k!} \\ &= 2 \sum_{k=2}^L \frac{1}{k(k-1)} = 2 \sum_{k=2}^L \left( \frac{1}{k-1} - \frac{1}{k} \right) \\ &= 2 \left[ \left( \frac{1}{1} - \frac{1}{2} \right) + \left( \frac{1}{2} - \frac{1}{3} \right) + \dots + \left( \frac{1}{1-L} - \frac{1}{L} \right) \right] \\ &= 2 \left( 1 - \frac{1}{L} \right) \end{aligned}$$

□

**c**

The generator for the rescaled process  $N_t/L$  is given by

$$(\mathcal{L}f)(x) = \binom{xL}{2} \left[ f\left(x - \frac{1}{L}\right) - f(x) \right],$$

where  $x = k/L$ . Using Taylor expansion for  $f(x - 1/L)$  around  $f(x)$  gives

$$\begin{aligned} f\left(x - \frac{1}{L}\right) &= f(x) - \frac{1}{L}f'(x) + \frac{1}{2L^2}f''(x) + \dots, \text{ so} \\ (\mathcal{L}f)(x) &= \binom{xL}{2} \left[ \frac{1}{2L^2}f''(x) - \frac{1}{L}f'(x) \right] \end{aligned}$$

Slowing the process down just applies a factor of  $1/L$  to the rates. The generator for  $X_t^L = \frac{1}{L}N_{t/L}$  is then

$$\begin{aligned} (\mathcal{L}f)(x) &= \frac{1}{L} \cdot \frac{(Lx)!}{2!(Lx-2)!} \left[ \frac{1}{2L^2}f''(x) - \frac{1}{L}f'(x) \right] \\ &= \frac{(\mathbb{X}x)(Lx-1)}{2\mathbb{X}} \left[ \frac{1}{2L^2}f''(x) - \frac{1}{L}f'(x) \right] \\ &= \frac{1}{2} \left[ \cancel{\frac{x^2}{2L}f''(x)} - x^2f'(x) - \cancel{\frac{x^2}{2L}f''(x)} + \cancel{\frac{x}{L}f'(x)} \overset{0}{\rightarrow} \right] \\ &= -\frac{x^2}{2}f'(x) \text{ as } L \rightarrow \infty. \end{aligned}$$

Since the process  $X_t^L$  starts at  $x = L/L = 1$  and always terminates at  $x = 1/L \neq 0$ , the state space  $(0, 1]$  is open at 0 and  $X_0 = 1$ .

We see that this corresponds to the generator of a diffusion process with only the drift term but no diffusion which specifies the randomness. Therefore the process is deterministic because there is no randomness.

In order to compute  $X_t$  explicitly, we start with an SDE with no diffusion term and integrate both sides

$$\begin{aligned} dX_t &= a(X_t, t)dt \\ \int_0^t dX_s &= \int_0^t \frac{-X_s^2}{2} ds \\ \int_0^t X_s^{-2} dX_s &= - \int_0^t \frac{1}{2} ds \\ X_s^{-1} \Big|_0^t &= \frac{1}{2} s \Big|_0^t \end{aligned}$$

By evaluating this integral and substituting  $X_0 = 1$ , we arrive at the explicit solution

$$X_t = \frac{2}{2+t}.$$

This is compatible with the result from part b) because the results are for differently scaled processes. i.e.,  $2/(2+t)$  describes the path of  $X_t^L$  which we have defined to be  $\frac{1}{L}N_{t/L}$ , whereas the mean time to absorption calculated in part b) is for the unscaled process  $N_t$ . By applying the appropriate scaling, we have

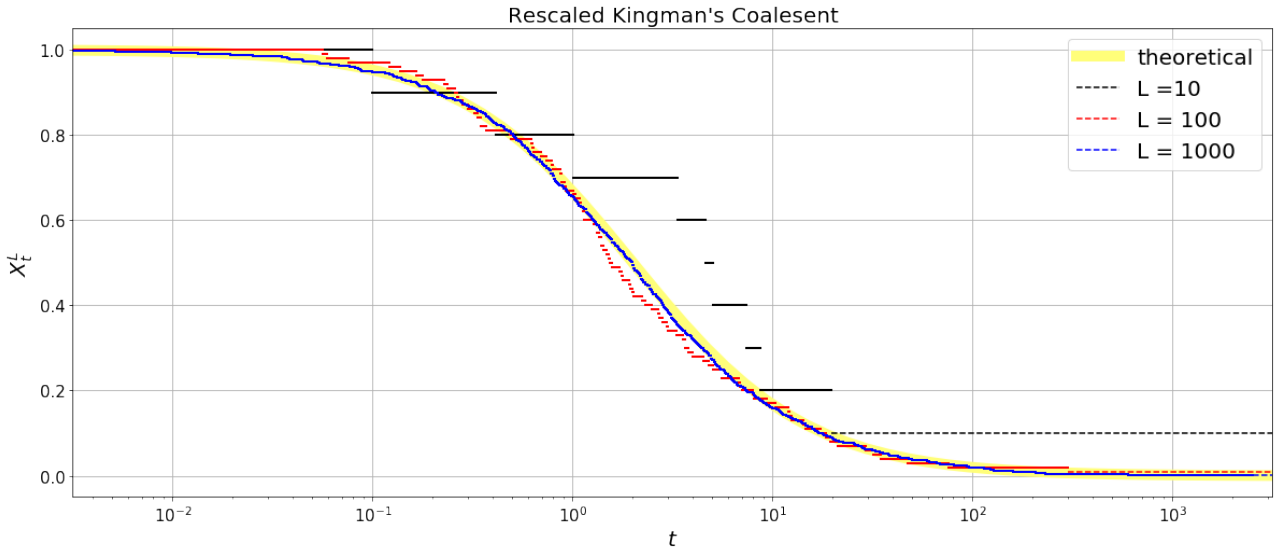
$$N_t = L \cdot \frac{2}{2+tL},$$

from which we can deduce that

$$N_t = 1 \implies t = \mathbb{E}(T) = 2\left(1 - \frac{1}{L}\right),$$

which is in line with our results from part b).

**d**



## 2 Orstein-Uhlenbeck process

**a**

Let  $f(X_t) = X_t$ , then  $f'(X_t) = 1$  and  $f''(X_t) = 0$ , and using the evolution equation,

$$\frac{d}{dt}\mathbb{E}[X_t] = \mathbb{E}[-\alpha X_t] = -\alpha\mathbb{E}[X_t],$$

so

$$\frac{d}{dt}\mathbb{E}[X_t] + \alpha\mathbb{E}[X_t] = 0$$

is the required ODE. Solving this gives  $m(t) := \mathbb{E}[X_t] = Ae^{-\alpha t}$ , where  $A$  is an arbitrary constant.

Similarly, let  $f(X_t) = X_t^2$ , then  $f'(X_t) = 2X_t$  and  $f''(X_t) = 2$ , and

$$\frac{d}{dt}\mathbb{E}[X_t^2] = \mathbb{E}[-\alpha X_t \cdot 2X_t + \frac{1}{2}\sigma^2 \cdot 2] = -\alpha\mathbb{E}[X_t^2] + \sigma^2,$$

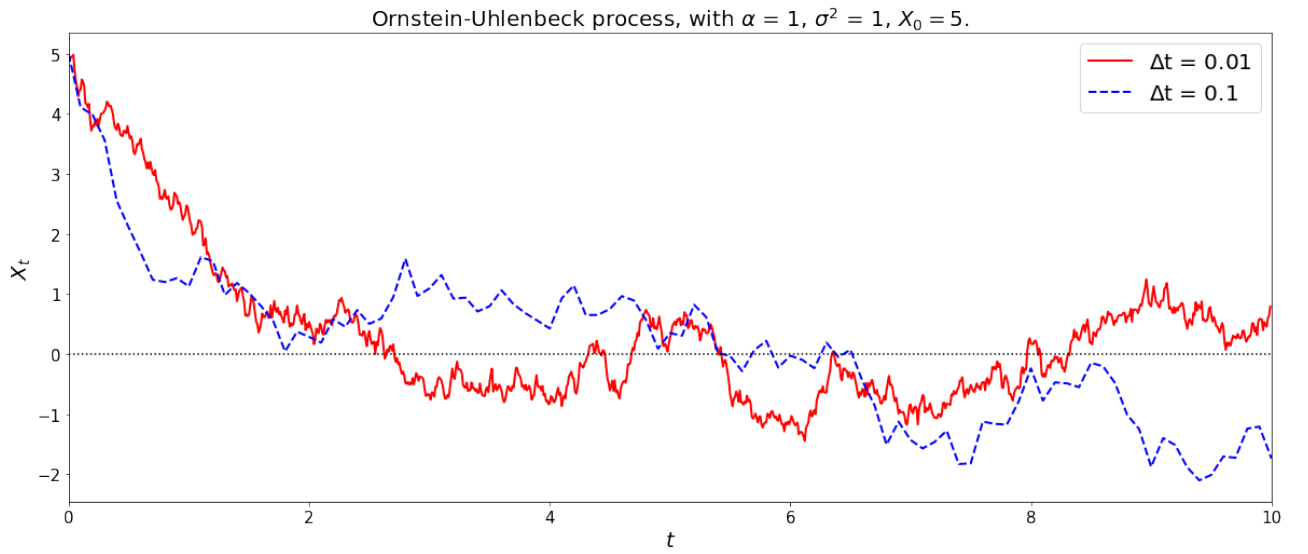
so

$$\frac{d}{dt}\mathbb{E}[X_t^2] + 2\alpha\mathbb{E}[X_t^2] = \sigma^2,$$

which can be solved to give  $\mathbb{E}[X_t^2] = \sigma^2/2\alpha + Be^{-2\alpha t}$ , where  $B$  is an arbitrary constant. Therefore,

$$\begin{aligned} v(t) = \mathbb{E}[X_t^2] - (\mathbb{E}[X_t])^2 &= \frac{\sigma^2}{2\alpha} + Be^{-2\alpha t} - (Ae^{-\alpha t})^2 \\ &= \frac{\sigma^2}{2\alpha} + Ce^{-2\alpha t} \end{aligned}$$

**c**



### 3 Moran model and Wright-Fisher diffusion

**a**

The state space of  $(X_t : t \geq 0)$  is  $\{(x_1, x_2, \dots, x_L) : x_i \in \{1, 2, \dots, L\}\}$ . The process is not reducible since once a type has been completely killed off, the process can no longer return to a state where one of the individuals is of that type.

The stationary distributions are  $\{(x, x, \dots, x) : x \in \{1, 2, \dots, L\}\}$  because the process is only stationary once only a single type survives.

**b**

$(N_t : t \geq 0)$  is a Markov process and its state space is  $\{0, 1, \dots, L\}$ .

The rates of this process for  $x \neq 0$  or  $L$  is

$$\begin{cases} g(x, x+1) &= x \cdot \frac{L-x}{L-1} \\ g(x, x-1) &= (L-x) \cdot \frac{x}{L-1} \end{cases} \dots 28????$$

The reasoning is that in order to gain an individual, one of the  $x$  current individuals must impose its type on someone of a different type. Since it does so randomly, the probability of killing someone of a different type is  $\frac{L-x}{L-1}$ . We also get  $g(x, x-1)$  by applying similar logic.

The generator is

$$(\mathcal{L}f)(x) = \frac{x(L-x)}{L-1} [f(x+1) - 2f(x) + f(x-1)].$$

The process is, once again, not irreducible since  $x = 0$  and  $x = L$  are absorbing states. This also means that there exist infinitely many stationary distributions of the form

$$(\alpha, 0, \dots, 1-\alpha), \alpha \in [0, 1].$$

The limiting distribution as  $t \rightarrow \infty$  is

$$\left( \frac{L-1}{L}, 0, \dots, 0, \frac{1}{L} \right)$$

since the probability of any one individual dominating is exactly the same due to symmetry, and there was only one individual of type  $k$  at  $t = 0$ .

## c

By setting  $f(N_t) = N_t$  and using the evolution equation, we obtain

$$\frac{d}{dt} \mathbb{E}[N_t] = \mathbb{E} \left[ \frac{N_t(L-N_t)}{L-1} \left( (N_t+1) - 2(N_t) + (N_t-1) \right) \right] = 0,$$

which means that  $N_t$  is constant in time. Given that  $N_0 = n$ , we may deduce that  $m_1(t) \equiv n$ .

Similarly, by setting  $f(N_t) = N_t^2$ ,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[N_t^2] &= \mathbb{E} \left[ \frac{N_t(L-N_t)}{L-1} \left( (N_t^2 + 2N_t + 1) - 2(N_t^2) + (N_t^2 - 2N_t + 1) \right) \right] \\ &= \mathbb{E} \left[ \frac{2N_t(L-N_t)}{L-1} \right] \\ &= \frac{2L}{L-1} \mathbb{E}[N_t] - \frac{2}{L-1} \mathbb{E}[N_t^2] \end{aligned}$$

i.e.,

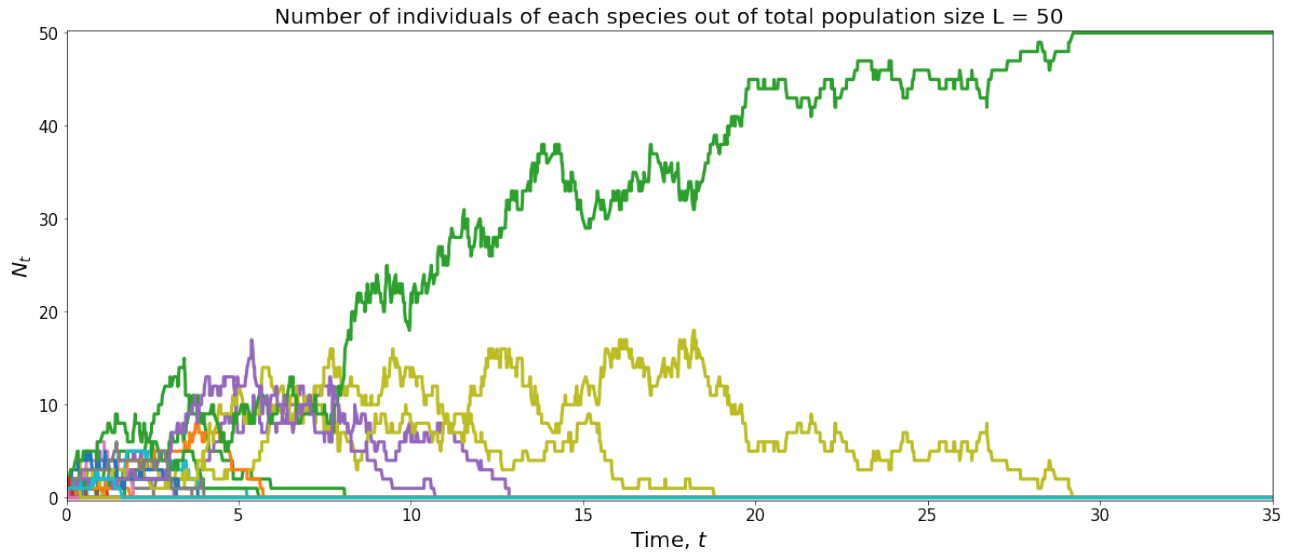
$$\frac{d}{dt} \mathbb{E}[N_t^2] + \frac{2}{L-1} \mathbb{E}[N_t^2] = \frac{2Ln}{L-1}.$$

Solving this gives

$$m_2(t) = \mathbb{E}[N_t^2] = Ln + Ae^{\frac{2t}{1-L}}$$

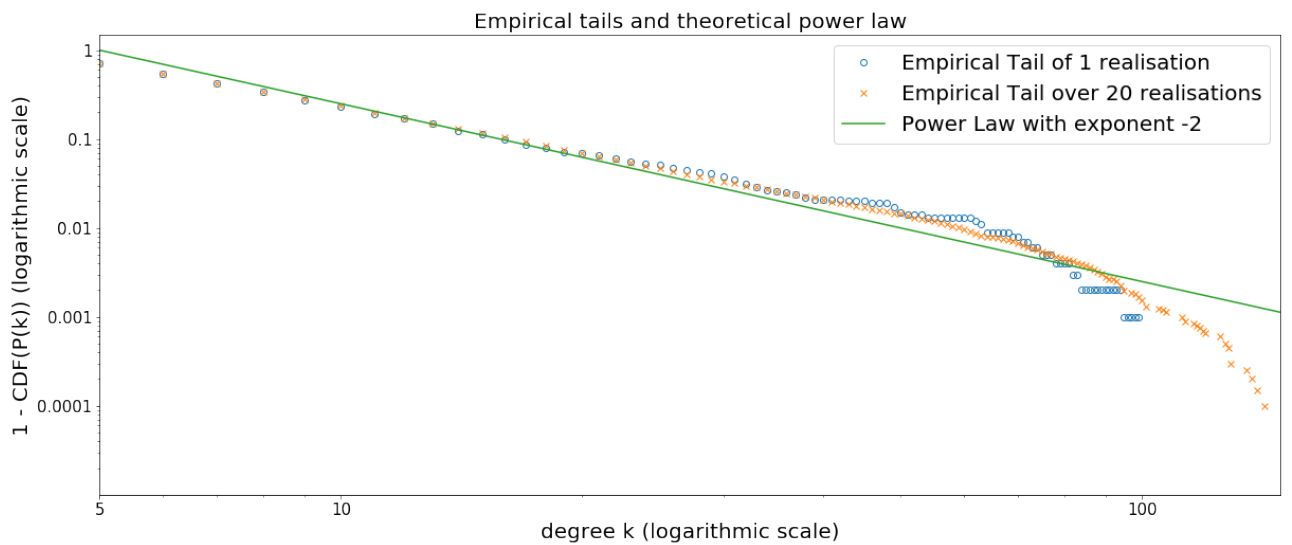
In the scaling limit  $t \rightarrow \infty$ , the exponential term vanishes so we just have  $m_2(t) \rightarrow Ln$ .

**f**

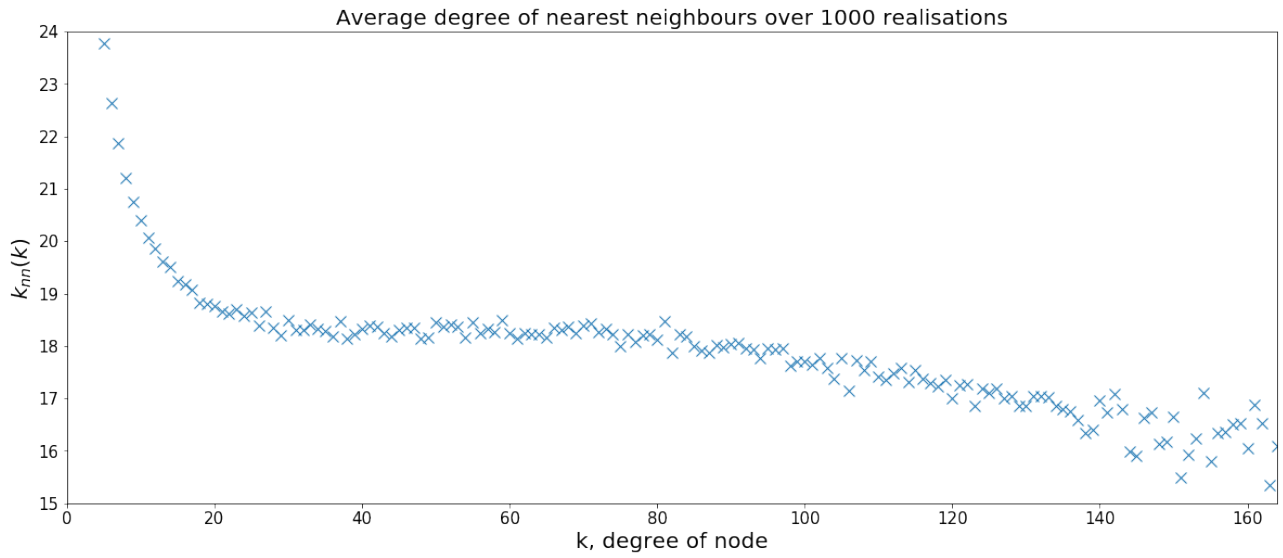


## 4 Barabási-Albert Model

**a**



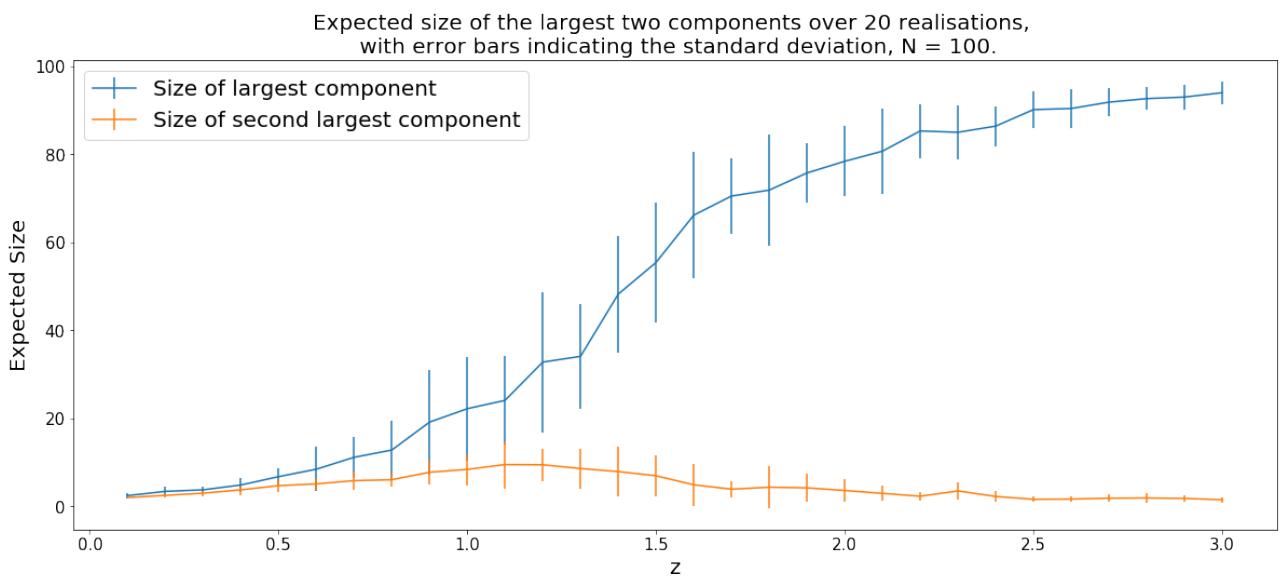
b

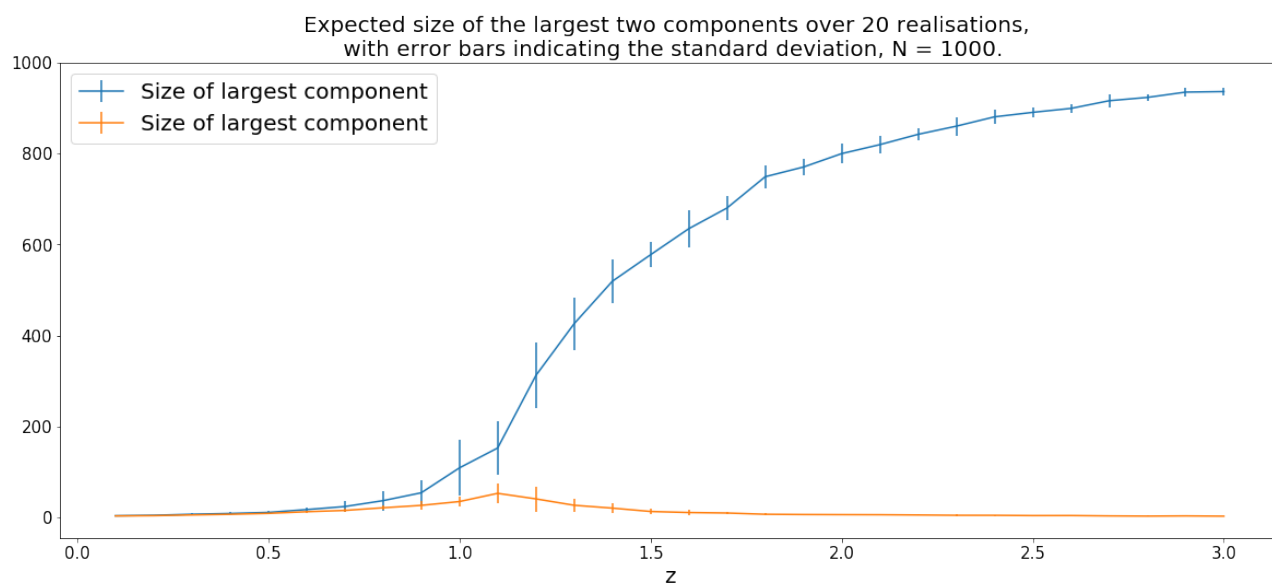


The graphs appear to be typically disassortative. It is apparent that  $k_{nn}(k)$  is high if  $k$  is low and vice versa, hence they are disassortative. While the plot showed a lot more spread in the  $t \geq 20$  range for lower numbers of realisations, running it over 1000 realisations seemed to help smooth out the plot a great amount.

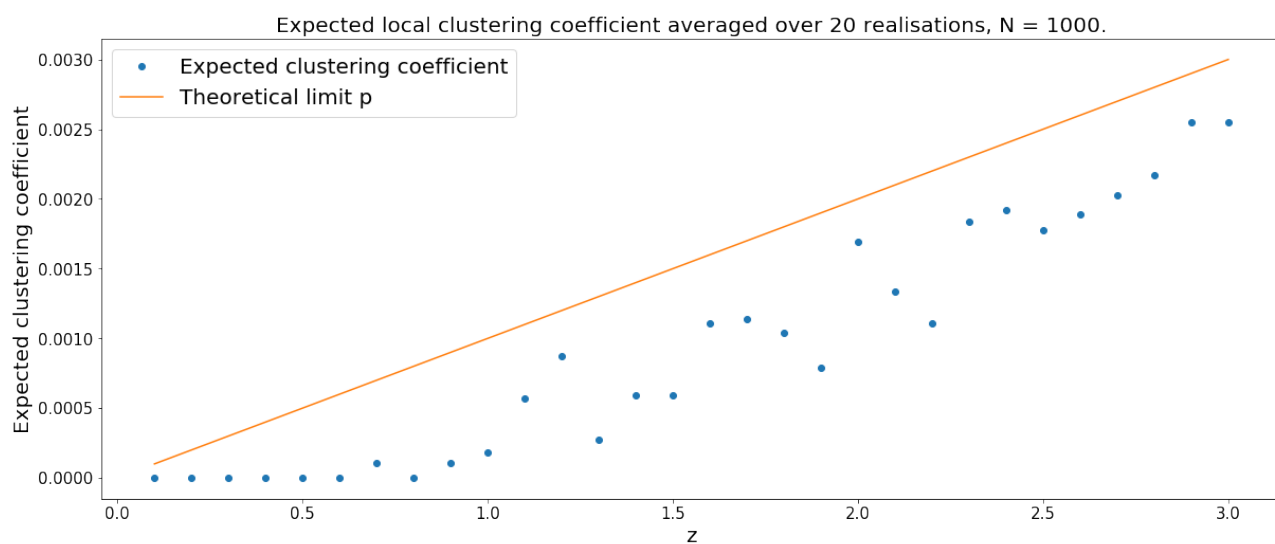
## 5 Erdős Rényi Random Graphs

a





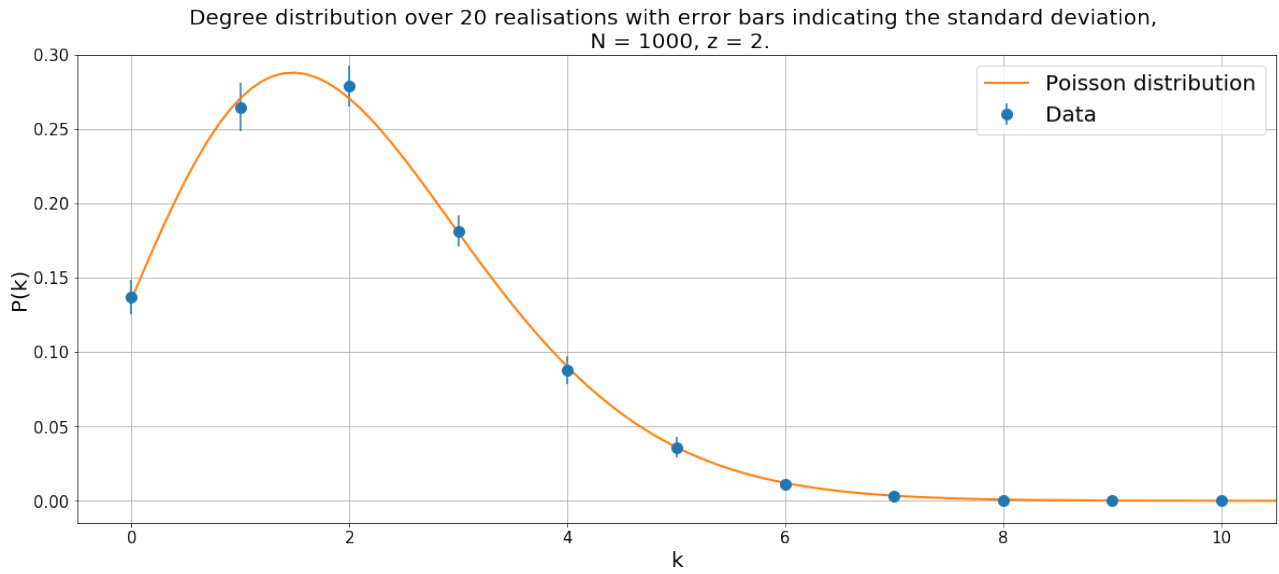
b



The theoretical limit  $p$  line fits a lot better when we allow for higher values of  $z$ .



**C**



I interpreted this question as collate all 20 realisations together to obtain one single degree distribution  $p(k)$  but then discovered that some people have plotted all the realisations separately. Therefore I settled and added some error bars.