

## 0.1 NOTATION

Any probability  $p$  is implicitly a function of  $n$ , so keep this in mind when taking limits as  $n \rightarrow \infty$ . You can't just treat  $p$  like a constant.

The  $O, \Omega, \Theta, o, \omega$  notation is all standard. If

$$\lim_{n \rightarrow \infty} \frac{A_n}{B_n} = 1$$

then we say  $A_n \approx B_n$ . We abuse notation a bit by saying  $A_n \leq B_n$  if there is some constant  $c$  such that  $A_n \leq cB_n$  for all  $n$ .

We say that  $A_n$  converges weakly to  $A$  if

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(A_n)] = \mathbb{E}[f(A)]$$

for all bounded continuous functions  $f$ . We denote this by  $A_n \Rightarrow A$ . (definition in terms of CDFs instead?)

## 0.2 ERDÖS-RÉNYI RANDOM CLIQUE COMPLEXES

**Important.** In the Erdős-Rényi clique complex, if  $p$  is in a certain regime, each expected Betti number will be nonzero with high probability. In addition, each Betti number will satisfy a central limit theorem.

**Note 1.** Throughout this section, assume  $p$  is in the following regime:

$$\begin{aligned} p &= \omega(n^{-1/k}), \\ p &= o(n^{-1/(k+1)}). \end{aligned}$$

**Lemma 1** (Morse Inequalities). Let  $f_k$  denote the number of  $k$ -dimensional faces of a simplicial complex  $\Delta$ , and let  $\beta_k$  denote the  $k$ -th Betti number of  $\Delta$ . Then

$$-f_{k-1} + f_k - f_{k+1} \leq \beta_k \leq f_k.$$

*Proof.* Do this. Uses definition of simplicial homology and the rank nullity theorem.  $\square$

**Theorem 1.**  $\beta_k$  is nonzero with high probability. Explicitly,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[\beta_k]}{n^{k+1} p^{\binom{k+1}{2}}} = \frac{1}{(k+1)!}.$$

*Proof.* The desired limit relation is actually straightforward to show for  $f_k$  instead of  $\beta_k$ . So we'll do that, then use our assumptions on  $p$  and the Morse inequalities to show that  $\beta_k$  has the same property.

Note that  $f_k$  also represents the number of  $(k+1)$ -cliques of our complex. Since the complex is Erdős-Rényi, each of the  $\binom{n}{k+1}$  possible  $(k+1)$ -cliques occur with the same probability. Since a  $(k+1)$ -clique has  $\binom{k+1}{2}$  distinct edges, this probability is  $p^{\binom{k+1}{2}}$ .

But  $f_k$  is really just the sum of  $\binom{n}{k+1}$  indicator functions, each tracking whether or not a particular  $(k+1)$ -clique is present in the complex. Since each clique has equal probability of forming, this means the expectation of  $f_k$  is

$$\mathbb{E}[f_k] = \binom{n}{k+1} p^{\binom{k+1}{2}} = \frac{n!}{(n-k-1)!(k+1)!} p^{\binom{k+1}{2}}.$$

The limit of our desired quantity, but with  $f_k$  substituted in place of  $\beta_k$ , is then

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[f_k]}{n^{k+1} p^{\binom{k+1}{2}}} = \lim_{n \rightarrow \infty} \frac{n \cdots (n-k)}{n^{k+1} (k+1)!} = \frac{1}{(k+1)!}.$$

We can then use our assumption on the regime of  $p$  and the Morse inequalities to show that  $\beta_k$  has the same property. Since  $p = \omega(n^{-1/k})$ ,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[f_{k-1}]}{\mathbb{E}[f_k]} = \lim_{n \rightarrow \infty} \frac{k+1}{np^k} = 0.$$

And since  $p = o(n^{-1/(k+1)})$ , we similarly have  $\lim_{n \rightarrow \infty} \mathbb{E}[f_{k+1}] / \mathbb{E}[f_k] = 0$ . Thus in this regime of  $p$ ,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[-f_{k-1} + f_k - f_{k+1}]}{\mathbb{E}[f_k]} = 1.$$

Then by the Morse inequalities,  $\beta_k$  must satisfy the desired limit property.  $\square$

**Corollary 1.** As  $n \rightarrow \infty$ ,

$$\mathbb{E}[-f_{k-1} + f_k - f_{k+1}] = \mathbb{E}[\beta_k] = \mathbb{E}[f_k].$$

*Proof.* This follows from the last equation of the previous proof.  $\square$

Now we know that  $\beta_k$  is very likely nonzero (at least in our specific regime of  $p$ ). Since it's not just some trivial quantity, a central limit theorem for it is a meaningful statement.

To do so, we'll use the Morse inequalities again. Proving central limit theorems for the non- $\beta_k$  terms in the inequalities is easier, and we can use them to get the desired property for  $\beta_k$  itself. To clean up notation, let  $g_k \doteq -f_{k-1} + f_k - f_{k+1}$ .

**Lemma 2.** As  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} \frac{\text{Var}(f_k)}{\text{Var}(g_k)} = 1.$$

*Proof.* Do this. □

**Corollary 2.** As  $n \rightarrow \infty$ ,

$$\text{Var}(g_k) = \text{Var}(\beta_k) = \text{Var}(f_k).$$

**Lemma 3.**  $f_k$  obeys a central limit theorem, i.e. as  $n \rightarrow \infty$ ,

$$\frac{f_k - \mathbb{E}[f_k]}{\sqrt{\text{Var}(f_k)}} \Rightarrow \mathcal{N}(0, 1).$$

*Proof.* Do this. □

**Lemma 4.**  $g_k$  obeys a central limit theorem, i.e. as  $n \rightarrow \infty$ ,

$$\frac{g_k - \mathbb{E}[g_k]}{\sqrt{\text{Var}(g_k)}} \Rightarrow \mathcal{N}(0, 1).$$

*Proof.* Do this. □

**Theorem 2.** Suppose  $p = \omega(n^{-1/k})$  and  $p = o(n^{-1/(k+1)})$ , then

$$\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \Rightarrow \mathcal{N}(0, 1).$$

*Proof.* I'm 99% sure that this has some serious logical errors. I don't think you can just move a limit inside of a probability, but that's what I'm doing I think. The original paper authors don't really explain what they did, though, so I'm confused. They say that the theorem has been proven once they have

$$\Phi(t) \leq \mathbb{P}\left(\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t + \varepsilon\right)$$

for arbitrary  $\varepsilon$ , but this isn't really what we're trying to prove. If we can move limits inside the probabilities, though, then our earlier corollaries give us the result. But if we can do that, then why didn't they do what I do below? It seems a lot simpler. Since  $\beta_k \leq f_k$ ,

$$\mathbb{P}\left(\frac{f_k - \mathbb{E}[f_k]}{\sqrt{\text{Var}(f_k)}} \leq t\right) \leq \mathbb{P}\left(\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t\right).$$

Now let  $n \rightarrow \infty$ . By Lemma 3, this becomes

$$\Phi(t) \leq \mathbb{P}\left(\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t\right).$$

By Corollaries 1 and 2, we get  $\mathbb{E}[\beta_k] = \mathbb{E}[f_k]$  and  $\text{Var}(f_k) = \text{Var}(\beta_k)$  in the limit. Thus as  $n \rightarrow \infty$ ,

$$\Phi(t) \leq \mathbb{P}\left(\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t\right),$$

so the CDF of  $\beta_k$  is bounded below by the CDF of the standard normal distribution. Similarly, since  $g_k \leq \beta_k$ ,

$$\mathbb{P}\left(\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t\right) \leq \mathbb{P}\left(\frac{g_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t\right).$$

Letting  $n \rightarrow \infty$  and applying Corollaries 1 and 2 and Lemma 4, this becomes

$$\mathbb{P}\left(\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \leq t\right) \leq \Phi(t).$$

Since the lower and upper bounds are both  $\Phi(t)$ , we conclude

$$\frac{\beta_k - \mathbb{E}[\beta_k]}{\sqrt{\text{Var}(\beta_k)}} \Rightarrow \mathcal{N}(0, 1),$$

as desired. □