We would like to thank the reviewers for their positive and interesting comments that will help us to enhance our manuscript. Please find our answers below.

To Reviewer 1. *Wasting information under random sampling.* Indeed random sampling is inefficient and waste information since it probes edges between clusters uniformly; whereas intuitively of course, an optimal adaptive algorithm gathers 'more' edge information between two clusters that are hard to distinguish.

Experiments with different a and b. We thank the reviewer for this nice suggestion. In the revised paper, we will include a plot showing the proportion of misclassified nodes as a function of a and b, and compare the plot to the theoretical results. Doing such a figure requires a lot of time, since we have to run our algorithm for a large number of problem instances. From our past experiences, we are confident that the plot will match and illustrate the theoretical results well.

Complexity of the ASP algorithm. Thanks for this question that we should have answered in the paper. The complexity of the ASP algorithm is polynomial to both n and T. Indeed, Step 1 (see Algorithm 1), including the Spectral Clustering Algorithm, requires $O(T\log(n))$ operations. Step 2 requires $O(T)$ operations to estimate parameters and Step 3 solves a linear program where the number of variables is k^2 which does not scale with n and T. The remaining steps simply check the log-likelihood values of each node, which requires $O(T)$ computations. Overall, the computational complexity of ASP is $O(T\log n)$.

Derivation of the bound on Page 2 from the main theorem. Let $p = p_{11} = p_{22}$ and $q = p_{12} = p_{21}$. When $KL(p,q) \geq KL(q,p)$, $D(\boldsymbol{p},\boldsymbol{\alpha}) = 2KL(p,q)$. Since both $p = \frac{a\log(n)}{n}$ and $q = \frac{b\log(n)}{n}$ are $o(1)$, we can derive $\frac{KL(p,q)}{\log(n)/n} = a\log(\frac{a}{b}) + (b-a)(1+o(1))$. Therefore, $\frac{nD(\boldsymbol{p},\boldsymbol{\alpha})}{\log(n)} \geq 1$ if and only if $a\log(\frac{a}{b}) + (b-a) \geq 1$. Analogously, we can conclude $b\log(\frac{b}{a}) + (a-b) \geq 1$ is equivalent to $\frac{nD(\boldsymbol{p},\boldsymbol{\alpha})}{\log(n)} \geq 1$ when $KL(p,q) \leq KL(q,p)$. We will add this discussion to the revised paper.

Which clustering algorithm did the author use for the red group in Figure 1? The spectral partition algorithm described in [16] is used. This algorithm is proved to be optimal in terms of error rate in [16].

To Reviewer 2. *Readability of the paper.* The paper is on the theoretical side, and requires rather technical and novel proofs. We agree that for readers not familiar to the stochastic block model and spectral clustering methods, it may be difficult to follow. We will make significant efforts to simplify when appropriate. Note that Rev. 1 and 3 found the paper easy to read and well written, but again, it depends on the reader's background.

Missing definitions and motivations. Note that in Page 2, x_{ij} is just a "dummy" variable for the optimization problem $D(\boldsymbol{p},\boldsymbol{\alpha})$, so it does not require any definition. Later in Page 3, we provide an insightful interpretation of x_{ij}. In each subsection of Section 4.2, we provide the motivation of the corresponding step of Algorithm 1. For instance, Section 4.2.1 explains how to divide the total sampling budget T and Section 4.2.2 states the meaning of $\hat{\boldsymbol{p}}$. We will extend these parts to motivate the algorithm in even more detail.

Numerical validation. The right graph of Figure 1 in our main manuscript Page 3, we compare the error rates of the adaptive spectral partition to that of a non-adaptive spectral partition algorithm described in [16] known to be optimal (in absence of adaptive sampling). We agree that the paper would benefit from more numerical results and will add more experiments in the revised paper.

To Reviewer 3. To explain the novelty of our contributions, we can say that:

1. the paper derives for the first time necessary and sufficient conditions for both asymptotically accurate detection and exact recovery in the Stochastic Block Model (SBM) with adaptive sampling. This was an important open problem in the community interested in the SBM.
2. Our proof techniques are novel and also allow for the first time the derivation of a necessary and sufficient condition holding *with high probability* (see the discussion the paragraph "Deriving fundamental limits" on Page 3 for an explanation why it is challenging). Note that most often, researchers are able to derive fundamental limits for the expected number of misclassified nodes, and devise an algorithm with performance guarantees holding with high probability (with probability tending to 1 as n goes large, the number of misclassified nodes is small).

In summary, in this paper, we manage not only to deal with adaptive sampling, but we also fix the aforementioned gap between fundamental limits and performance guarantees (both hold with high probability). We will emphasize and state the contributions more clearly in the revised paper.