We thank the reviewers for the insightful comments.

On the novelty of our approach (R1,R4). As it is for [14], our article takes inspiration from the works considering the problem in the static regime [9,18,19,28] that, although not formally rigorous, all converge to similar conclusions, recently partially proved in *e.g.* [10-13]. To the best of our knowledge, the proposed article and [14] provide the only methods capable of making non-trivial reconstruction as soon as theoretically possible. However, as opposed to [14], in our article we study much more in depth the spectrum of the matrix B, consequently proposing a dynamical Bethe-Hessian that, compared to B, is smaller in size and symmetric, hence allows for a faster computation of the relevant eigenvectors. The choice $\xi = \lambda_d$, introduces further advantages: i) the value of λ (needed to build B) does not need to be known; ii) as claimed in the article, Algorithm 1 is well defined for $k \geq 2$ classes of *arbitrary size* (unlike [14]). These aspects suggest that Algorithm 1 should be preferred over the spectral algorithm of [14]. In agreement with the points raised by reviewers 2 and 4, given the growing interest in dynamical networks and the vast use of spectral clustering, our submission has the potential to strongly impact the field of dynamical community detection.

Simulation of [14] (R1). After a personal communication with the authors of [14], it appears that the algorithm deployed in [14] to obtain Figure 3 uses a regularization of B to suppress the *uninformative* modes and locate the informative eigenvalues in dominant position. This regularization step unfortunately is not mentioned in the published version of [14] and the authors are currently at work to clarify this point. We eventually learned of the existence of an unpublished version of [14] (arXiv:1506.06179) in which the authors mention the need of a regularization and show in Figure 2 that the spectrum of B for $k = 2$ can indeed have more than two isolated eigenvalues. In Figure 3 we show the performance of B when using the second largest eigenvector, as prescribed in [14]. We further underline that the results of our Appendix C imply that the matrix B can make non-trivial partitions for $\alpha > \alpha_c(T, \eta)$, in agreement with the conjectures of [14]. To make this point clearer, we will move the simulation of *opt. dyn B* of Figure 10 to Figure 3.

Comprehensiveness of the simulation on synthetic data (R1, R2). Our theoretical case study was limited to two classes of equal size, typical setting used in the literature (see e.g. [18,19]) for which the transition is well defined. For consistency, we chose to show only the most significant case of $k = 2$ in the main article and explore the behavior of the algorithm for different values of η, k, Φ in the Appendix. The reviewers suggested to further simulate the behavior of the algorithm for larger values of T and for classes of different sizes. Algorithm 1 is well defined also in these cases and relevant simulations will be added to the Appendix showing that: i) the performance increases as T increases; ii) Algorithm 1 can reconstruct communities of very different sizes (ratio 9/1).

Applicability and analysis on real data-sets. (R3, R4) For any given temporal graph and a value of η, Algorithm 1 is well defined. Other notable works in the static regime [18,19,28] showed that the non-backtracking and Bethe-Hessian matrices perform well also for graphs not generated from the DC-SBM and we expect the same to hold for Algorithm 1. If η is not known, as we indicate in lines 219-225, it can be selected through cross validation as the value leading to the largest number of classes detected. This aspect will be clarified in the updated version and it makes Algorithm 1 applicable to any temporal graph. As suggested, we tested its performance on some real datasets taken from the SocioPattern webpage (*e.g.* Primary, LyonSchool) observing that: i) for the same k it outperforms [26] in terms of modularity; ii) for similar modularity values, it outputs an estimate of k much closer to the ground truth than [41].

Explicit derivation of $\alpha_c(T, \eta)$ (R1). Appendix A consists of an explicit re-derivation of the results of [14], generalized to the case of $\Phi \neq 1$. This is clearly stated in the Appendix, but it certainly deserves to be properly addressed in the main text as well. A clear comment will be added in the new version.

Definition of ω_{ij} for the temporal edges (R1). As per Definition 1 and lines 151-153, the non-edges do not take part in the definition of $B_{\xi,h}$ and the weight $\omega_{ij} = h$ effectively only concerns the time edges. We will clarify this definition.

Regime of sparsity considered (R2). The regime of sparsity in which we are able to operate is an implicit consequence of the detectability condition. In fact, Algorithm 1 can be adopted for all $c_{\rm in}, c_{\rm out} = O_n(1)$ that satisfy $\alpha > \alpha_c(T, \eta)$ and that guarantee the existence of giant component (as stated in the footnote 3). We will make this aspect more explicit.

Figure 3 (R2). The overlap is defined in footnote 7. For clarity the definition might be moved to the main text. The overlap expresses how much a label assignment performs better than random guess, relatively to the planted assignment. It ranges between 0 (random assignment) to 1 (perfect assignment). We chose this measure that is adopted in many other works in the field, *e.g.* [14,18,19,28]. The left plot is a representation of the overlap (heatmapped) against α and η, while the right plot is a horizontal cut of the left plot at $\eta = 0.7$ for different techniques. For both plots, $k = 2$.

Extension to non-symmetric matrices (R2). The approach based on the Bethe-Hessian matrix can be easily generalized to weighted and directed graphs, as initially suggested in [19]. The use of a different generative model (giving a non symmetric matrix), however, would affect most of the analytical results obtained and the crucial aspect of the detectability threshold. The extension of the concept of detectability threshold and of Proposition 1 would need a dedicated analysis, but we do not predict any major complication in this extension.