We thank all the reviewers for their high-quality and constructive feedback! We hope that we address all the concerns below in a satisfactory manner.

Reply to Reviewer 2. • Regarding the normalization with $(k-1)$ in Eq. (1), we do not assume that inter-group and intra-group edge densities are equal. Instead, we motivate our normalization by the following reasoning: Suppose that group sizes and inter-group and intra-group edge densities stay fixed (but not necessarily equal) as k increases. Since the number of inter-group edges grows quadratically with k and the number of intra-group edges grows linearly with k, if Eq. (1) is unweighted, the objective will be quickly dominated by the number of inter-group edges. We would also like to point out that, unlike clusters in unsigned networks where the intra-group edge density is usually larger than the inter-group edge density, this is not the case for conflicting groups in signed networks. This is indeed the case for the solutions found in all our problem instances, except one (SCG-B with $k = 6$). • Thank you for mentioning scalability. We followed the approach described in the WWW2020 paper to augment `WikiCon` to 1.1 M nodes and 32.7 M edges, while preserving the ratio of negative edges, and run SCG to detect $k = 6$ conflicting groups on the same machine we reported in our submission. SCG-MA, SCG-MO, SCG-R complete in 1.7 h, 1.1 h, and 2.3 h, respectively, and SCG-B fails to complete in 1 day. We will include additional scalability experiments in the next version of our paper. • Regarding evaluation on datasets with ground truth, unfortunately, we are not aware of such data. • Finally, we will balance our references and consider citing the CIKM2017 paper you pointed out.

Reply to Reviewer 3. • Thank you for the suggestion to analyze the recovery condition in m-SSBM and sparse regime, and providing detailed examples and references! We will present this interesting direction in the future-work section. • Q: *Is there any particular reason why the proposed approach does not fail by returning a large component as a cluster/group?* A: We first remind our objective in Eq. (6), which, after ignoring the weighting between inter/intra-group edges, can be expressed as $(\#\{\text{edges satisfying Property 1}\} - \#\{\text{edges violating Property 1}\})$ divided by $\#\{\text{nodes in all conflicting groups}\}$. Intuitively we are seeking for small-size conflicting groups with many "consistent" edges. For simplicity, consider m-SSBM with $\eta = 0$ and let $\{S_j^*\}_{j=1}^k$ be the ground-truth groups. Adding any additional node to a group S_h^* will only decrease the objective score. • Finally, for polarity-score error bars and a comparison to the ground-truth groups, please see below Figures (a) and (b).

Reply to Reviewer 4. • Q: *Can experiments be done to analyze the mean ratio of the number of negative edges in each group and the number of positive edges between groups after search algorithm?* A: Assuming that the suggested measure is the following

$$\texttt{MeanDisagreementRatio}(S_1, \cdots, S_k) = \frac{1}{k}\sum_{h=1}^{k} \underbrace{\frac{|\{(i,j) \in E_- \cap E(S_h)\}| + |\{(i,j) \in E_+ \cap E(S_h, \cup_{\ell \neq h} S_\ell)\}|}{|E(S_h, \cup_{\ell=1}^{k} S_\ell)|}}_{\texttt{DisagreementRatio}\text{ of group } S_h},$$

we have computed the measure in our data, and the results are shown below in Figure (c). The error bars show the variance of the `MeanDisagreementRatio` measure on graphs generated by m-SSBM. We observe that KOCG-top-1 has almost always the lowest `MeanDisagreementRatio` score, even lower than the score of the ground-truth solution (in aqua blue). This result suggests that the `MeanDisagreementRatio` measure can be easily hacked by methods like KOCG-top-1, which find solutions of very small size. The reason that `MeanDisagreementRatio` is not a reliable measure, in our opinion, is that it can favor very small/sparse solutions. For example, for $k = 2$, a solution consisting of a single negative edge would be optimal. More generally, groups with purely positive intra-group and/or negative inter-group edges would be optimal for any k, regardless of their size or density. On the contrary, our Polarity score cannot be easily hacked by small groups with few edges, since it is unlikely for groups with few edges to have large score. Hence, thank you for this suggestion, but we believe that our proposed objective is preferable. • Regarding some formulations are hard to follow without referencing the Supplementary, we guess you mean Section 6.1. If the paper gets accepted, there will be an additional page and we will expand the discussion and try to make the section self-contained.