We sincerely thank all of you for the detailed, thoughtful, and constructive comments and feedback. We have incorporated all your suggestions in our paper, which has significantly improved from them. We elaborate below.

Reviewer #2: (I) Our algorithm can handle > 2 *protected groups*: in our numerical results, there are up to five protected (racial) groups. It can also handle > 2 *protected attributes* (e.g., race, age, gender) by either: a) partitioning the network based on joint values of the protected attributes, and imposing max-min fairness constraint for each group; e.g., constraints on (White, Young, Female) people, etc.; or b) imposing a max-min fairness constraint for each protected attribute, separately. **(II)** We added a table of racial composition data for all networks. For instance, the `MFP` networks consisted of 16.5% White, 35% Black, 21% Hispanic, 18.5% Mixed, and 9% Others. Each individual belongs to a single race level. **(III)** The computational complexity of the problem increases exponentially with K, limiting us to increase K beyond 3 for the considered instances. As demonstrated by our results, $K \sim 3$ was sufficient to considerably improve fairness of the covering at moderate cost. **(IV)-(V)** Covering schemes are not inputs but rather *decision variables* of the K-adaptability problem. The optimization problem will identify the best K covering schemes that satisfy all the constraints including fairness constraints. **(VI)** In Section 5, we vary W from 0 to 1, in increments of 0.04; we employ the largest W for which the problem is feasible (see lines 152-154). By construction, this choice of W guarantees that all of the fairness constraints are satisfied. The choice of W varies with the network structure, no. of monitors, no. of failed nodes and K. In Table 2, for network `MFP2`, and for $J = 1, \ldots, 5$, W was: 0.64, 0.56, 0.48, 0.4 and 0.32, respectively. We will report the values of W in a table in the appendix. **(V)** We clarify Figs. 2(b)-(d) by changing the y-axis label to "Average normalized objective value" and adding to the caption that "it corresponds to the ratio of objective value of the master problem (appendix, line 582) to the network size, averaged over the five network instances."

Reviewer #3 (I)-(II) We incorporated all the recommendations. We included proof sketches, in the main text, after Props. 1 and 2 and Th. 1. We improve clarity of Th. 1 by adding "In this formulation, there are two sets of variables: a) The decision variables of the original problem; b) Dual variables emerging from employing linear programming duality to reformulate the inner minimization problem in Problem (4) ". We explain the role of the dual variables, and the two sets of constraints corresponding to different values of the parameter $\boldsymbol{l}$. **(III)** The memory overflow is due to the fact that the MILP formulation in Th. 1, although polynomial in all problem inputs, remains exponential in K. This is the main motivation to develop the Bender's decomposition approach in Section 4. Please see also response (III) to Reviewer #2. **(IV)** We will provide a head-to-head comparison with Table 1. For instance, the corresponding results of our approach ($K = 3$) for `MFP2` network are: White: 56%; Black: 80%; Hispanic: 70%; Mixed: 71%; Other: 72%. **(V)** We improved the K-adaptability formularization by adding to Section 4: "the MILP reformulation relies on three key components: a) partitioning of the uncertainty set (achieved by introducing the parameter $\boldsymbol{l}$), b) continuous relaxation of each subset of the uncertainty set , and c) linear programming duality theory, to reformulate the robust optimization formulation over each subset." **(VI)** We will release the code and a "readme" file with instructions, detailing the sequence of the runs. **(VII)** The 2-hour time limit is justified by the "flattening" in the "Objective Bound" (Figures 2(b)-(d)); this is a common approach in optimization to terminate the algorithm when the change in objective is small. **(VIII)** We apologize for the confusion caused by Line 281. We now write "…by imposing fairness constraints for each group. We set the number of monitors to I=N/3." Please see also our answer (VI) to Reviewer #2. **(IX)** We now add a section on future work. **(X)** The Bernoulli distribution of the random variables Y_n and Z_{ni} is due to the Erdős-Rényi network generation process (see lines 418-419). Therefore, the probability of Z_{ni} (similarly Y_n) taking the value of 1 is a known constant. **(XI)** The "budget regime" refers to the assumptions on the values of I, which we made more explicit. **(XII)-(XIII)** The remaining comments were addressed; we also added a part that was inadvertently deleted in the proof of Prop. 2.

Reviewer #4 (I) Please see answer (I) to Reviewer #2. **(II)** The paper [31] does not handle the uncertainty in node availability, which is one of the main contributions of our framework. **(III)** We have added the discussion of the worst-case PoF (Lemma 2) to the main text. **(IV)** We clarified, in the text, that we investigate the ratio of expected coverage rather than expectation of ratios for analytical tractability. **(V)** The assumption on I can be interpreted as a "small budget assumption" that helps simplify the evaluation of the coverage. Please also see answer (II) to reviewer #3. **(VI)** Intuitively, uncertainty sets involving constraints as lower bounds on the (sums of) uncertain parameters satisfy the upward-closeness property. We now provide three examples of such sets, that are of practical relevance.**(VII)** The value of K determines the approximation quality, enabling the decision-maker to trade-off the optimality with computation time. The choice of K is mainly guided by the available computational resources (e.g., time) and is domain specific. Particularly, in low-resource settings (e.g., suicide prevention for homeless youth), we may be restricted to use low values of K. **(VIII)** Please refer to answer (II) to Reviewer #3. **(IX)** We have incorporated all your comments to improve the interpretability of Table 2. **(X)** Bender's decomposition is an *exact* iterative algorithm that converges to an optimal solution provided subproblems are LPs as in our case (Bertsimas, Dimitris, John N. Tsitsiklis. Intro. to linear optimization, 1994). In practice, it is run until a termination criterion, such as time, optimality gap, etc. is satisfied. We chose time limit for practical purposes. **(XI)** From discussion with our social work partners, $I \in [20, 30]\%N$ is typically seen in the context of suicide prevention. We now have added more rows in Table 2, reporting the average coverage improvement and average PoF for different values of I. For instance, for $I = 20\%N$, the average "Improvement in Min. Fraction Covered" for $J = 0, \ldots, 5$ is 17.2%, 13.8%, 14.0%, 10.0%, 9.0% and 6.7%, respectively. The "PoF" values are all less than 4%. The value of γ can be inferred from J ($\gamma = J/I$), we replace "Size" with N in Table 2 for clarity.