Supporting Information

Zhou et al. 10.1073/pnas.1000488107

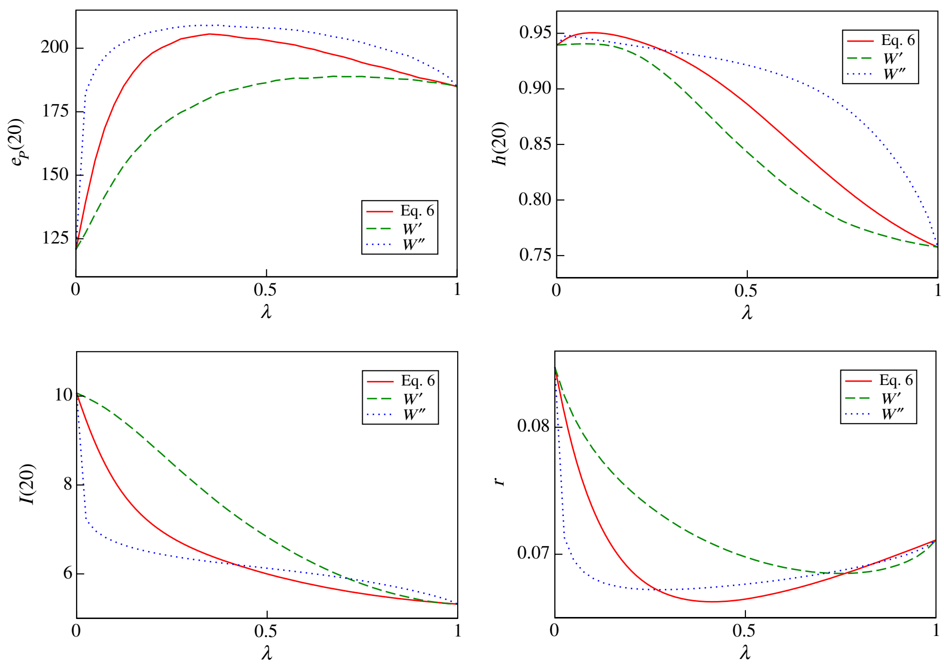

Fig. S1. Elegant hybrids of the HeatS and ProbS algorithms can be created in several ways besides that given in Eq. **6** of the paper. For example $W'_{\alpha\beta} = \left(\frac{1-\lambda}{k_\alpha} + \frac{\lambda}{k_\beta}\right) \sum_{j=1}^{u} a_{\alpha j} a_{\beta j}/k_j$, or $W''_{\alpha\beta} = \frac{1}{(1-\lambda)k_\alpha + \lambda k_\beta} \sum_{j=1}^{u} a_{\alpha j} a_{\beta j}/k_j$. While $W'_{\alpha\beta}$ performs well only with respect to $I(20)$, Eq. **6** and $W''_{\alpha\beta}$ both have their advantages. However, Eq. **6** is somewhat easier to tune to different requirements since it varies more slowly and smoothly with λ. The results shown here are for the RateYourMusic dataset.

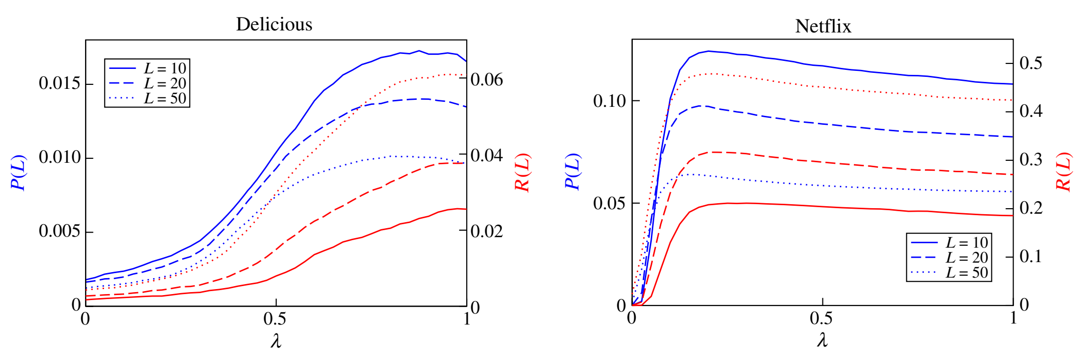

Fig. S2. Precision $P(L)$ and recall $R(L)$ provide complementary but contrasting measures of accuracy: the former considers what proportion of selected objects (in our case, objects in the top L places of the recommendation list) are relevant, the latter measures what proportion of relevant objects (deleted links) are selected. Consequently, recall (red) grows with L, whereas precision (blue) decreases. Here we compare precision and recall for the HeatS + ProbS hybrid algorithm on the Delicious and Netflix datasets. While quantitatively different, the qualitative performance is very similar for both measures.

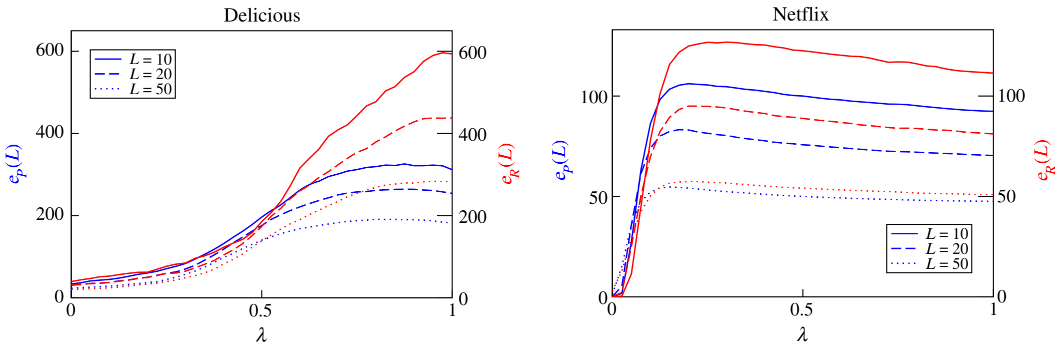

Fig. S3. A more elegant comparison can be obtained by considering precision and recall *enhancement*, that is, their values relative to that of randomly sorted recommendations: $e_P(L) = P(L) \cdot ou/D$ and $e_R(L) = R(L) \cdot o/L$ (Eqs. **7***a*, *b* in the paper). Again, qualitative performance is close, and both of these measures decrease with increasing L, reflecting the inherent difficulty of improving either measure given a long recommendation list.

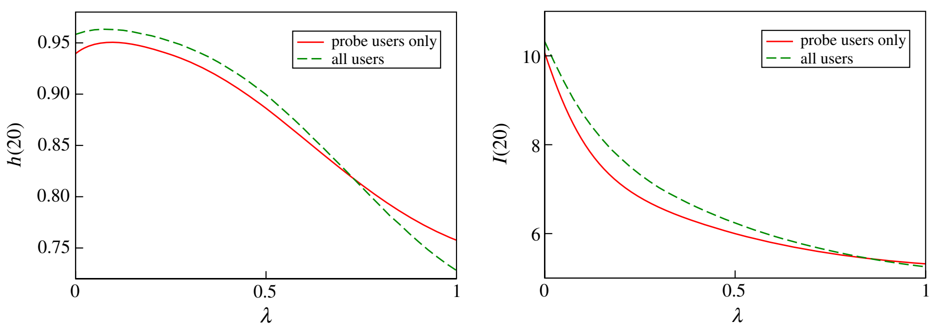

Fig. S4. Comparison of the diversity-related metrics $h(20)$ and $I(20)$ when two different averaging procedures are used: averaging only over users with at least one deleted link (as displayed in the paper) and averaging over all users. The different procedures do not alter the results qualitatively and make little quantitative difference. The results shown are for the RateYourMusic dataset.

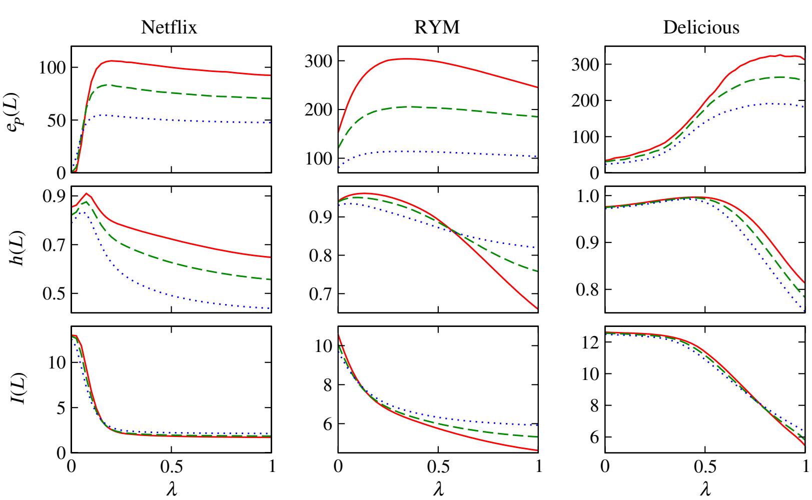

Fig. S5. Comparison of performance metrics for different lengths L of recommendation lists: $L = 10$ (red), $L = 20$ (green), and $L = 50$ (blue). Strong quantitative differences are observed for precision enhancement $e_P(L)$ and personalization $h(L)$, but their qualitative behavior remains unchanged. Much smaller differences are observed for surprisal $I(L)$.

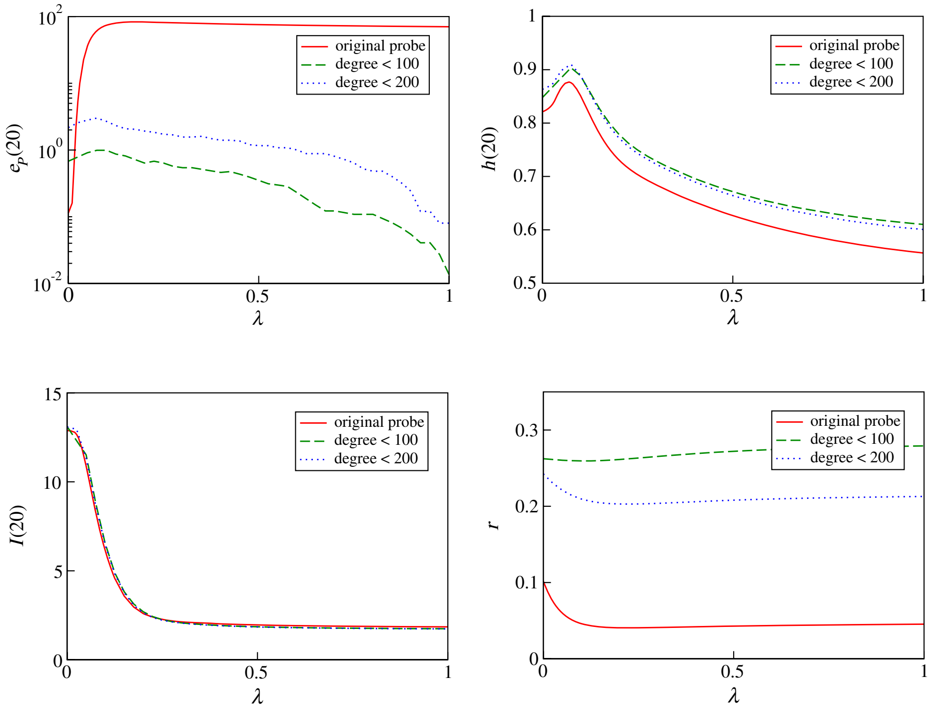

Fig. S6. Our accuracy-based metrics all measure in one way or another the recovery of links deleted from the dataset. Purely random deletion will inevitably favor high-degree (popular) objects, with their greater proportion of links, and consequently methods that favor popular items will appear to provide higher accuracy. To study this effect, we created two special probe sets consisting of links only to objects whose degree was less than some threshold (either 100 or 200): links to these objects were deleted with probability 0.5, while links to higher-degree objects were left untouched. The result is a general decrease in accuracy for all algorithms—unsurprisingly, since rarer links are inherently harder to recover—but also a reversal of performance, with the low-degree-favoring HeatS now providing much higher accuracy than the high-degree-oriented ProbS, USim, and GRank. The results shown here are for the Netflix dataset.

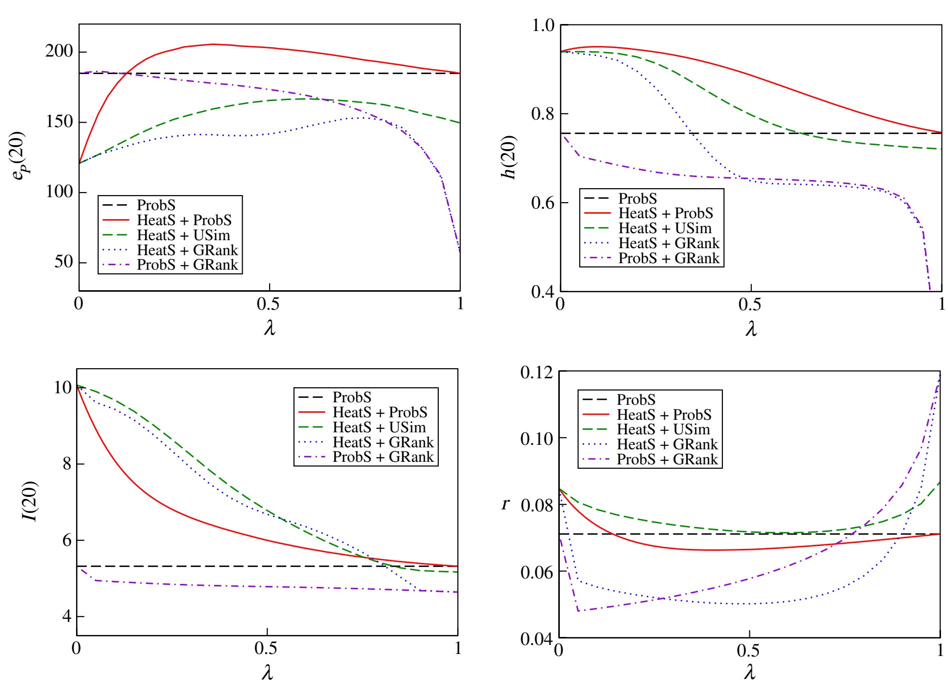

Fig. S7. In addition to HeatS + ProbS, various other hybrids were created and tested using the method of Eq. **5** in the paper, where for hybrid X + Y, $\lambda = 0$ corresponds to pure X and $\lambda = 1$ pure Y. The results shown here are for the Netflix dataset. The HeatS + USim hybrid offers similar but weaker performance compared to HeatS + ProbS; combinations of GRank with other methods produce significant improvements in r, the recovery of deleted links, but show little or no improvement of precision enhancement $e_P(L)$ and poor results in diversity-related metrics. We can conclude that the proposed HeatS + ProbS hybrid is not only computationally convenient but also performs better than combinations of the other methods studied.