



**Fig. 5. Ensemble clustering overview.** Finishing techniques were applied to random toy data (see file S1 for analysis details). (A) Set of clustering results obtained using the  $k$ -means algorithm with various values of  $k$  (a  $k$ -sweep). (B) Hierarchically clustered (Ward linkage) co-occurrence matrix for the ensemble of results in (A). The heatmap represents the percentage of times any pair of data points coclusters across the ensemble. (C) A majority vote analysis was applied (left panel) using a threshold of 50% on the co-occurrence matrix in (B). Six clusters (see dendrogram color groupings) result from the majority vote (right panel). (D) Application of fuzzy clustering to the ensemble. The left panel shows the details of the co-occurrence matrix for the blue, gray, and orange clusters, and the right shows the clustering assignments. The gray cluster provides an example of partially fuzzy clustering because it shares membership with the orange and dark blue clusters.