Figure 1. Cluster distribution and cluster purity resulting from employing the icluster integrative clustering method on synthetic data.

Figure 1 above presents the cluster distribution by employing the *iCluster*[1–3] integrative methodology on a subset of the synthetic data. The technique was implemented by utilizing the R package *iClusterPlus*[4] (*lambda penalty = 0.2 for both datasets, maximum number of iterations allowed = 150 and number of subtypes =4*). A subset of 50 relationships of each type was chosen from the initial 6000 since methodology with the aforementioned constraints was unable to complete clustering for the entire dataset (number of observations = 6000 and number of features = 100). As visualized above, the clustering purity is significantly lower than the clustering achieved by the scagnostics workflow, which in addition to correctly classifying all relationships in its own cluster, also reduces the dimensionality to ensure applicability of all standard unsupervised clustering techniques.

Reference:
1.　Mo Q, Wang S, Seshan VE, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci*. 2013;110(11):4245-4250. doi:10.1073/pnas.1208949110
2.　Shen R, Olshen AB, Ladanyi M. (iCluster)Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*. 2009;25(22):2906-2912. doi:10.1093/bioinformatics/btp543
3.　Shen R, Mo Q, Schultz N, et al. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One*. 2012;7(4). doi:10.1371/journal.pone.0035236
4.　Mo Q, Shen R. iClusterPlus: Integrative clustering of multi-type genomic data. 2018.