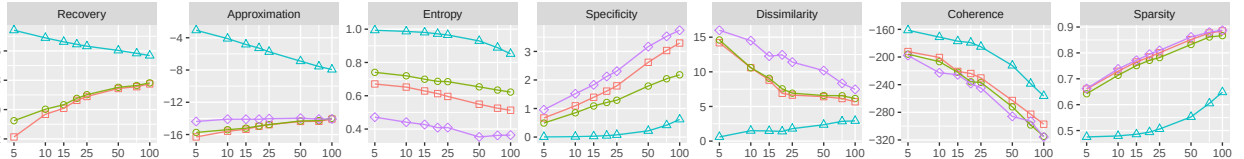
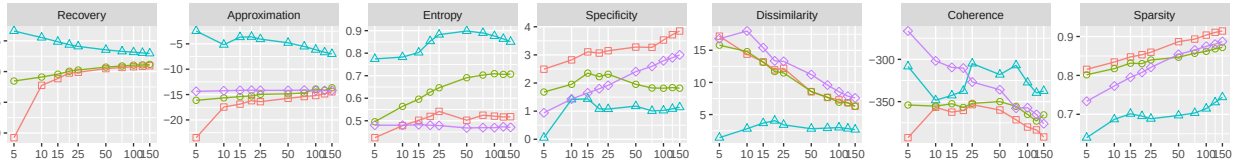
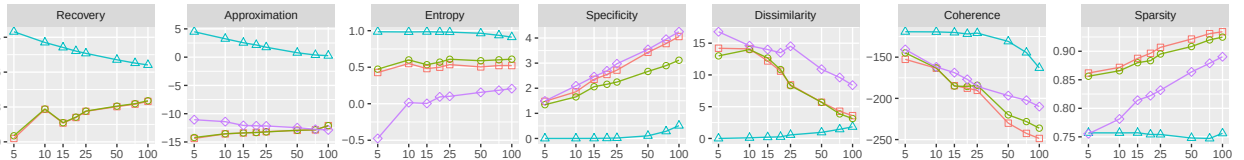
NIPS (# documents M=1,348 / # vocabulary N=5k / average document length=380.5)



NYTimes (# documents M=263,325 / # vocabulary N=15k / average document length=204.9)



Movies (# documents M=63,041 / # vocabulary N=10k / average document Length=142.8)



Songs (# documents M=14,653 / # vocabulary N=10k / average document length=119.2)

