Data set	N	$|\mathcal{S}|$	D	Train	Valid	Test	$N \times V/D$
NLTCS	16	13,733	1,716	16,181	2,157	3,236	150.871
MSNBC	17	54,839	24,452	291,326	38,843	58,265	202.541
KDD 2k	64	48,279	14,292	180,092	19,907	34,955	806.457
Plants	69	132,959	58,853	17,412	2,321	3,482	20.414
Audio	100	739,525	196,103	15,000	2,000	3,000	7.649
Jester	100	314,013	180,750	9,000	1,000	4,116	4.979
Netflix	100	161,655	51,601	15,000	2,000	3,000	29.069
Accidents	111	204,501	74,804	12,758	1,700	2,551	18.931
Retail	135	56,931	22,113	22,041	2,938	4,408	134.560
Pumsb-star	163	140,339	63,173	12,262	1,635	2,452	31.638
DNA	180	108,021	52,121	1,600	400	1,186	5.526
Kosarak	190	203,321	53,204	33,375	4,450	6,675	119.187
MSWeb	294	68,853	20,346	29,441	3,270	5,000	425.423
Book	500	190,625	41,122	8,700	1,159	1,739	105.783
EachMovie	500	522,753	188,387	4,524	1,002	591	12.007
WebKB	839	1,439,751	879,893	2,803	558	838	2.673
Reuters-52	889	2,210,325	1,453,390	6,532	1,028	1,540	3.995
20 Newsgrp	910	14,561,965	8,295,407	11,293	3,764	3,764	1.239
BBC	1058	1,879,921	1,222,536	1,670	225	330	1.445
Ad	1556	4,133,421	1,380,676	2,461	327	491	2.774