

Supplement materials

1. Characteristics of networks used in the experiments

Below characteristics of all used in the experiments datasets are presented in a form of boxplots. For each class within a given network such metrics as: indegree centrality, outdegree centrality, betweenness centrality,

page rank, clustering coefficient, hub centrality, and authority are considered.

Each boxplots shows: on each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

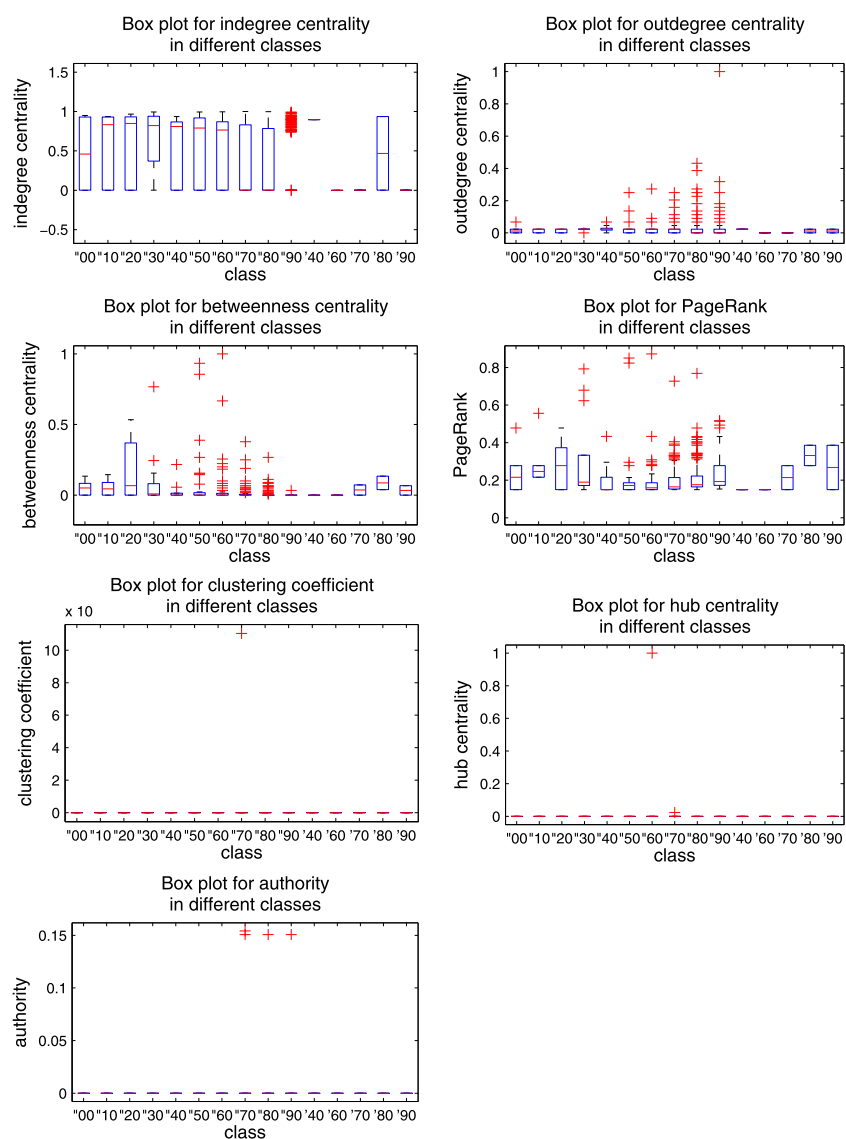


Fig. 14. Characteristics of CSPhd network.

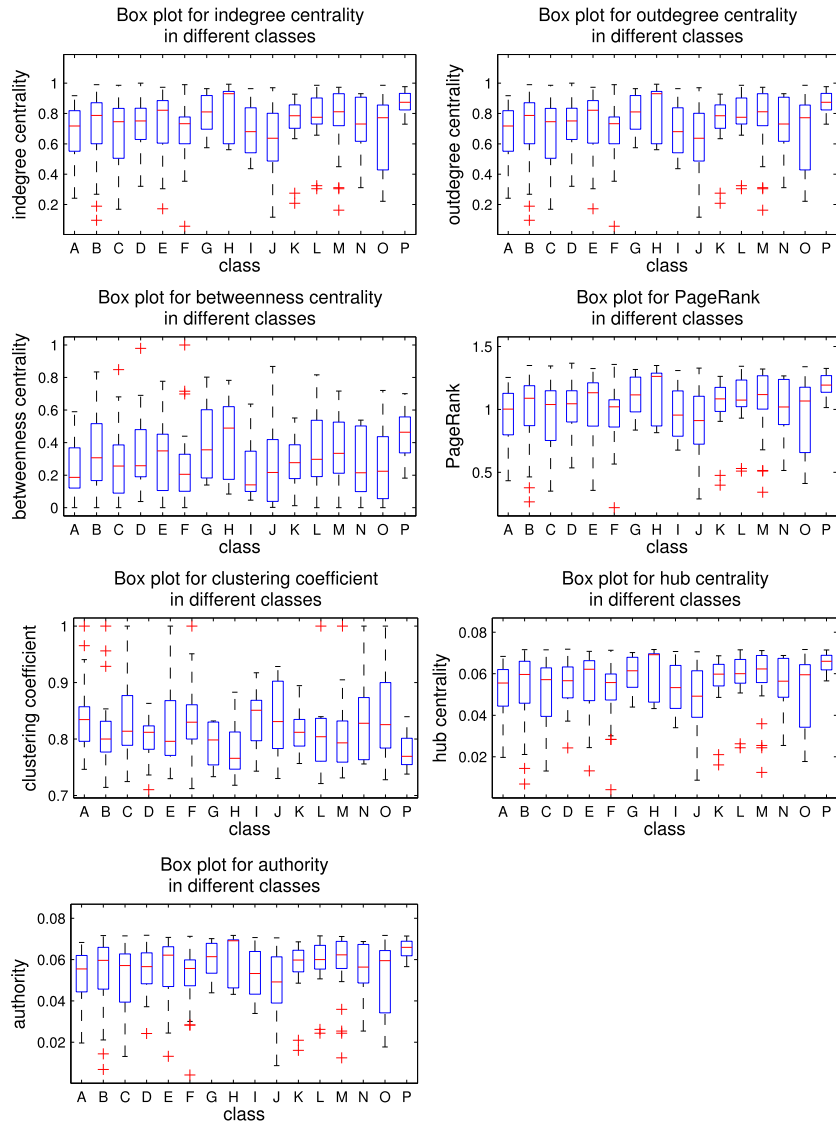


Fig. 15. Characteristics of AMD network.

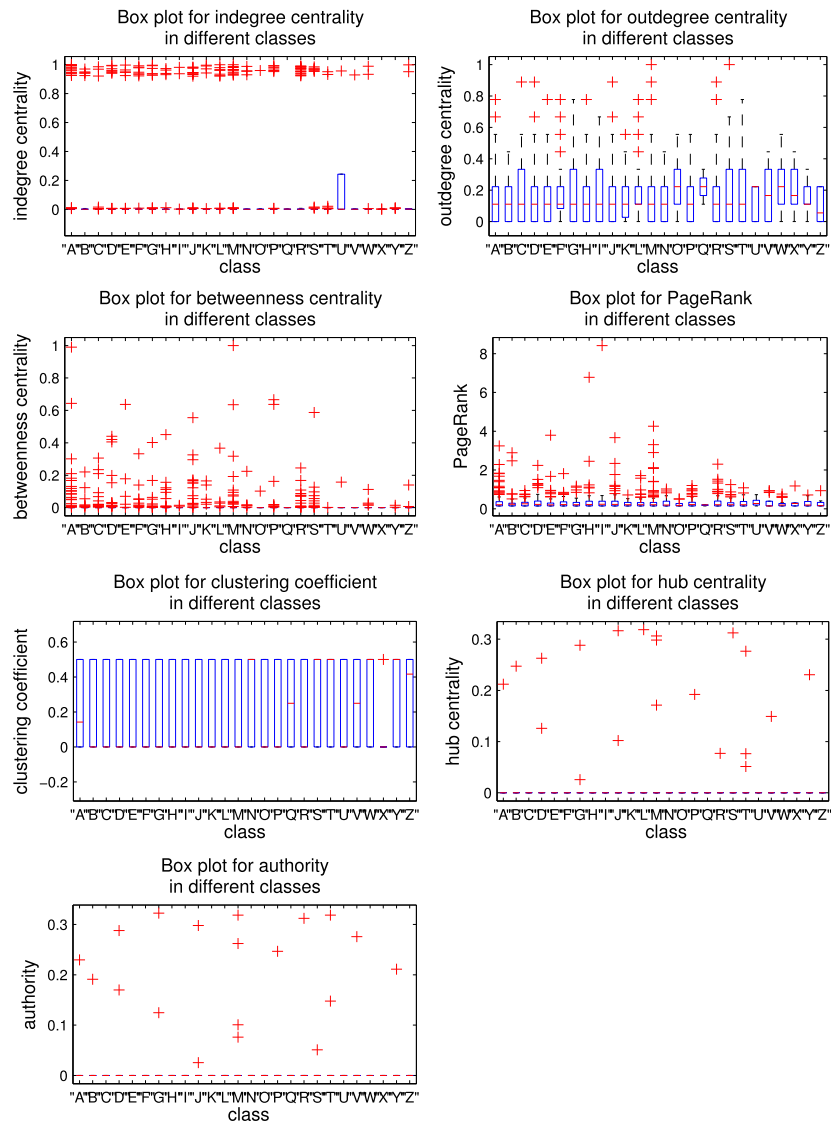


Fig. 16. Characteristics of Net Science network.

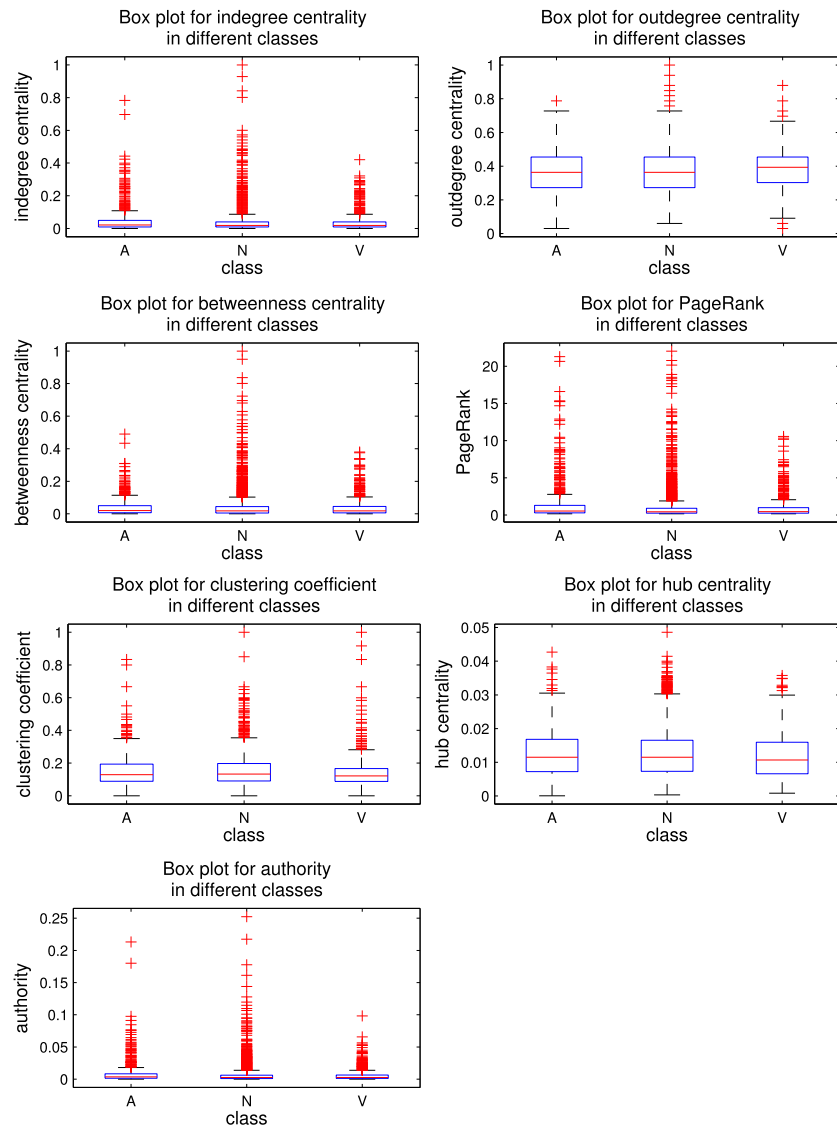


Fig. 17. Characteristics of Pairs FSG network.

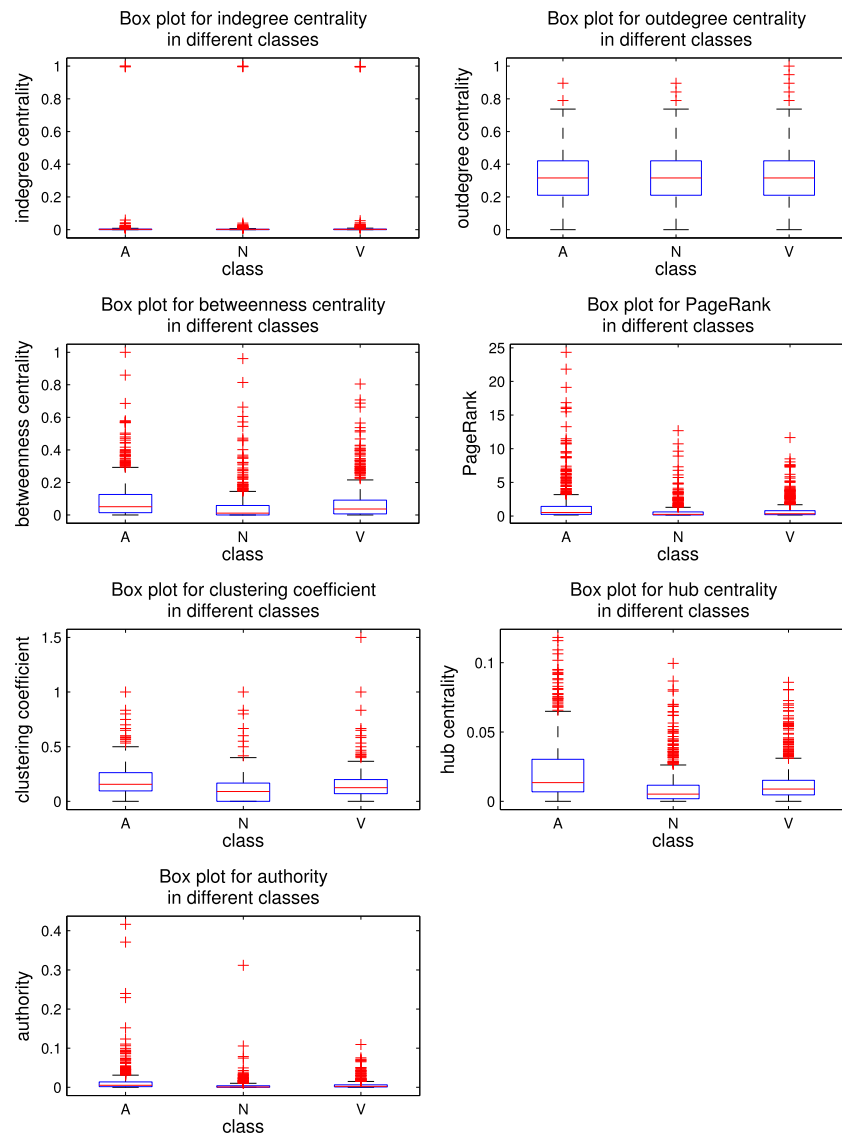


Fig. 18. Characteristics of Pairs FSG small network.

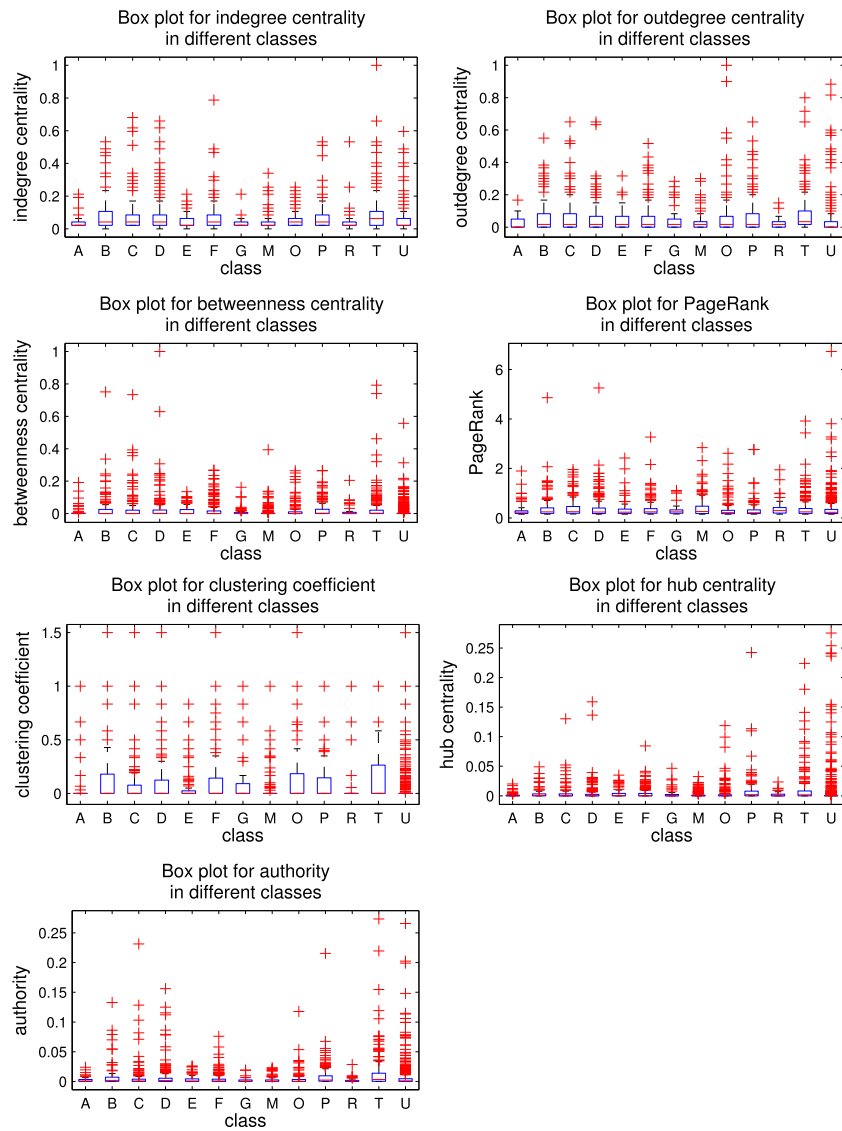


Fig. 19. Characteristics of yeast network.

2. Distribution of classes in analysed networks

Below distribution of classes within each analysed network is presented.

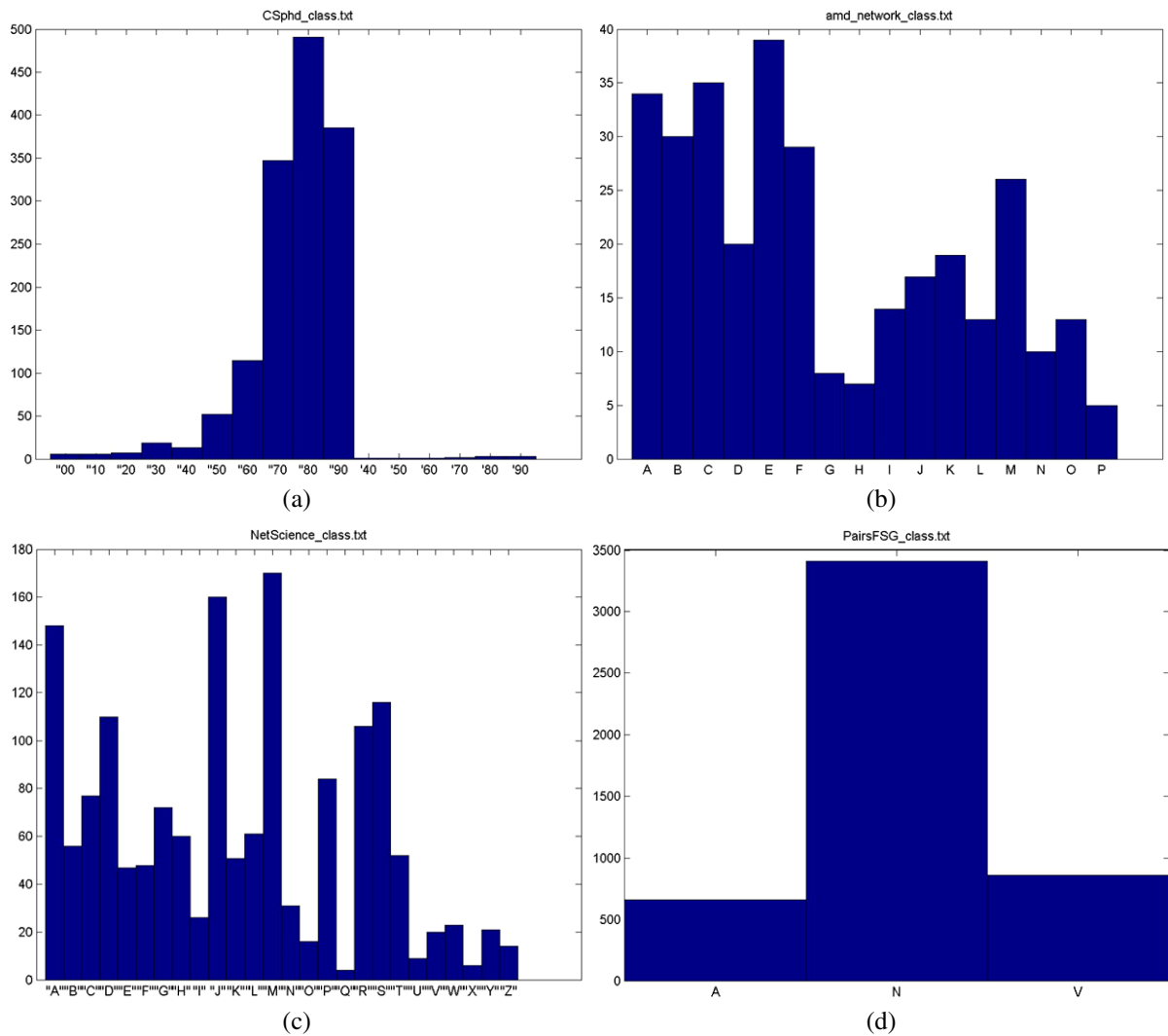


Fig. 20. Histograms of classes for all evaluated networks. (a) CSpHd network; (b) AMD network; (c) NetScience network; (d) PAIRS_FSG network; (e) PAIRS_FSG_small network; (f) YEAST network.

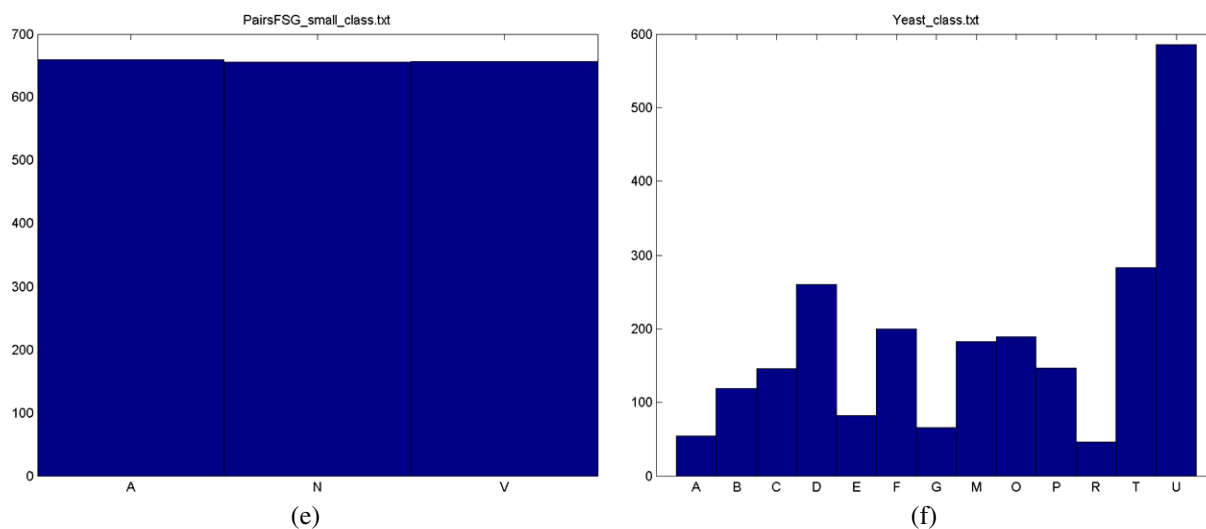


Fig. 20. (Continued.)

3. Representativeness of sampled data

The representativeness of a data sample is assessed using Kullback–Leibler divergence (a.k.a. relative entropy) which is a measure of the difference between two probability distributions. It measures how much information is lost when one probability distribution (in our case it is a distribution of-

classes in a given sample – 10%, ..., 90% of the whole dataset) is used to approximate another one (in this paper it is the probability distribution of classes in the whole dataset). The smaller the divergence the smaller loss; 0 means that no information is lost.

Below the Kullback–Leibler divergence for each analysed network is presented.

Network measure used for ranking nodes	Method of ranking creation	% of the whole dataset sampled								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
indegree	Descending	0.172	0.096	0.092	0.074	0.047	0.028	0.014	0.009	0.005
outdegree		0.057	0.033	0.020	0.016	0.015	0.017	0.013	0.004	0.002
betweenness		0.036	0.019	0.016	0.013	0.009	0.007	0.007	0.006	0.005
clustering coefficient		0.040	0.022	0.026	0.039	0.025	0.021	0.018	0.011	0.006
hub		0.054	0.037	0.020	0.013	0.016	0.008	0.006	0.008	0.005
authority		0.039	0.023	0.015	0.013	0.009	0.008	0.008	0.006	0.004
PageRank		0.514	0.354	0.140	0.105	0.071	0.045	0.027	0.012	0.007
indegree	Ascending	0.039	0.040	0.025	0.012	0.008	0.009	0.008	0.012	0.007
outdegree		0.053	0.020	0.019	0.010	0.007	0.007	0.007	0.004	0.003
betweenness		0.071	0.063	0.071	0.077	0.048	0.030	0.023	0.011	0.007
clustering coefficient		0.072	0.063	0.070	0.077	0.048	0.030	0.023	0.011	0.007
hub		0.056	0.028	0.020	0.017	0.010	0.007	0.008	0.006	0.006
authority		0.071	0.064	0.070	0.077	0.048	0.030	0.023	0.011	0.007
PageRank		0.071	0.061	0.069	0.076	0.047	0.029	0.022	0.011	0.007
indegree	Random	0.056	0.023	0.027	0.021	0.011	0.008	0.006	0.005	0.004
outdegree		0.056	0.034	0.041	0.058	0.043	0.031	0.022	0.011	0.007
betweenness		0.072	0.065	0.069	0.073	0.048	0.033	0.021	0.013	0.007
clustering coefficient		0.065	0.020	0.011	0.010	0.008	0.007	0.006	0.005	0.003
hub		0.114	0.067	0.059	0.043	0.036	0.031	0.027	0.017	0.007
authority		0.584	0.286	0.161	0.072	0.045	0.027	0.020	0.010	0.005
PageRank		0.041	0.026	0.013	0.015	0.012	0.012	0.007	0.005	0.004

Fig. 21. Kullback–Leibler divergence for CSPhd network.

Network measure used for ranking nodes	Method of ranking creation	% of the whole dataset sampled								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
indegree	Descending	0.338	0.135	0.073	0.046	0.026	0.019	0.011	0.005	0.001
outdegree		0.172	0.103	0.065	0.046	0.027	0.022	0.016	0.010	0.003
betweenness		0.254	0.153	0.044	0.039	0.017	0.017	0.006	0.003	0.003
clustering coefficient		0.338	0.135	0.073	0.046	0.026	0.019	0.011	0.005	0.001
hub		0.172	0.103	0.065	0.046	0.027	0.022	0.016	0.010	0.003
authority		0.220	0.146	0.083	0.055	0.024	0.017	0.014	0.006	0.001
PageRank		0.305	0.201	0.072	0.039	0.021	0.017	0.011	0.005	0.002
indegree	Ascending	0.223	0.100	0.065	0.036	0.023	0.019	0.013	0.012	0.003
outdegree		0.296	0.114	0.065	0.032	0.013	0.008	0.006	0.005	0.003
betweenness		0.298	0.152	0.103	0.050	0.022	0.024	0.016	0.013	0.003
clustering coefficient		0.322	0.184	0.084	0.045	0.021	0.022	0.015	0.007	0.002
hub		0.241	0.074	0.055	0.019	0.014	0.018	0.011	0.007	0.003
authority		0.375	0.135	0.065	0.042	0.024	0.019	0.009	0.003	0.001
PageRank		0.172	0.097	0.058	0.051	0.026	0.018	0.013	0.009	0.005
indegree	Random	0.180	0.070	0.072	0.054	0.026	0.025	0.012	0.012	0.004
outdegree		0.375	0.135	0.065	0.042	0.024	0.019	0.009	0.003	0.001
betweenness		0.172	0.097	0.058	0.051	0.026	0.018	0.013	0.009	0.005
clustering coefficient		0.156	0.079	0.033	0.029	0.017	0.013	0.008	0.004	0.002
hub		0.318	0.135	0.067	0.044	0.028	0.019	0.011	0.005	0.001
authority		0.172	0.103	0.065	0.050	0.028	0.021	0.016	0.009	0.003
PageRank		0.262	0.037	0.041	0.023	0.008	0.010	0.008	0.002	0.001

Fig. 22. Kullback–Leibler divergence for AMD network.

Network measure used for ranking nodes	Method of ranking creation	% of the whole dataset sampled								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
indegree	Descending	0.056	0.040	0.023	0.011	0.012	0.009	0.005	0.003	0.002
outdegree		0.077	0.023	0.024	0.014	0.006	0.006	0.003	0.002	0.001
betweenness		0.072	0.027	0.016	0.009	0.007	0.006	0.005	0.002	0.001
clustering coefficient		0.086	0.051	0.020	0.012	0.005	0.005	0.004	0.003	0.002
hub		0.083	0.035	0.019	0.018	0.008	0.004	0.003	0.004	0.002
authority		0.064	0.030	0.023	0.017	0.011	0.008	0.005	0.003	0.001
PageRank		0.092	0.048	0.015	0.010	0.009	0.004	0.004	0.002	0.001
indegree	Ascending	0.083	0.019	0.015	0.011	0.007	0.005	0.004	0.002	0.002
outdegree		0.080	0.045	0.017	0.015	0.011	0.007	0.005	0.003	0.002
betweenness		0.062	0.021	0.016	0.012	0.008	0.007	0.005	0.003	0.001
clustering coefficient		0.043	0.028	0.021	0.013	0.006	0.005	0.004	0.002	0.001
hub		0.105	0.059	0.030	0.017	0.016	0.013	0.006	0.004	0.001
authority		0.056	0.035	0.010	0.008	0.010	0.005	0.004	0.002	0.001
PageRank		0.099	0.035	0.012	0.010	0.009	0.005	0.004	0.002	0.001
indegree	Random	0.104	0.044	0.020	0.017	0.009	0.004	0.003	0.003	0.001
outdegree		0.057	0.034	0.010	0.008	0.010	0.005	0.004	0.002	0.001
betweenness		0.099	0.035	0.012	0.010	0.009	0.005	0.004	0.002	0.001
clustering coefficient		0.072	0.033	0.017	0.012	0.012	0.007	0.003	0.002	0.002
hub		0.088	0.020	0.012	0.009	0.007	0.004	0.003	0.002	0.001
authority		0.066	0.050	0.015	0.010	0.007	0.005	0.003	0.002	0.001
PageRank		0.081	0.037	0.029	0.017	0.015	0.009	0.006	0.004	0.002

Fig. 23. Kullback–Leibler divergence for Net Science network.

Network measure used for ranking nodes	Method of ranking creation	% of the whole dataset sampled								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
indegree	Descending	0.014	0.005	0.004	0.003	0.002	0.001	0.000	0.000	0.000
outdegree		0.007	0.003	0.003	0.003	0.002	0.001	0.001	0.000	0.000
betweenness		0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
clustering coefficient		0.002	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
hub		0.000	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000
authority		0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
PageRank		0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
indegree	Ascending	0.017	0.006	0.001	0.001	0.000	0.000	0.000	0.000	0.000
outdegree		0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
betweenness		0.008	0.008	0.006	0.003	0.001	0.001	0.000	0.000	0.000
clustering coefficient		0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.000	0.000
hub		0.001	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000
authority		0.003	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000
PageRank		0.001	0.003	0.001	0.001	0.000	0.000	0.000	0.000	0.000
indegree	Random	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
outdegree		0.010	0.006	0.003	0.003	0.002	0.001	0.000	0.000	0.000
betweenness		0.008	0.004	0.003	0.002	0.003	0.002	0.001	0.000	0.000
clustering coefficient		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hub		0.014	0.009	0.004	0.002	0.001	0.001	0.000	0.000	0.000
authority		0.005	0.004	0.002	0.001	0.001	0.001	0.001	0.001	0.000
PageRank		0.006	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000

Fig. 24. Kullback–Leibler divergence for Pairs FSG network.

Network measure used for ranking nodes	Method of ranking creation	% of the whole dataset sampled								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
indegree	Descending	0.024	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000
outdegree		0.002	0.001	0.002	0.002	0.001	0.001	0.000	0.000	0.000
betweenness		0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000
clustering coefficient		0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hub		0.003	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000
authority		0.002	0.004	0.004	0.002	0.001	0.000	0.000	0.000	0.000
PageRank		0.099	0.048	0.040	0.031	0.025	0.017	0.015	0.009	0.001
indegree	Ascending	0.188	0.122	0.065	0.033	0.021	0.011	0.006	0.002	0.001
outdegree		0.003	0.002	0.003	0.002	0.000	0.000	0.000	0.000	0.000
betweenness		0.049	0.031	0.029	0.012	0.014	0.016	0.014	0.013	0.003
clustering coefficient		0.187	0.171	0.070	0.031	0.012	0.004	0.005	0.002	0.001
hub		0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
authority		0.143	0.105	0.068	0.053	0.037	0.026	0.021	0.011	0.004
PageRank		0.276	0.146	0.101	0.053	0.033	0.022	0.012	0.006	0.002
indegree	Random	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
outdegree		0.175	0.091	0.079	0.058	0.043	0.029	0.021	0.011	0.002
betweenness		0.195	0.140	0.104	0.062	0.041	0.025	0.014	0.005	0.002
clustering coefficient		0.009	0.002	0.003	0.002	0.001	0.001	0.000	0.000	0.000
hub		0.059	0.043	0.029	0.023	0.020	0.015	0.011	0.007	0.002
authority		0.176	0.088	0.051	0.030	0.018	0.009	0.005	0.002	0.000
PageRank		0.002	0.003	0.001	0.001	0.001	0.000	0.000	0.000	0.000

Fig. 25. Kullback–Leibler divergence for Pairs small FSG network.

Network measure used for ranking nodes	Method of ranking creation	% of the whole dataset sampled								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
indegree	Descending	0.127	0.088	0.065	0.040	0.018	0.013	0.014	0.007	0.001
outdegree		0.085	0.040	0.014	0.016	0.018	0.009	0.007	0.003	0.001
betweenness		0.019	0.009	0.004	0.002	0.003	0.003	0.003	0.001	0.000
clustering coefficient		0.118	0.081	0.050	0.040	0.032	0.017	0.013	0.007	0.001
hub		0.058	0.023	0.030	0.036	0.016	0.017	0.009	0.005	0.001
authority		0.012	0.010	0.006	0.004	0.002	0.002	0.001	0.001	0.000
PageRank		0.080	0.047	0.040	0.032	0.028	0.020	0.017	0.007	0.001
indegree	Ascending	0.073	0.033	0.015	0.021	0.023	0.014	0.007	0.003	0.001
outdegree		0.008	0.007	0.003	0.002	0.001	0.002	0.001	0.000	0.000
betweenness		0.054	0.073	0.056	0.037	0.039	0.029	0.012	0.005	0.001
clustering coefficient		0.075	0.036	0.017	0.008	0.007	0.008	0.010	0.004	0.001
hub		0.027	0.014	0.008	0.005	0.005	0.002	0.001	0.001	0.000
authority		0.067	0.064	0.057	0.039	0.027	0.020	0.016	0.006	0.001
PageRank		0.073	0.024	0.024	0.030	0.028	0.019	0.012	0.005	0.001
indegree	Random	0.018	0.013	0.004	0.001	0.001	0.001	0.002	0.001	0.000
outdegree		0.156	0.089	0.043	0.019	0.013	0.008	0.007	0.004	0.001
betweenness		0.092	0.062	0.039	0.020	0.014	0.009	0.009	0.006	0.002
clustering coefficient		0.018	0.005	0.006	0.006	0.004	0.002	0.001	0.001	0.000
hub		0.042	0.020	0.010	0.006	0.003	0.002	0.002	0.001	0.000
authority		0.035	0.015	0.009	0.005	0.003	0.003	0.002	0.001	0.001
PageRank		0.054	0.017	0.010	0.005	0.004	0.003	0.001	0.000	0.000

Fig. 26. Kullback–Leibler divergence for yeast network.