

School of Computing and Information Technology













• World-Wide Web can reveal more information than just the information contained in documents. For example, links pointing to a document indicate the popularity of the document, while links coming out of a document indicate the richness or perhaps the variety of topics covered in the document. This can be compared to bibliographical citations. When a paper is cited often, it ought to be important. The PageRank and CLEVER methods take advantage of this information conveyed by the links to find pertinent web pages. By means of counters, higher levels cumulate the number of artifacts subsumed by the concepts they hold. Counters of hyperlinks, in and out documents, retrace the structure of the web artifacts summarized.



























































\bigvee		On the quality of the clus (Paths generated as in Sh	tering ahabi et al)
	Error	rates	
n	Matrix-based	Medoid-based-crips	95% confidence
100	34%	25%	interval is $\pm 3.4\%$
500	54%	27%	
1000	n/a	27%	
1500	n/a	27%	
Pa in pa	aths are randomly a predefined gra aths that are not io	y generated around nucleu ph. Error rate measures the dentified as perturbations	as paths he percentage of of nucleus paths.
	N		

similarity n	neasure	es			
Dissimilarity	n	Harmonic		Crisp	
		Error	95% CI	Error	95%C
Usage	100	24%	2.2	27%	3.4
	500	24%	2.2	27%	3.4
Frequency	100	30%	2.9	33%	4.4
	500	30%	2.9	33%	4.4
Order	100	26%	2.7	29%	4.4
	500	26%	2.7	29%	4.4





Dissimilarity	n Harmonic		Crisp	
		Error	Error	
Usage	100	7%	11%	
	500	7.12%	7.35%	
Frequency	100	6.60%	10.20%	
	500	6.01%	7.20%	
Order	100	5.75%	8.75%	
	500	4.15%	6.34%	

Harmonic is more robust to random initialization, producing more consistent results across independent runs

© Vladimir Estivill-Castro



