1 appendix

1.1 Experimental Setup

The experimental setup for this research work comprises the hardware requirements of CPU @ 2.90 GHz with Intel Core i7-7500 CPU over 64-bit Operating System having 8.00 GB RAM. We use the software of *Python 3* with library modules of *networkx* for graphical analysis, *NLTK* for text processing, *pandas* to handle the data, *matplotlib* for graph plot, and many other relevant modules. We implement the baselines by using existing modules[1] which are further modified to incorporate the settings for EdgeGraph. We use the default parameter settings of random walk based GKET which are available in the existing implementation.

1.2 Performance evaluation for long text documents

1.3 Statistical Significance

We use the following symbolic representation for four categories of statistical analysis:

1. *EdgeGraph significantly outperforms WCN*: The resulting values for $t-test$ are positive and the $p-value < 0.05$. It is represented as *Bold* in $p-value$ column.
2. *EdgeGraph is better than WCN, but not statistically significant*: The resulting values for $t-test$ are positive and the $p-value \not< 0.05$. It is represented as *bold + italics* in $p-value$ column.
3. *WCN is better than EdgeGraph, but not statistically significant*: he resulting values for $t-test$ are negative and the $p-value \not< 0.05$. It is represented as *italics* in $p-value$ column.
4. *WCN significantly outperforms EdgeGraph*: The resulting values for $t-test$ are negative and the $p-value < 0.05$. It is represented as normal text in $p-value$ column.

[1] https://github.com/boudinfl/pke

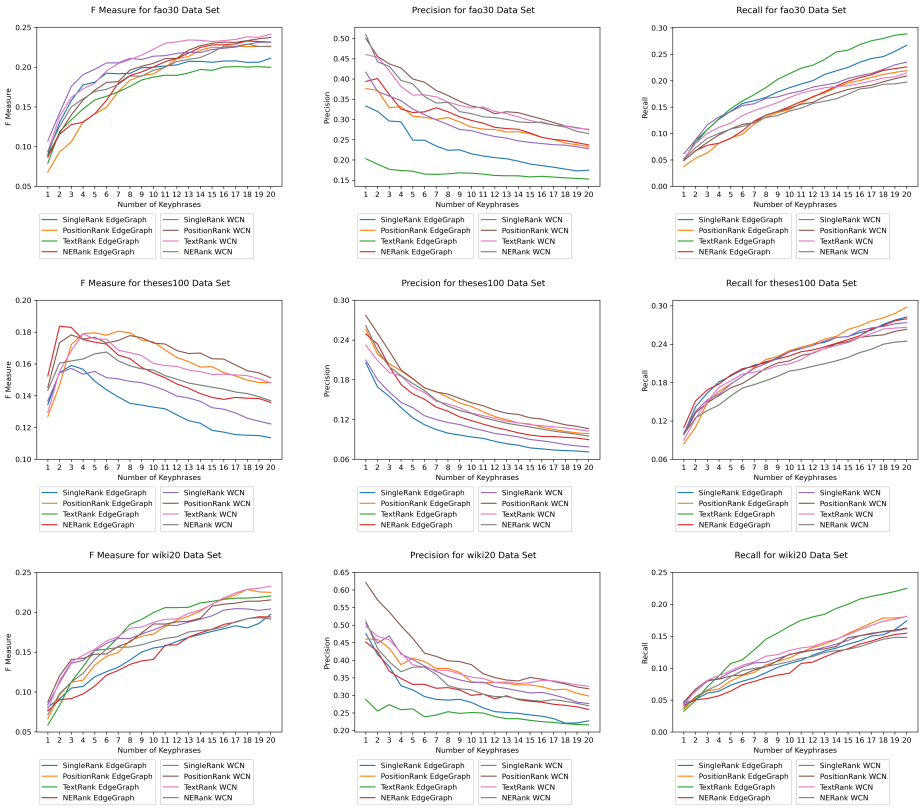

Fig. 1. Performance evaluation of the random walk based GKET over WCN and EdgeGraph representation of long-texts for varying values of k: F-measure, Precision and Recall.

Table 1. Statistical Significance for different keyphrase Extraction over WCN and EdgeGraph.

Dataset Theses-100						
Methods	Recall		Precision		F Measure	
	t_test	p_value	t_test	p_value	t_test	p_value
Text Rank	2.9808	**0.007679**	-15.5573	2.89E-12	-7.2403	7.14E-07
NE Rank	-9.0110	2.74E-08	0.1505	***0.8818***	-4.9817	8.28E-05
Position Rank	24.9520	**5.52E-16**	-7.0996	9.41E-07	1.3439	***0.19478***
Single Rank	10.2857	**3.34E-09**	2.1093	**0.04840**	4.8088	**0.00012**
Dataset FAO 30						
Methods	Recall		Precision		F Measure	
	t_test	p_value	t_test	p_value	t_test	p_value
Text Rank	3.8896	**0.00098**	-29.1109	3.17E-17	-18.0404	2.06E-13
NE Rank	7.7256	**2.80E-07**	-17.7406	2.78E-13	-19.0711	7.55E-14
Position Rank	2.5301	**0.02039**	-6.3402	4.39E-06	-0.6857	0.5011
Single Rank	2.3643	**0.0288**	-2.5759	0.0185	0.6899	***0.4985***
Dataset 500N-KPC						
Methods	Recall		Precision		F Measure	
	t_test	p_value	t_test	p_value	t_test	p_value
Text Rank	8.8480	**3.64E-08**	-11.3725	6.38E-10	10.5934	**2.06E-09**
NE Rank	16.1771	**1.45E-12**	0.8804	***0.3896***	21.4380	**8.98E-15**
Position Rank	12.1549	**2.09E-10**	-14.8203	6.80E-12	20.5013	**2.03E-14**
Single Rank	9.2492	**1.82E-08**	-16.0827	1.61E-12	12.9572	**7.01E-11**
Dataset PAK 2018						
Methods	Recall		Precision		F Measure	
	t_test	p_value	t_test	p_value	t_test	p_value
Text Rank	4.0972	**0.00061**	-0.8953	*0.3817*	-0.4678	*0.6452*
NE Rank	23.8044	**1.31E-15**	13.5307	**3.33E-11**	27.1391	**1.17E-16**
Position Rank	-5.7231	1.62E-05	-4.0292	0.0007	-7.1178	9.08E-07
Single Rank	3.9588	**0.00084**	-1.3992	*0.1778*	-0.3888	*0.7016*
Dataset PT BN KP						
Methods	Recall		Precision		F Measure	
	t_test	p_value	t_test	p_value	t_test	p_value
Text Rank	11.3315	**6.78E-10**	-1.1464	*0.2658*	16.6051	**9.09E-13**
NE Rank	17.2039	**4.83E-13**	-0.7460	*0.4647*	30.9321	**1.02E-17**
Position Rank	22.7884	**2.93E-15**	-12.3763	1.54E-10	19.9476	**3.34E-14**
Single Rank	8.3208	**9.30E-08**	1.4777	***0.1558***	9.008	**2.75E-08**
Dataset WIKI 20						
Methods	Recall		Precision		F Measure	
	t_test	p_value	t_test	p_value	t_test	p_value
Text Rank	-5.0775	6.69E-05	-17.7666	2.71E-13	-12.0685	2.35E-10
NE Rank	5.5577	**2.32E-05**	-16.2917	1.28E-12	-1.1878	*0.2495*
Position Rank	-3.3987	0.0030	-4.0372	0.00070	-3.9784	0.0008
Single Rank	0.6545	***0.5206***	-5.6385	1.95E-05	-0.9103	*0.3740*