



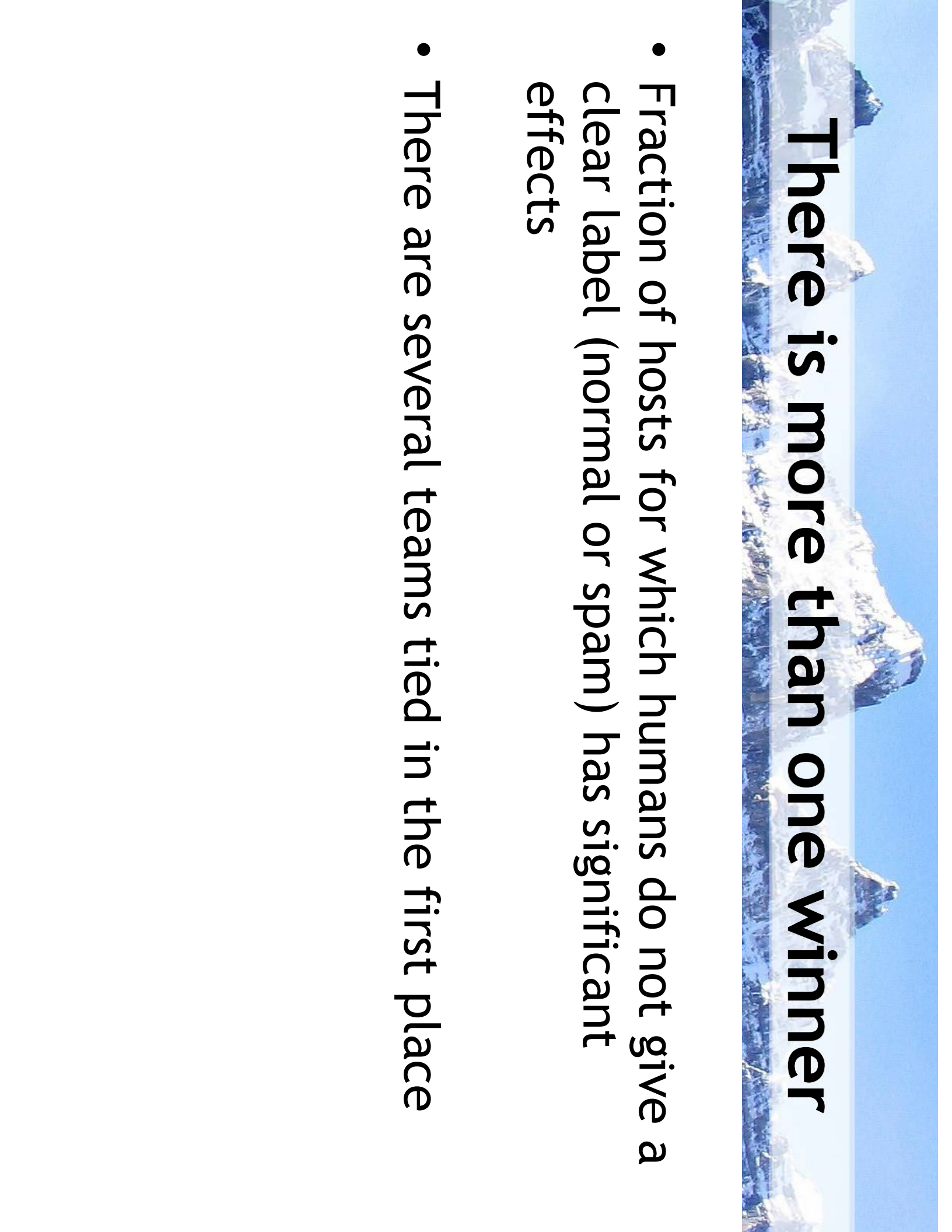
AIRWeb'07

Third International Workshop on
Adversarial Information Retrieval on the Web

Web Spam Challenge Track I

<http://webspam.lip6.fr/>

Supported by the EU PASCAL Network of Excellence Challenge Program



There is more than one winner

- Fraction of hosts for which humans do not give a clear label (normal or spam) has significant effects
- There are several teams tied in the first place



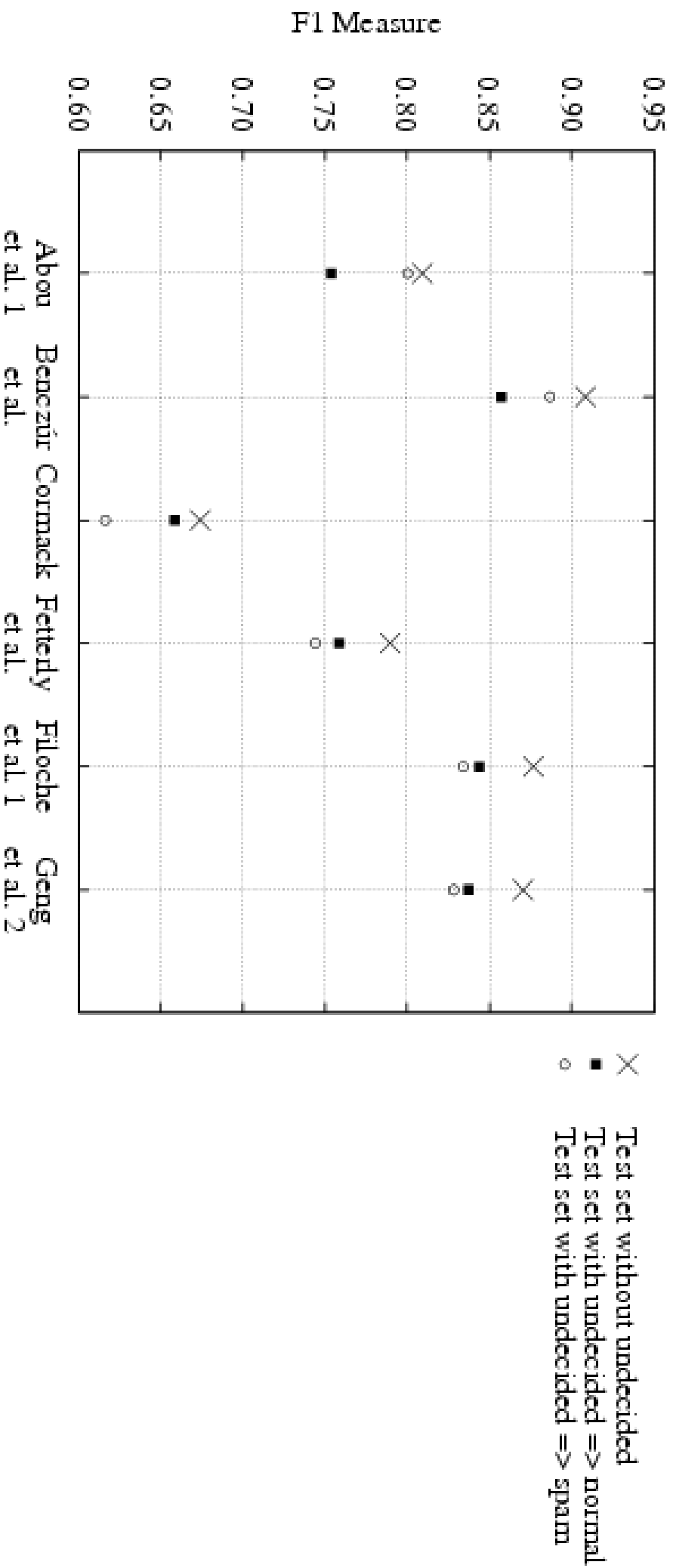
Evaluation by F-Measure

Handling of undecided hosts

	Discard	Normal	Spam
Benczúr et al.	0.91	0.86	0.89
Filoché et al. 1	0.88	0.84	0.83
Geng et al. 2	0.87	0.84	0.83
Abou et al. 1	0.81	0.73	0.80
Fetterly et al.	0.79	0.76	0.74
Cormack	0.67	0.66	0.62

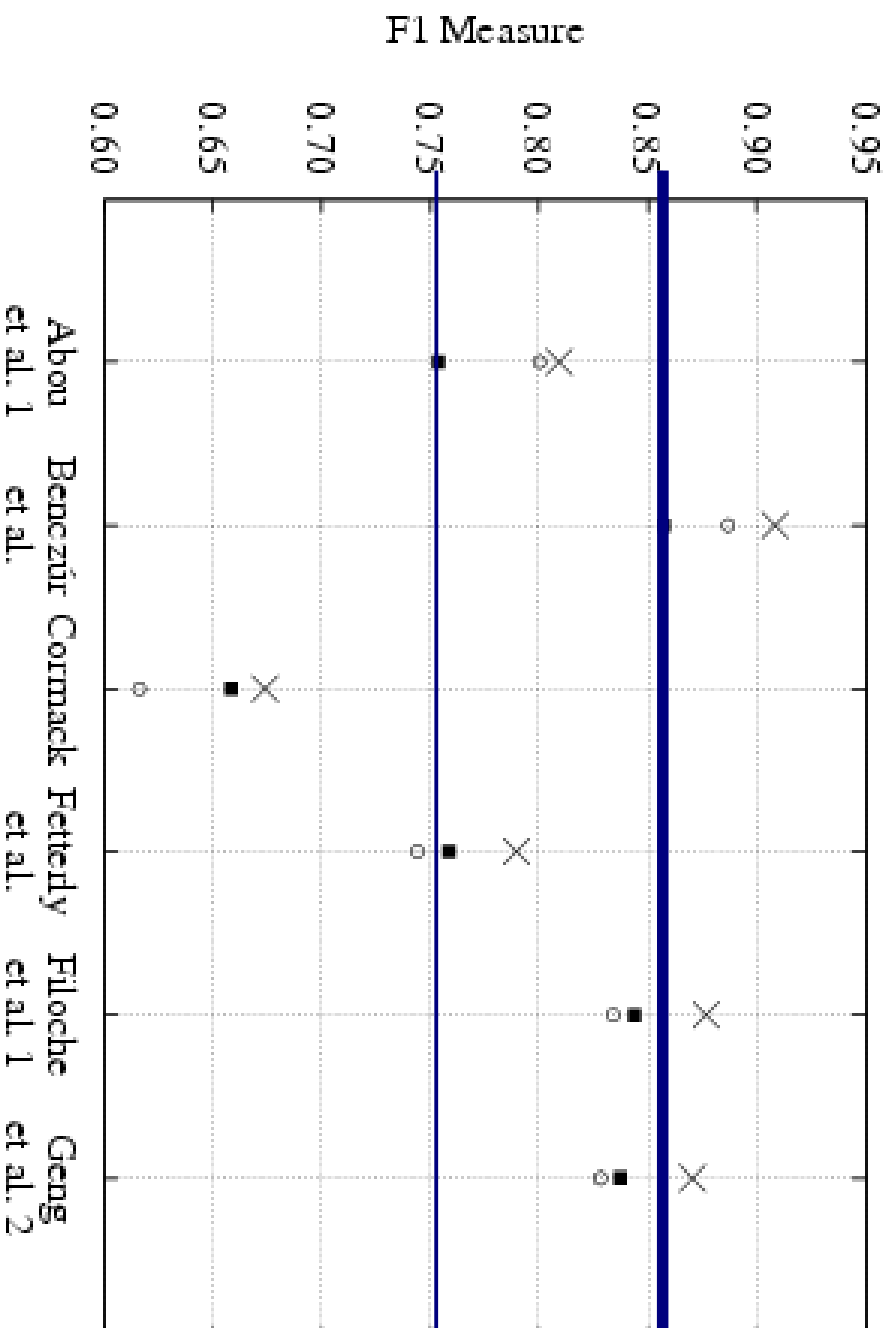
F-Measure

F1 Measure of Spam Classification in Test Set

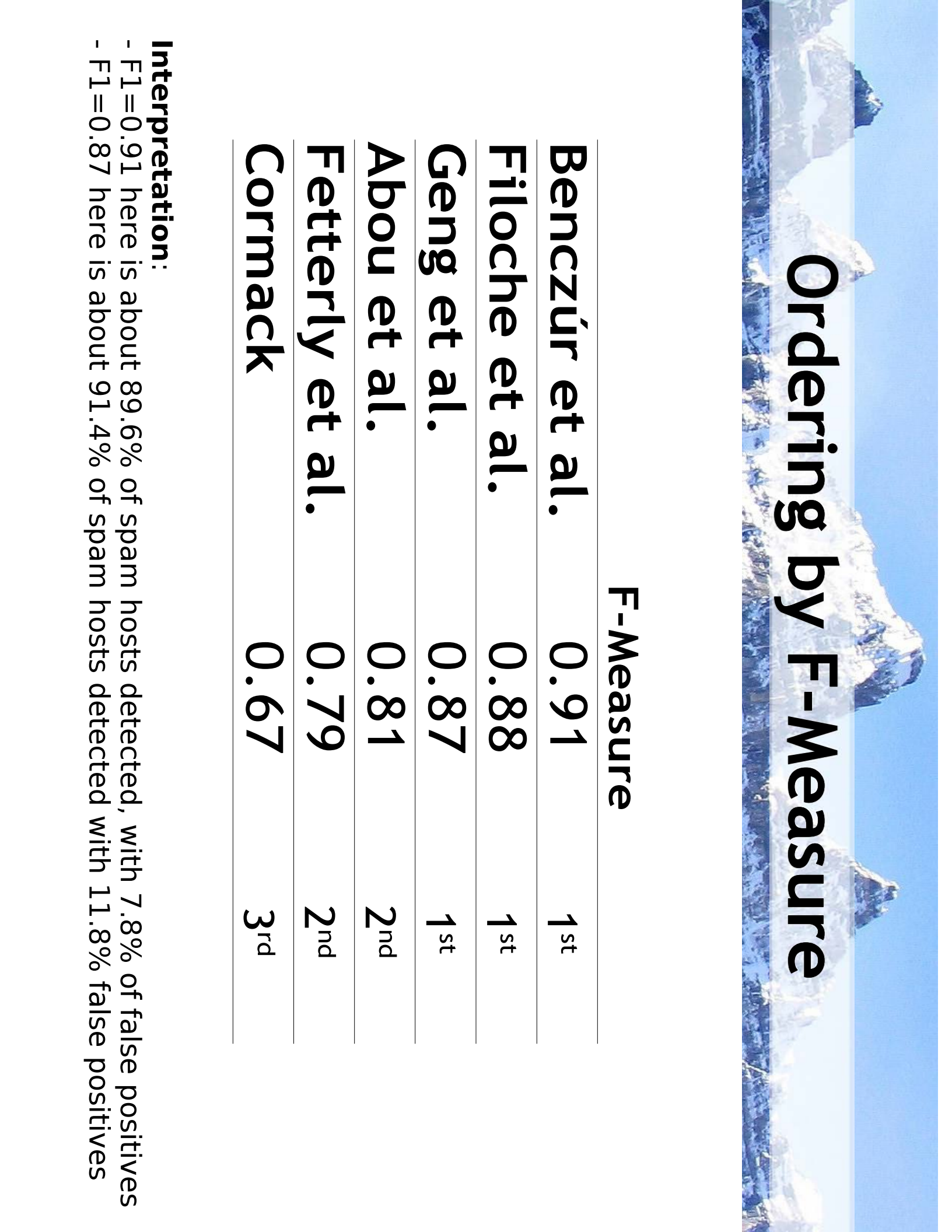


F-Measure

F1 Measure of Spam Classification in Test Set



1st place by F1

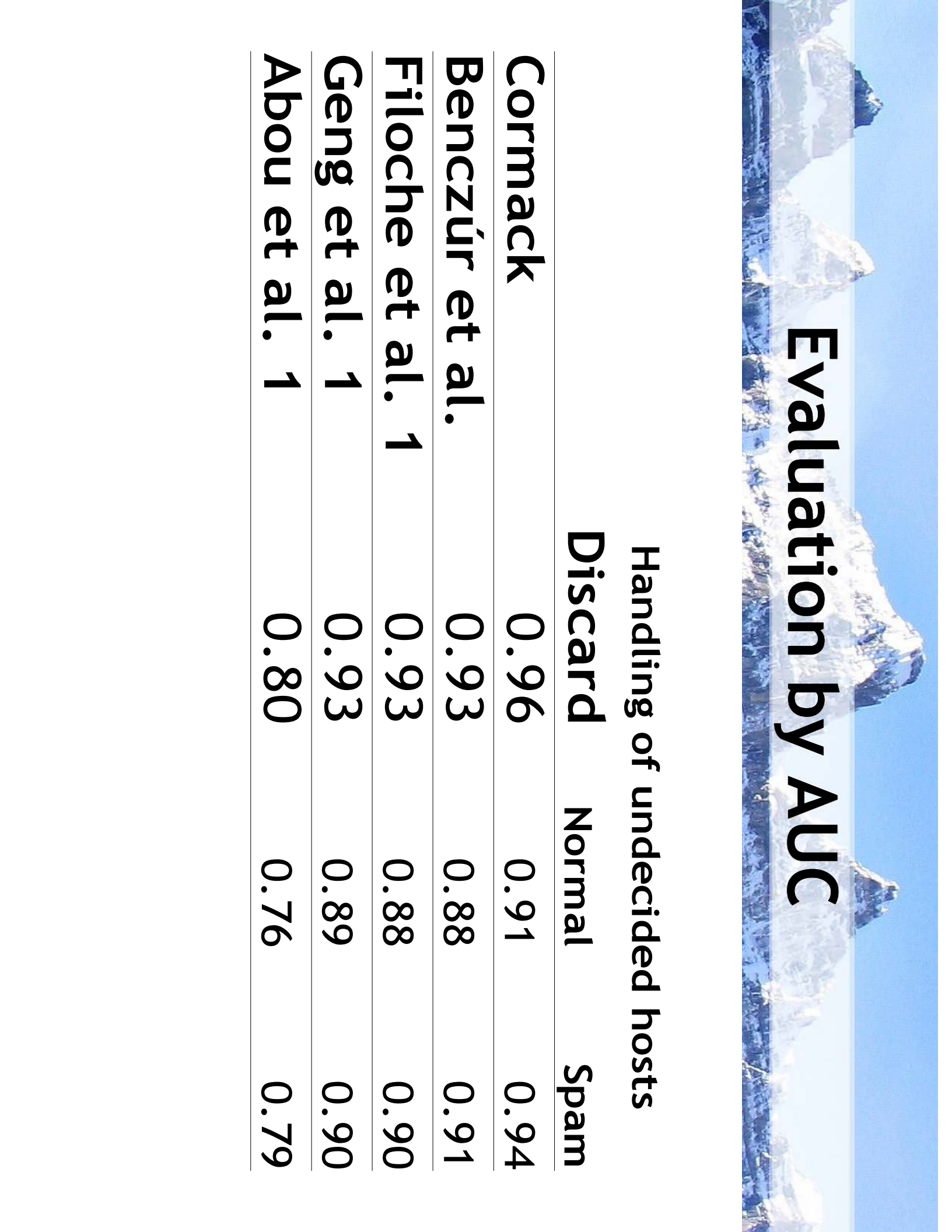


Ordering by F-Measure

F-Measure		
Benczúr et al.	0.91	1 st
Filoché et al.	0.88	1 st
Geng et al.	0.87	1 st
Abou et al.	0.81	2 nd
Fetterly et al.	0.79	2 nd
Cormack	0.67	3 rd

Interpretation:

- $F1=0.91$ here is about 89.6% of spam hosts detected, with 7.8% of false positives
- $F1=0.87$ here is about 91.4% of spam hosts detected with 11.8% false positives



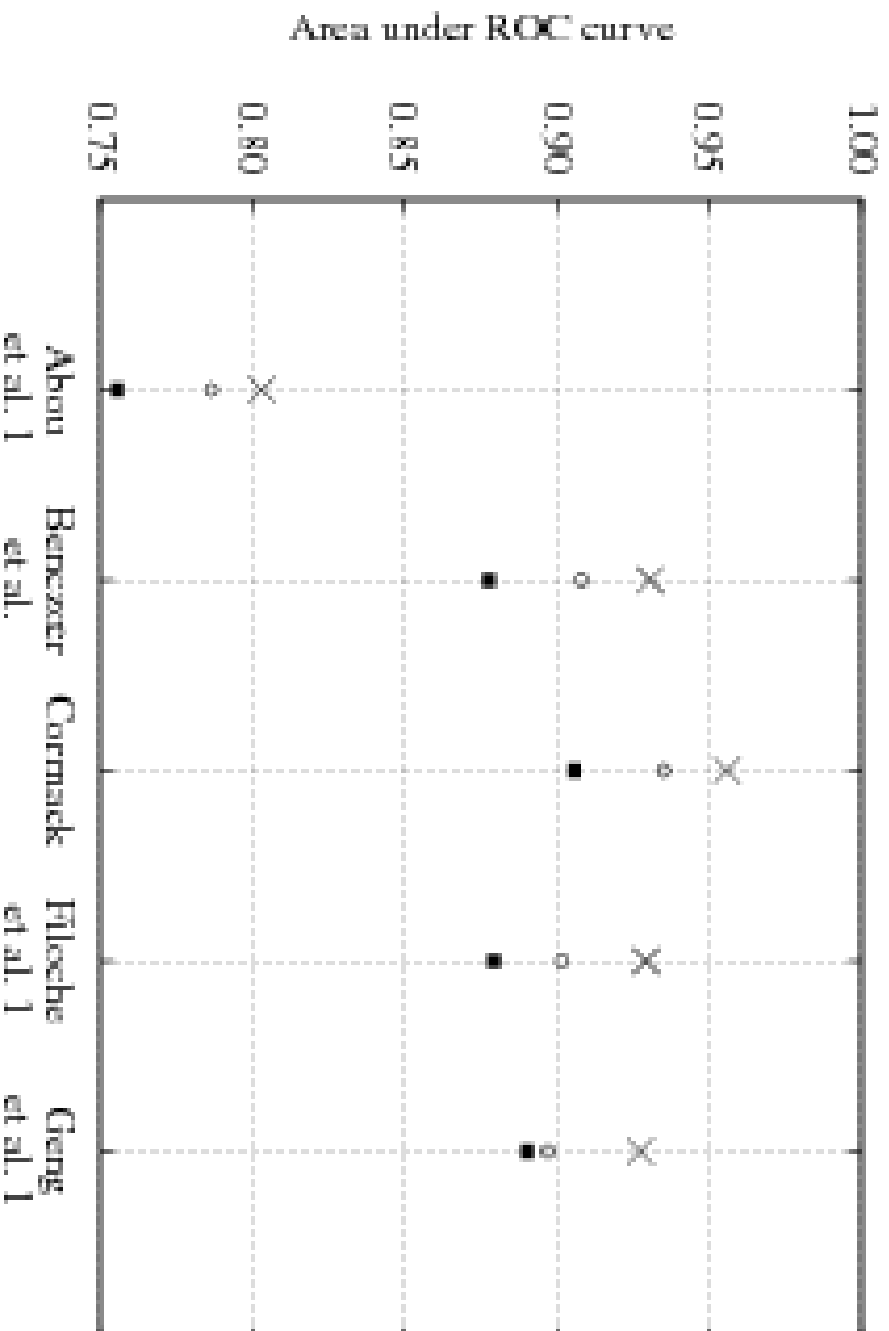
Evaluation by AUC

Handling of undecided hosts

	Discard	Normal	Spam
Cormack	0.96	0.91	0.94
Benczúr et al.	0.93	0.88	0.91
Filoché et al. 1	0.93	0.88	0.90
Geng et al. 1	0.93	0.89	0.90
Abou et al. 1	0.80	0.76	0.79

Area under ROC curve

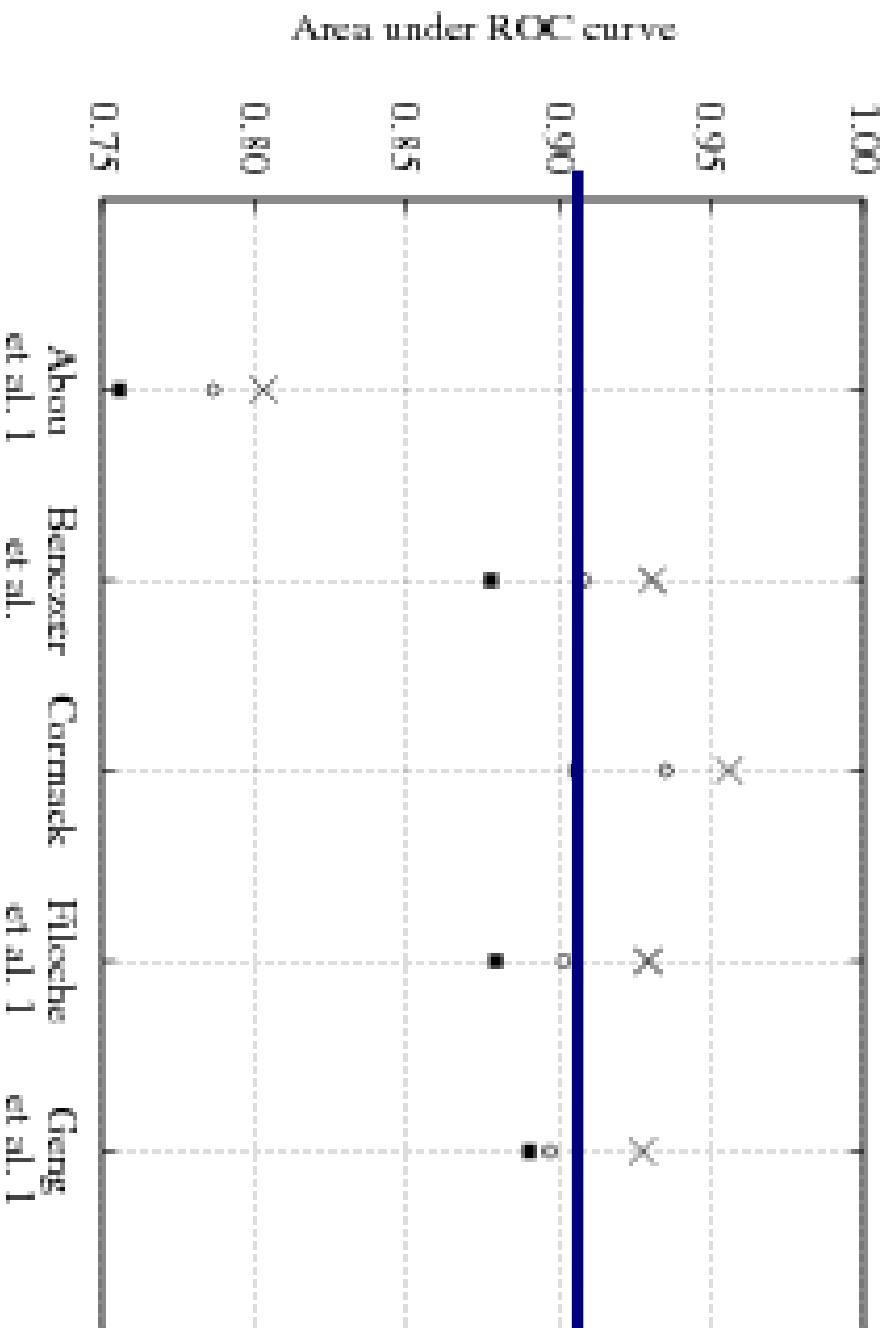
AUC of Spam Classification in Test Set



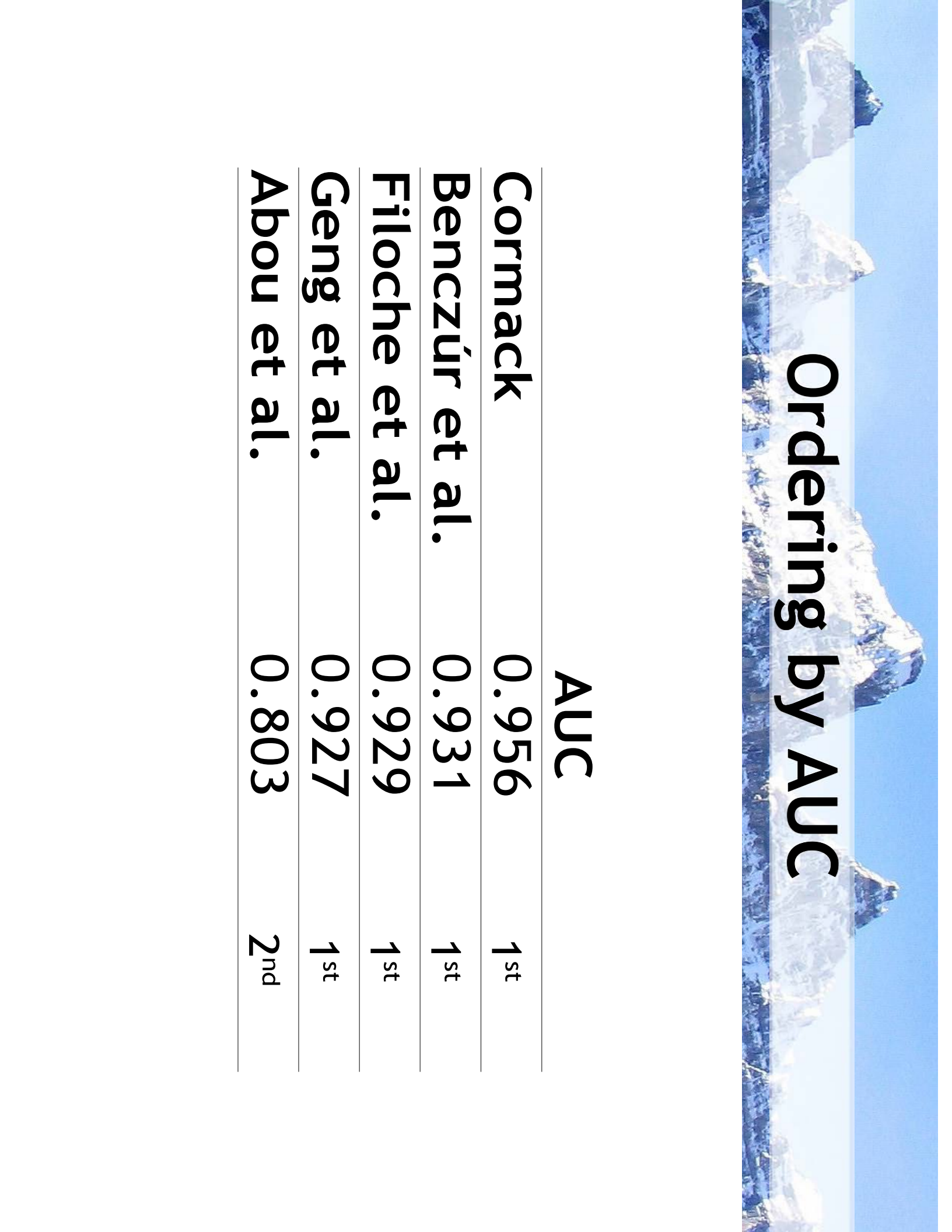
X Test set without undecided
■ Test set with undecided => normal
● Test set with undecided => spam

Area under ROC curve

AUC of Spam Classification in Test Set

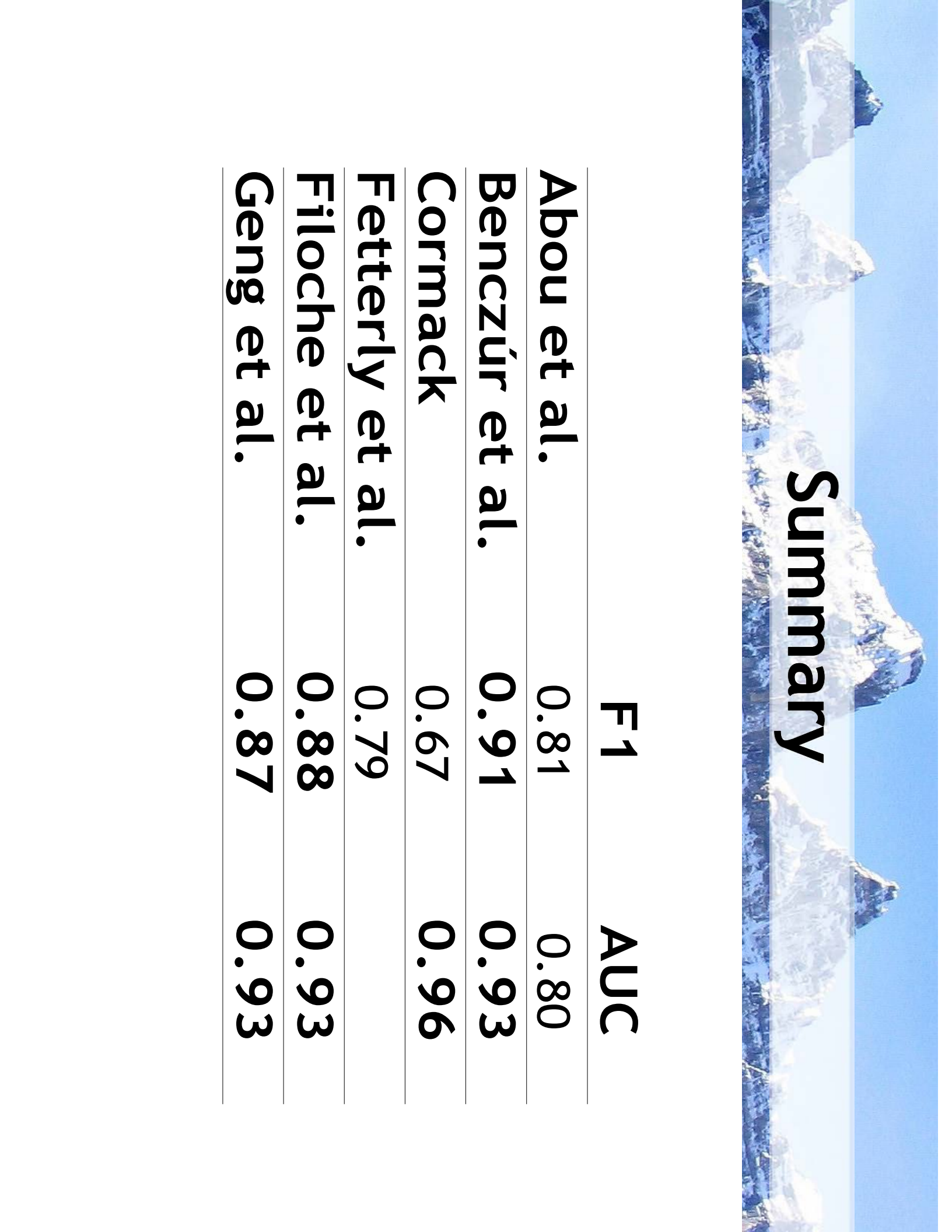


1st place by AUC



Ordering by AUC

AUC		
Cormack	0.956	1 st
Benczúr et al.	0.931	1 st
Filoché et al.	0.929	1 st
Geng et al.	0.927	1 st
Abou et al.	0.803	2 nd



Summary

	F1	AUC
Abou et al.	0.81	0.80
Benczúr et al.	0.91	0.93
Cormack	0.67	0.96
Fetterly et al.	0.79	
Filoché et al.	0.88	0.93
Geng et al.	0.87	0.93



Thanks to all the participants!

- Tony Abou-Assaleh and Tapaigyoti Das
- András A. Benczúr, István Bíró, Károly Csalogány, Miklós Kurucz and Tamás Sarlós
- Gordon Cormack
- Pascal Filoche, Tanguy Urvoy, Chauveau Emmanuel and Lavergne Thomas
- Dennis Fetterly, Steve Chien, Marc Najork, Mark Manasse and Alexandros Ntoulas
- Guanggang Geng, Chunheng Wang, Xiaobo Jin, Qiudan Li and Lei Xu

ECML/PKDD Graph Labeling Workshop

Warsaw, Poland - September 2007

Web Spam Challenge Track II

<http://webspam.lip6.fr/>

Supported by the EU PASCAL Network of Excellence Challenge Program

- Also supported by the EU Pascal Network of Excellence Challenge Program
- Jointly organized with the ECML/PKDD Workshop on Graph Labeling 2007
- A single set of features, with all the labeling done before the competition
- Feedback from Track I is welcome
- See <http://graphlab.lip6.fr/>