

**Subject:** [Seminario lunedì 11 mar 2019 ore 15:30] A. Funel - "Analysis of the Web Graph Aggregated by Host and Pay-Level Domain"  
**From:** "eneagrid administrators" <eneagrid.noreply@enea.it>  
**Date:** 08/03/19, 09:30  
**To:** griduser

Buongiorno,

lunedì 11 marzo 2019, alle ore 15.30, Agostino Funel (ENEA) terrà un seminario dal titolo "Analysis of the Web Graph Aggregated by Host and Pay-Level Domain".

Il seminario presenta i risultati di un articolo sull'analisi del grafo del web pubblicato e presentato alla conferenza internazionale "Complex Networks 2018" tenutasi a Cambridge (UK) nel periodo 11-13 Dicembre 2018.

Link della conferenza:  
<https://www.2018.complexnetworks.org/>

Tutti i lavori della conferenza sono disponibili mediante credenziali ENEAGRID/CRESCO al link:  
[https://www.eneagrid.enea.it/conferences/complex\\_networks\\_2018/](https://www.eneagrid.enea.it/conferences/complex_networks_2018/)

Link dell'articolo su arXiv:  
<http://arxiv.org/abs/1802.05435>

Il seminario sarà trasmesso online via Connect al seguente indirizzo: <https://connect.portici.enea.it/hpc>

=====

Good morning,

on Monday 11 March 2019 at 15:30 Agostino Funel (ENEA) will hold a seminar entitled "Analysis of the Web Graph Aggregated by Host and Pay-Level Domain".

The seminar will present the results of a published article presented at the international conference "Complex Networks 2018" held in Cambridge (UK) from 11 to 13 December 2018.

Conference link:  
<https://www.2018.complexnetworks.org/>

The proceedings of te conference are available at (needed ENEAGRID/CRESCO credentials):  
[https://www.eneagrid.enea.it/conferences/complex\\_networks\\_2018/](https://www.eneagrid.enea.it/conferences/complex_networks_2018/)

Link of the article on arXiv:  
<http://arxiv.org/abs/1802.05435>

The seminar will be transmitted online via Connect at:  
<https://connect.portici.enea.it/hpc>

=====

\*\*\*\*\* ABSTRACT \*\*\*\*\*

[Analysis of the Web Graph Aggregated by Host and Pay-Level Domain]

In this paper the web is analyzed as a graph aggregated by host and pay-level domain (PLD). The web graph datasets, publicly available, have been released by the Common Crawl Foundation and are based on a web crawl performed during the period May-June-July 2017. The host graph has ~1.3 billion nodes and ~5.3 billion arcs. The PLD graph has ~91 million nodes and ~1.1 billion arcs. We study the distributions of degree and sizes of strongly/weakly connected components (SCC/WCC) focusing on power laws detection using statistical methods. The statistical plausibility of the power law model is compared with that of several alternative distributions. While there is no evidence of power law tails on host level, they emerge on PLD aggregation for indegree, SCC and WCC size distributions. Finally, we analyze distance-related features by studying the cumulative distributions of the shortest path lengths, and give an estimation of the diameters of the graphs.

---

Per favore non rispondere a questo E-mail.  
In caso di necessità utilizzare <http://gridticket.enea.it> per contattare gli amministratori di eneagrid/CRESCO  
Informazioni su eneagrid/CRESCO sono disponibili a <http://www.eneagrid.enea.it> e <http://www.cresco.enea.it>

---

Please do not reply to this E-mail.  
In case use <http://gridticket.enea.it> to contact eneagrid/CRESCO administrators  
See <http://www.eneagrid.enea.it> and <http://www.cresco.enea.it> for information about eneagrid/CRESCO

=====

This message is sent to you because you are subscribed to the mailing list [<griduser@enea.it>](mailto:griduser@enea.it).

To unsubscribe, E-mail to: [<griduser-off@enea.it>](mailto:griduser-off@enea.it) but please take care that the membership to the list is required to have access as an user to ENEAGRID/CRESCO computational resources.

To contact ENEAGRID/CRESCO administrators please submit a ticket to <http://gridticket.enea.it/>