

# Modeling Grid Job Time Properties

*Lovro Ilijašić*

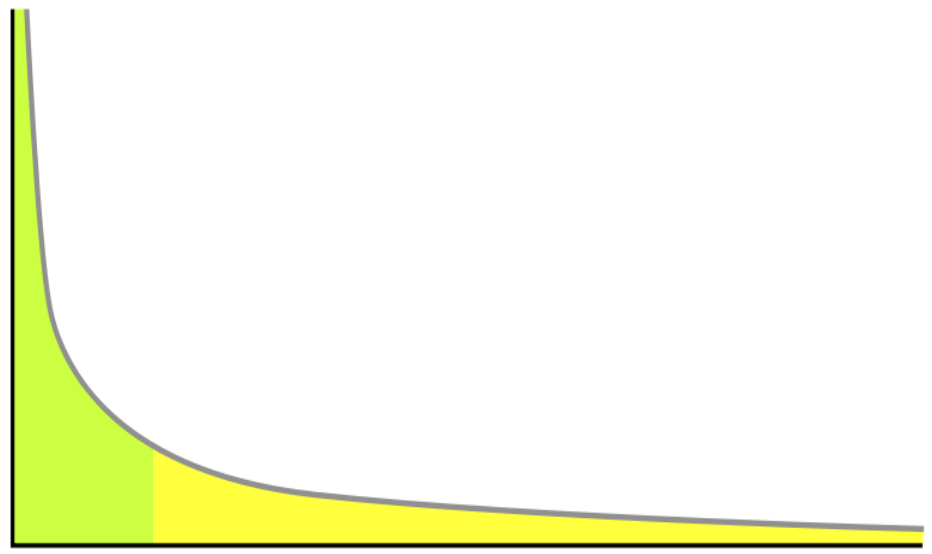
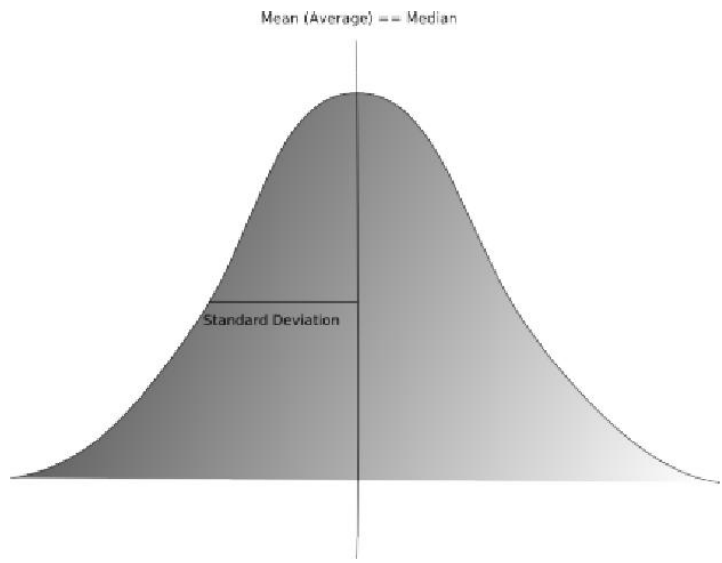
*Lorenza Saitta*

*University of Eastern Piedmont, Italy*

- **The Grid Observatory cluster of EGEE – the scientific view**
- **Data collection, analysis of behaviour and usage**
- **20 months of data, more than 28 million jobs**
- **Development of models**
  
- **Grid is more than just a sum of its parts**

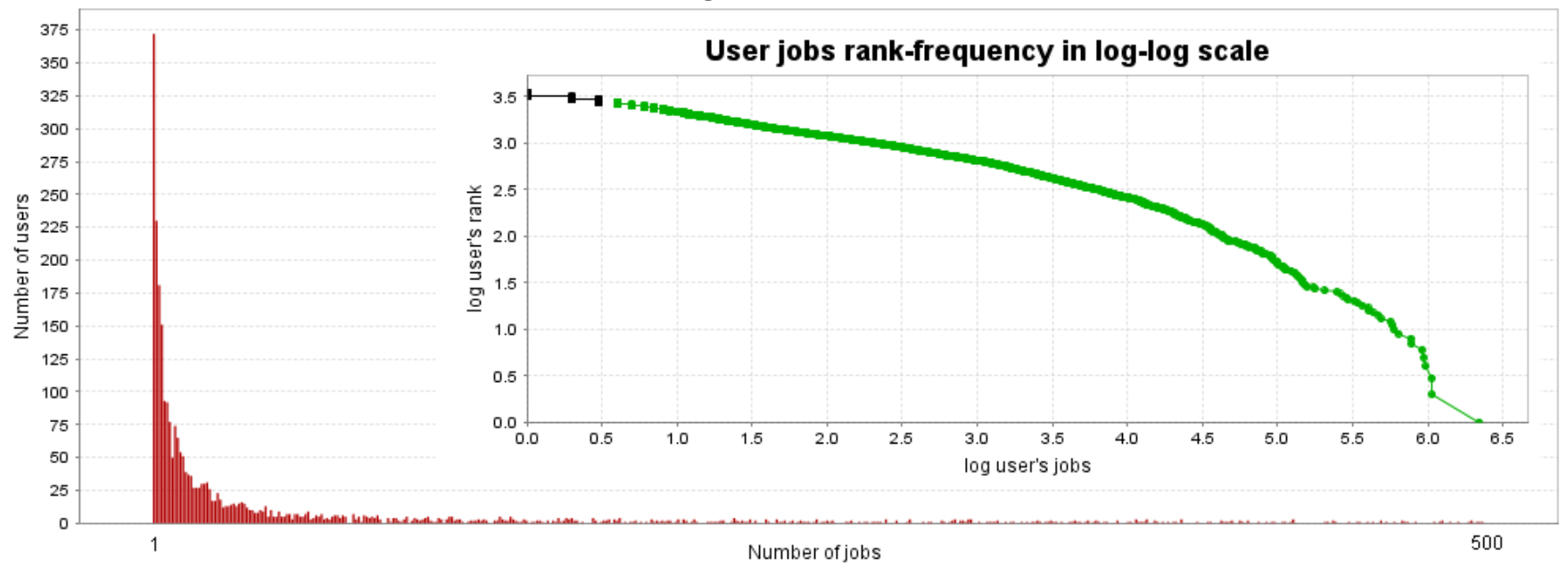
- **Properties that are apparent only on higher levels of organization and are not present on the lower ones**
- **Emergent Behaviour is observable on all levels of reality**

- Pareto distribution
- Zipf's law
- 80-20 rule
- Self similarity

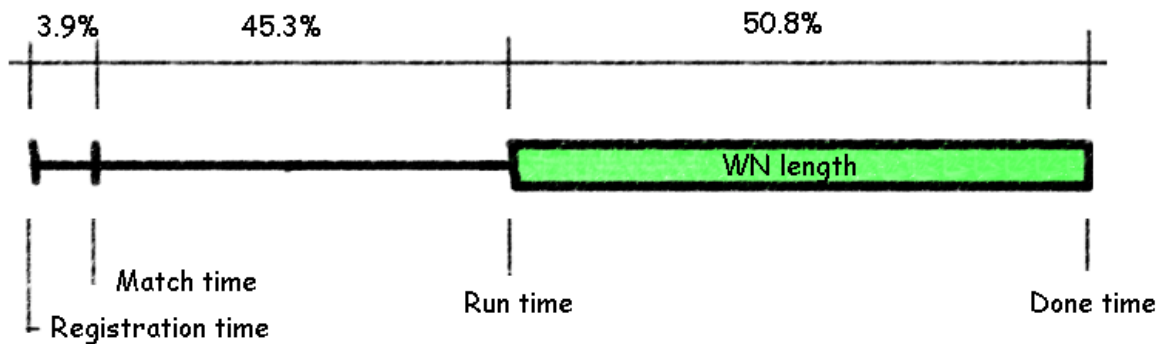
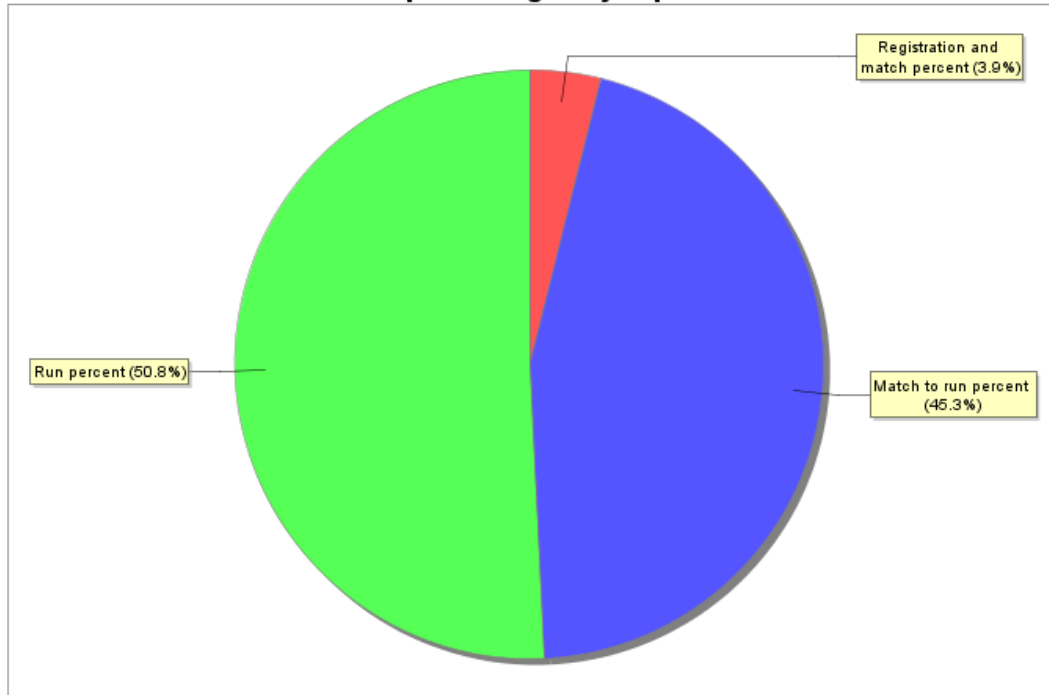


- In- and out-degree distributions: How users connect (use) CEs
- Weighted degrees: Distribution of number of jobs

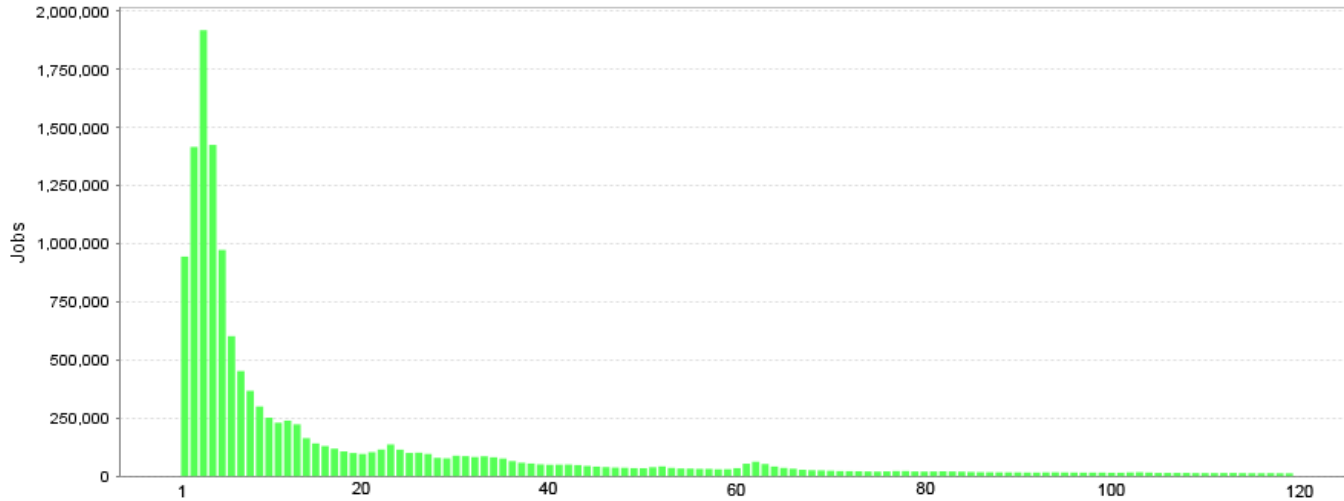
**User jobs distribution**



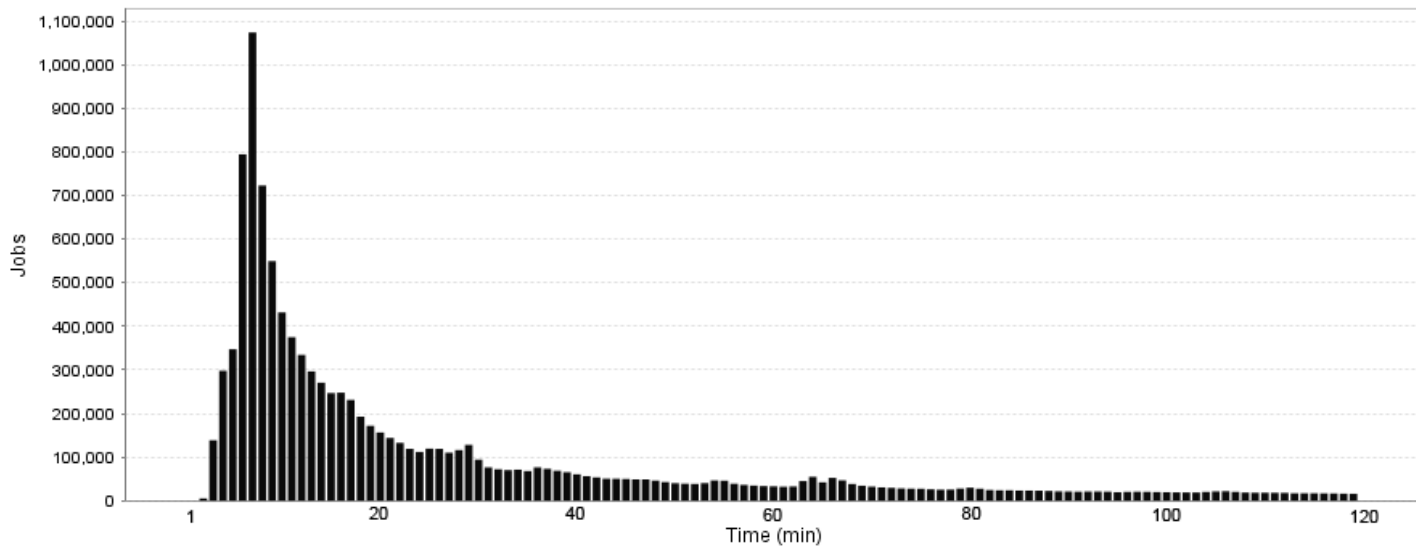
Mean percentage of job parts



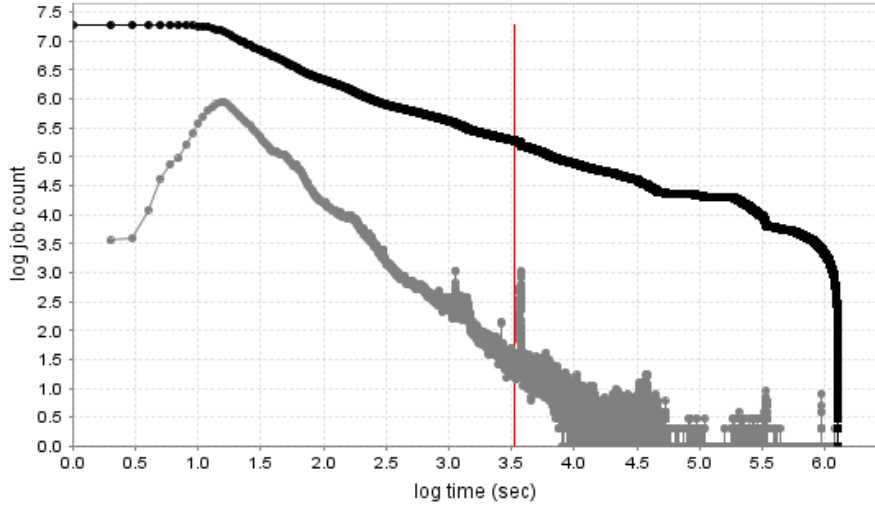
**Worker Node time linear distribution**



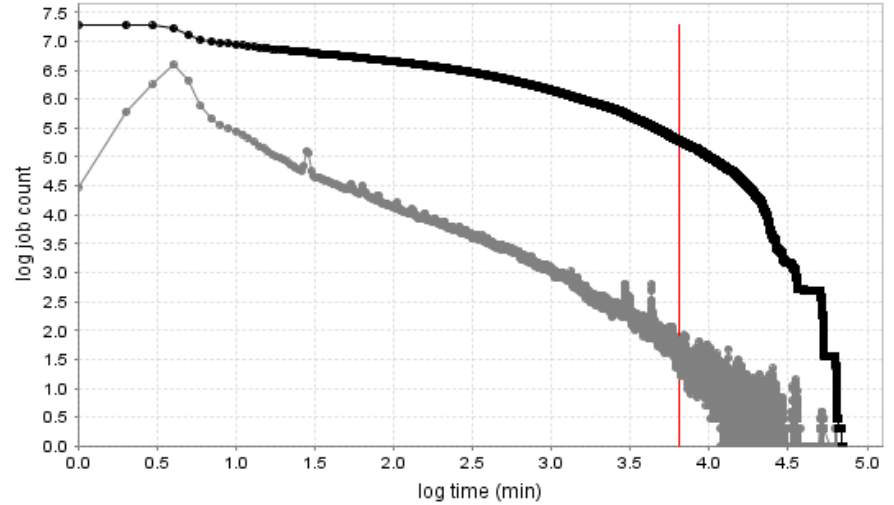
**Total time linear distribution**



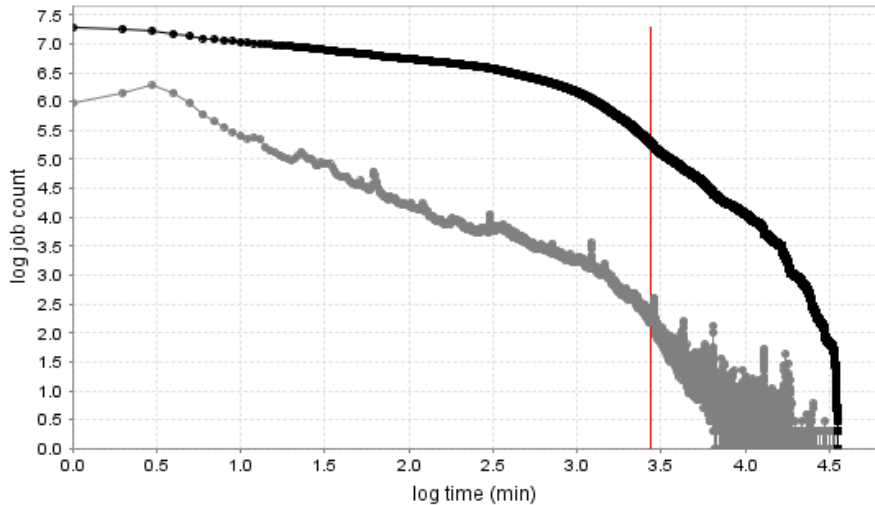
**a) Match time distributions in log-log scale**



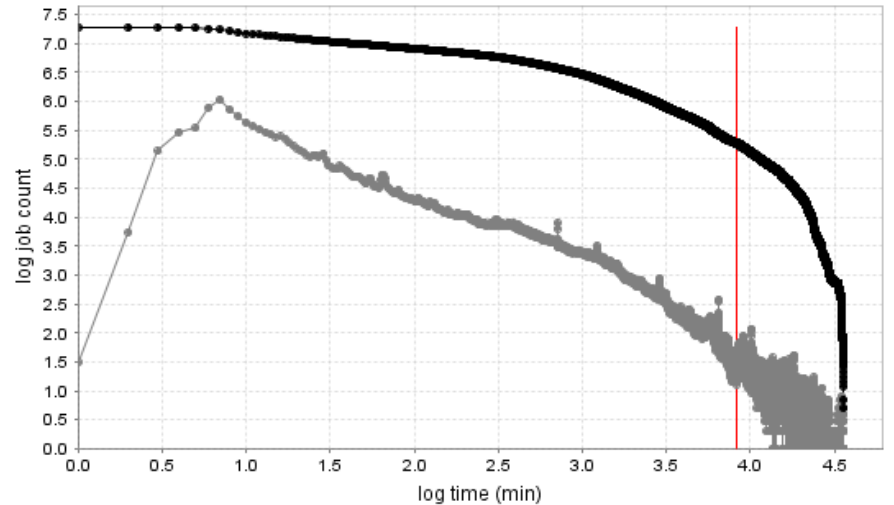
**b) Wait time distributions in log-log scale**



**c) Run time distributions in log-log scale**



**d) Total time distributions in log-log scale**





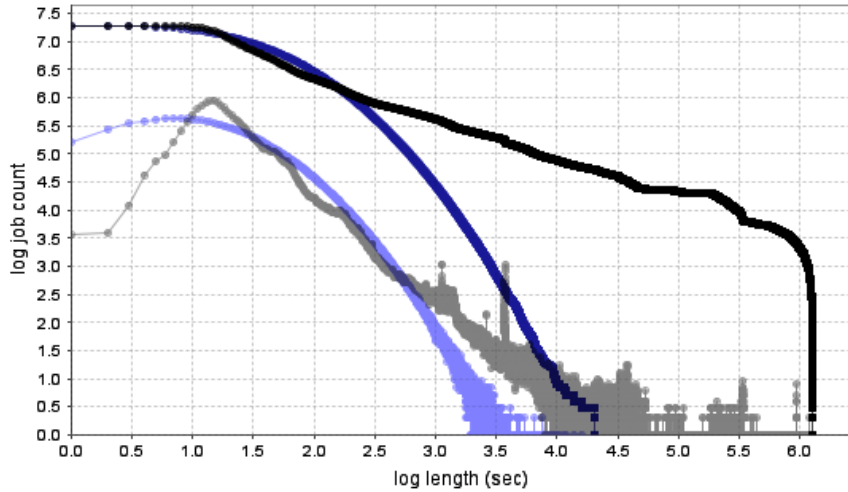
- **Power-law: preferential attachment**
- **Power-law: optimization of the average amount of information per unit transmission cost**
- **Power-law: monkeys typing randomly**
- **Probabilities of letters not equal: power-law or log-normal?**

$$p(x) \propto x^{-\alpha}$$

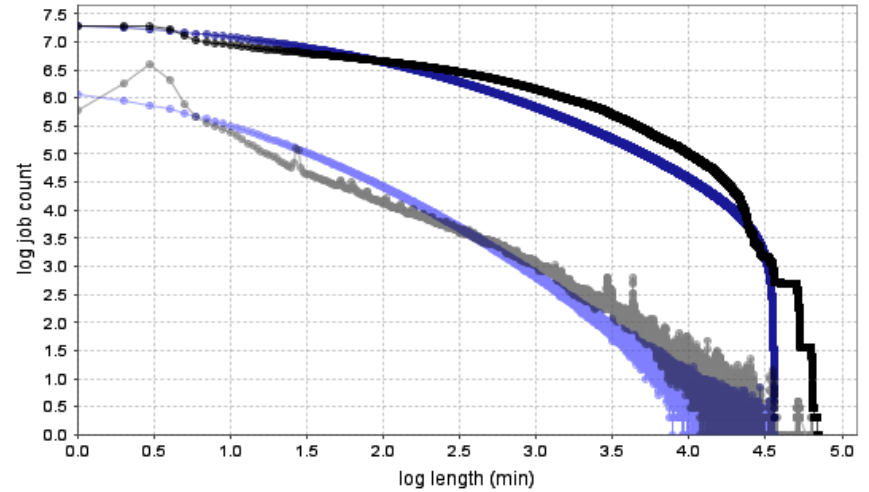
- **Log-normal: multiplicative processes**
- **At each step, the event ( $X_t$ ) may grow or shrink, according to a random variable  $F_t$ :  $X_t = F_t X_{t-1}$**
- **Multiplicative models can also generate Pareto distribution if there is not a minimum size of event. Otherwise it is log-normal**
- **Intermixing of generations, where  $t$  is random variable, leads to power law.**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

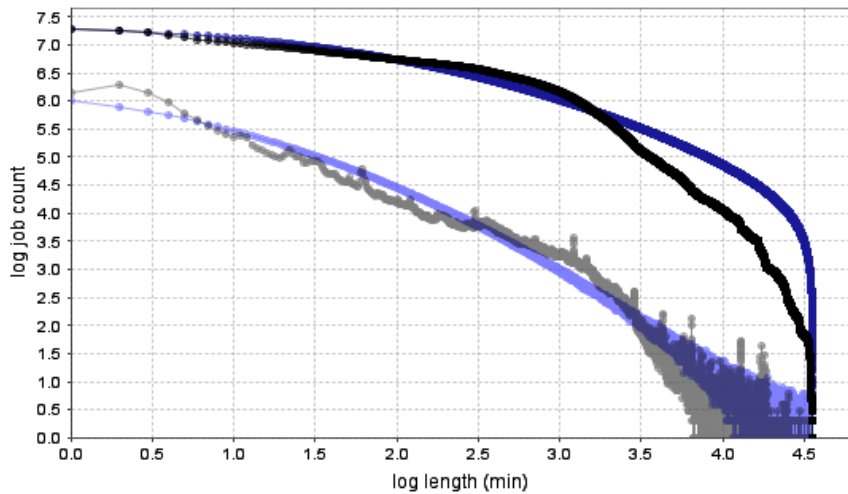
**a) Log-normal fitted match time**



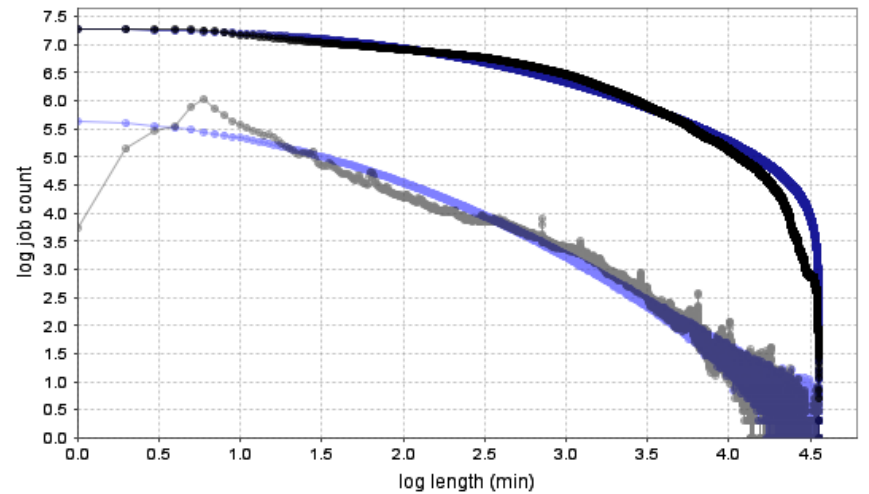
**b) Log-normal fitted wait time**



**c) Log-normal fitted run time**



**d) Log-normal fitted total time**



● Real data PDF    ● Real data CCDF    ● Fitted log-normal data PDF    ● Fitted log-normal data CCDF

- **Double Pareto distribution**
- **Double Pareto log-normal distribution**
- **More distribution parameters that allow better fitting**

# Modeling Grid Job Time Properties

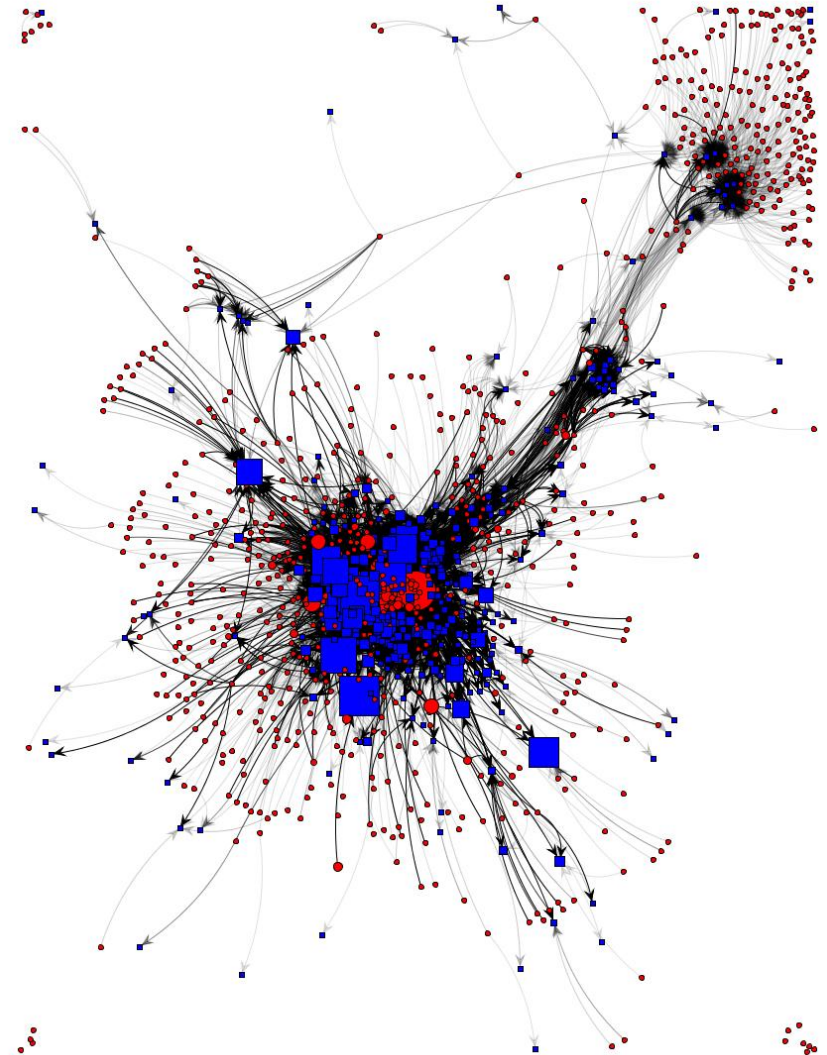
*Lovro Ilijašić*

*Lorenza Saitta*

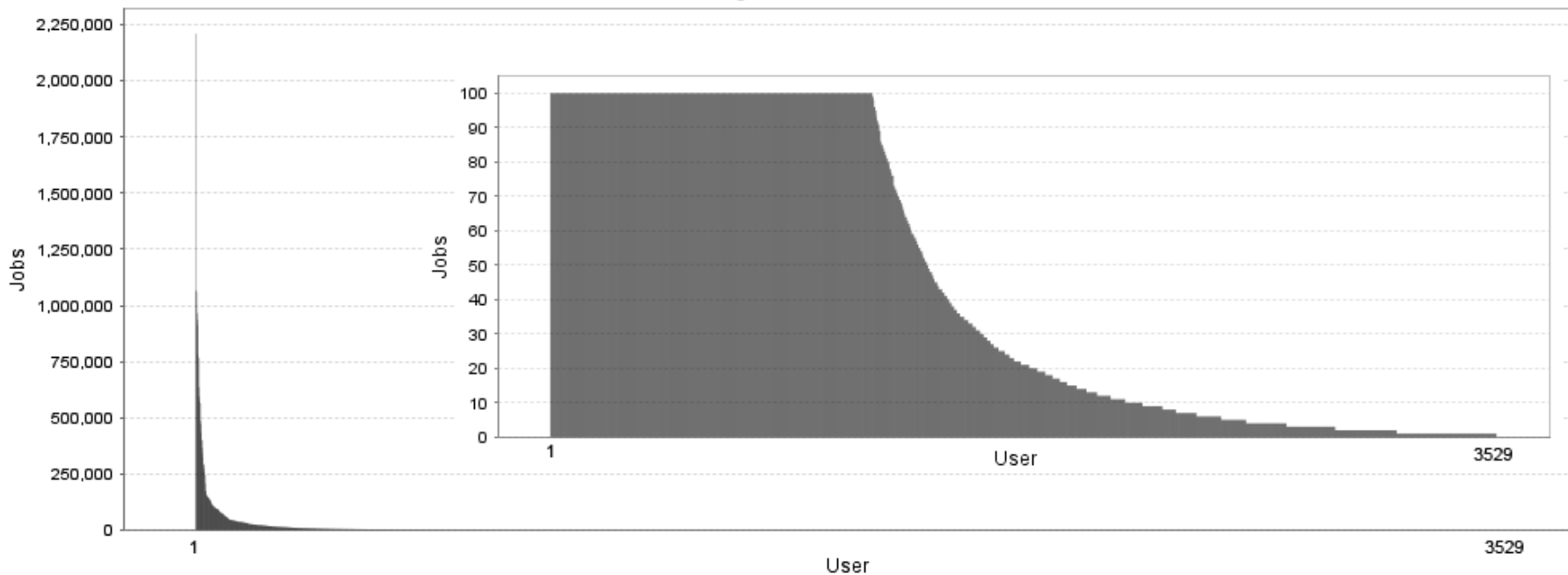
*University of Eastern Piedmont, Italy*



- **Complex Networks – Complex systems represented as graphs**
- **Gathered experiences from Physics, Chemistry, Biology, Computer Science, Sociology, Economics...**
- **Representing Grid as a Complex Network**
- **20 months of log data, more than 28 million jobs**
- **Edges representing jobs go from Users to CEs**

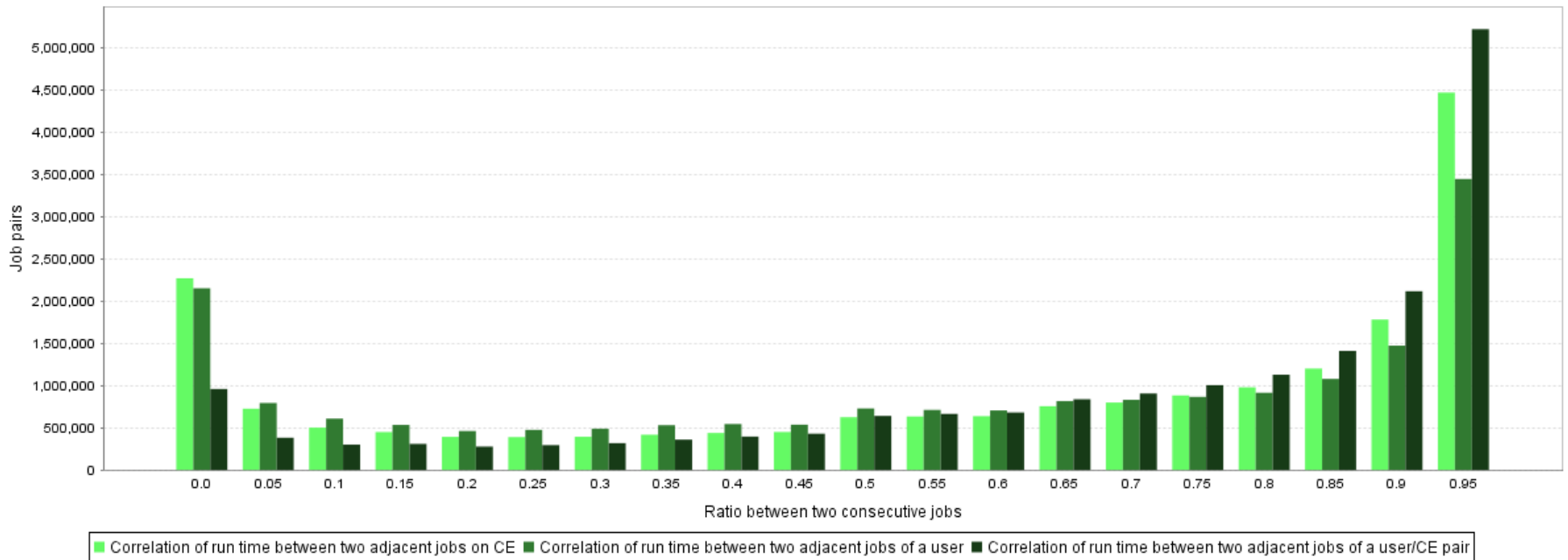


Number of jobs for each user

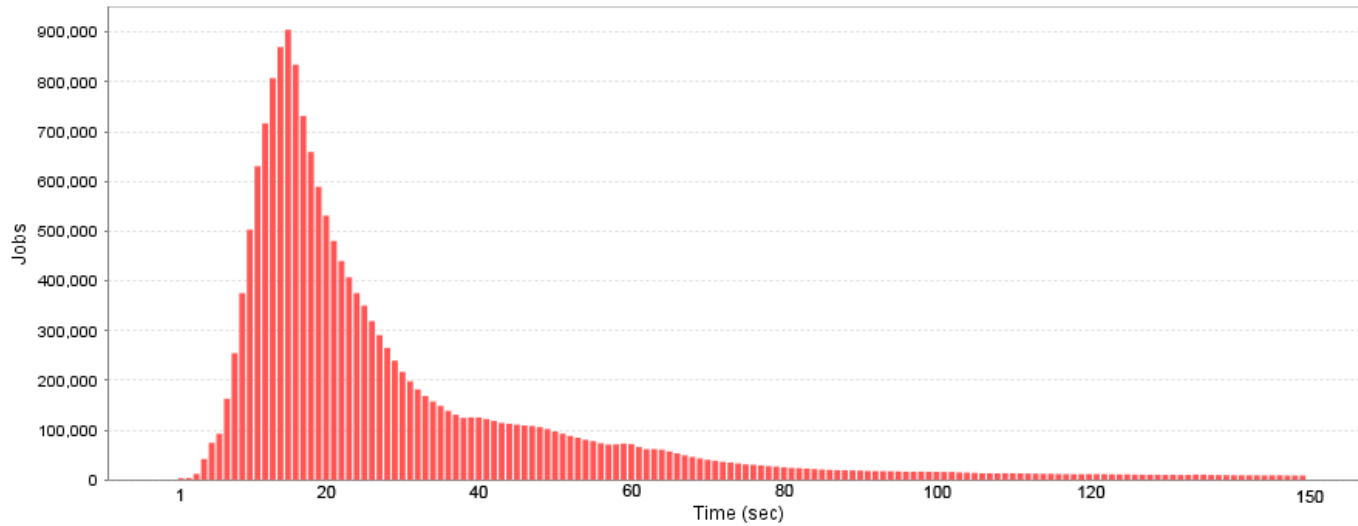




**Distribution of run length ratio between two consecutive jobs**



**Register and Match time distribution linear (1 second bin)**



**Match to Run time linear distribution**

