Analysis of the network of collaboration between European Institutions

Panos Argyrakis

University of Thessaloniki, Thessaloniki, Greece

KNOWeSCAPE

General idea:

Employing network tools and analysis on real data

♪
♪

What data:

European Commission Framework Programs,
the FP6 and FP7 datasets

How did we get the data:

Provided by the Office of Statistics of the European Commission

Form of the data:

- ✓ FP6 and FP7
- ✓ 2 databases for each
  - projects (accepted proposals)
  - all proposals (accepted and rejected combined)
- ✓ In anonymised form
- ✓ Microsoft Access format

Preparatory work:

Data records missing or faulty
Tedious, time-consuming procedures needed to clean the datasets

Goals:

- ✓ Apply network analysis tools on every available network level of the two datasets (FP6 and FP7)
- ✓ Examine the results
- ✓ Compare the corresponding networks
- ✓ Identify any possible differences in their structure

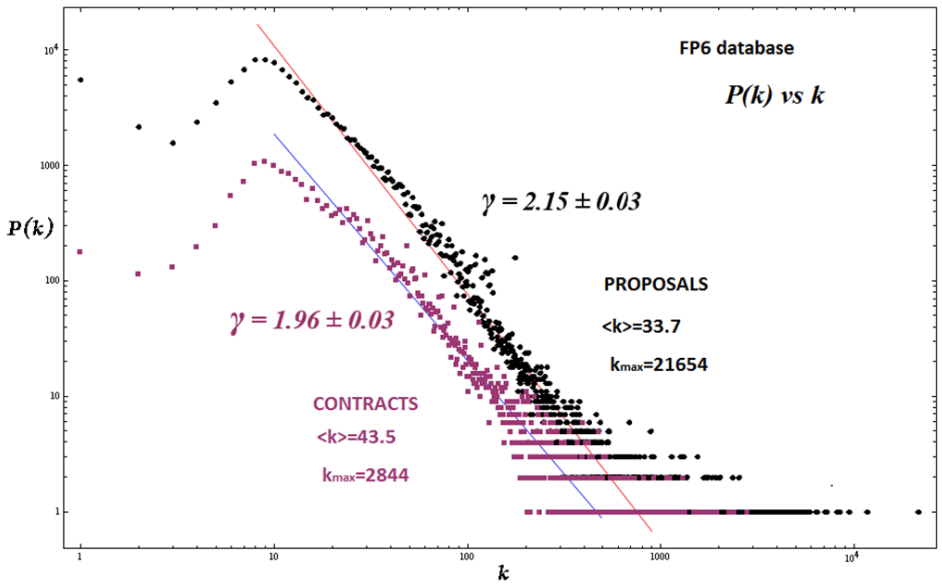
Level 1, participants - applicants:

For the FP6, FP7 projects and proposals datasets:

- ✓ Construct the network of participants/applicants for both datasets
- ✓ Compare the probability distribution of the two networks
- ✓ FP6: 43565 applicants, 22375 participants
- ✓ FP7: 52742 applicants, 24382 participants

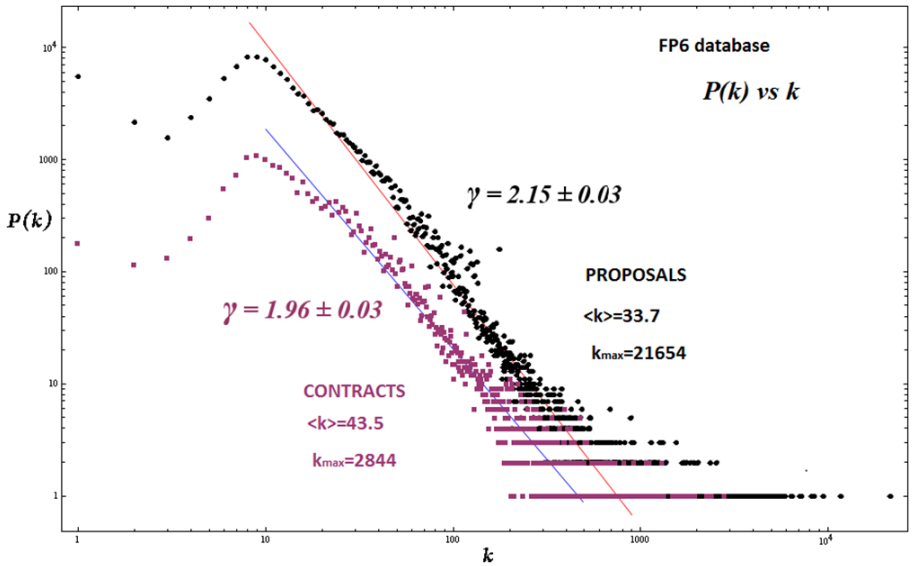
FP6: degree probability distribution comparison for the participants and applicants networks:

- ✓ The probability distribution of both networks exhibit power law behaviour



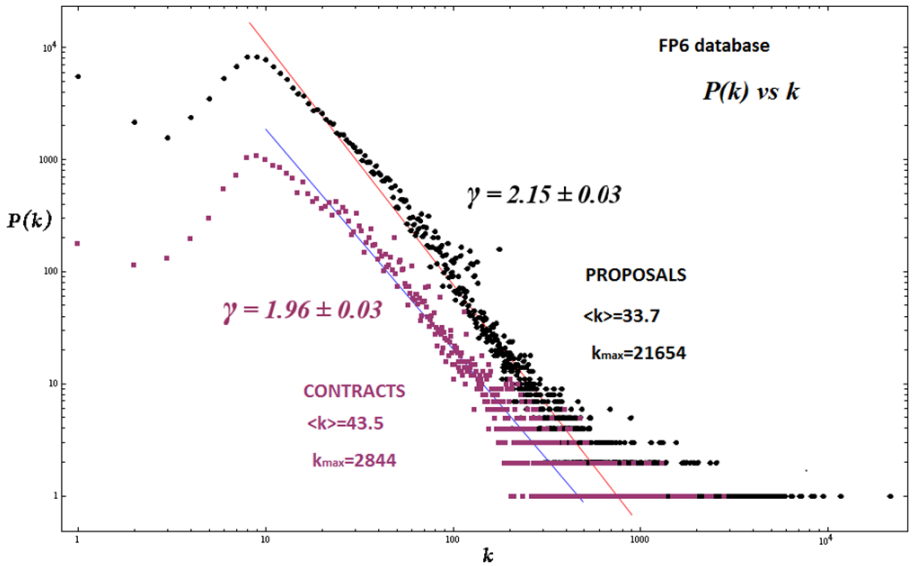
Some remarks:

- ✓ Both distributions have broad tails
- ✓ Broader tail for the distribution of the applicants network
- ✓ k_{max} larger for the distribution of the applicants network



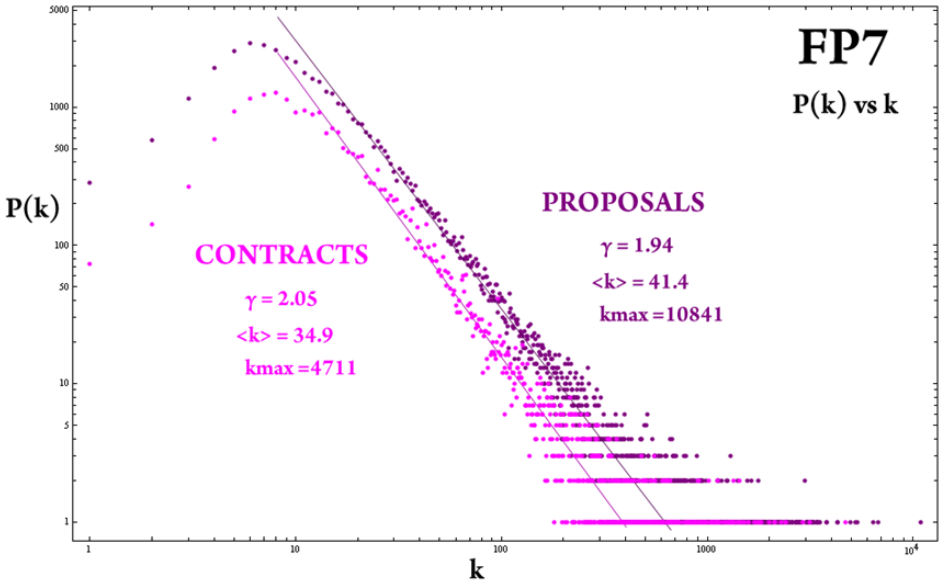
Some remarks:

- ✓ Exponent γ smaller for the distribution of the participants' network
- ✓ <k> larger for the participants network
- ✓ Participants network denser than that of the applicants



FP7 (level 1)
Probability distribution comparison for the participants and applicants* networks:

- ✓ The probability distribution of both networks exhibit power law behaviour



*Due to irrevocably missing data for this level, the probability distribution of the applicants network shown, was evaluated using the 5/6 of the records.

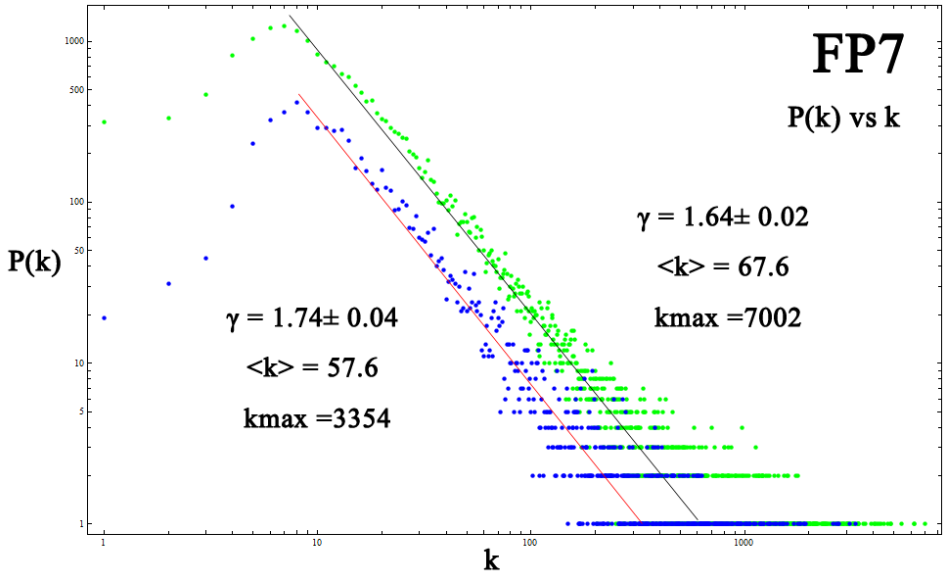
Level 2, cities:

For the FP6, FP7 projects and proposals datasets:

- ✓ Construct the network of cities involved in both datasets
- ✓ Compare the probability distribution of the two networks
- ✓ FP6: 6732 cities in projects, 19564 cities in proposals
- ✓ FP7: 7368 cities in projects, 23168 cities in proposals

FP7, cities (level 2) distribution comparison for the cities in the projects and proposals datasets:

- ✓ The probability distribution of both networks exhibit power law behavior
- ✓ Similar exponents γ



Level 3, NUTS 3*:

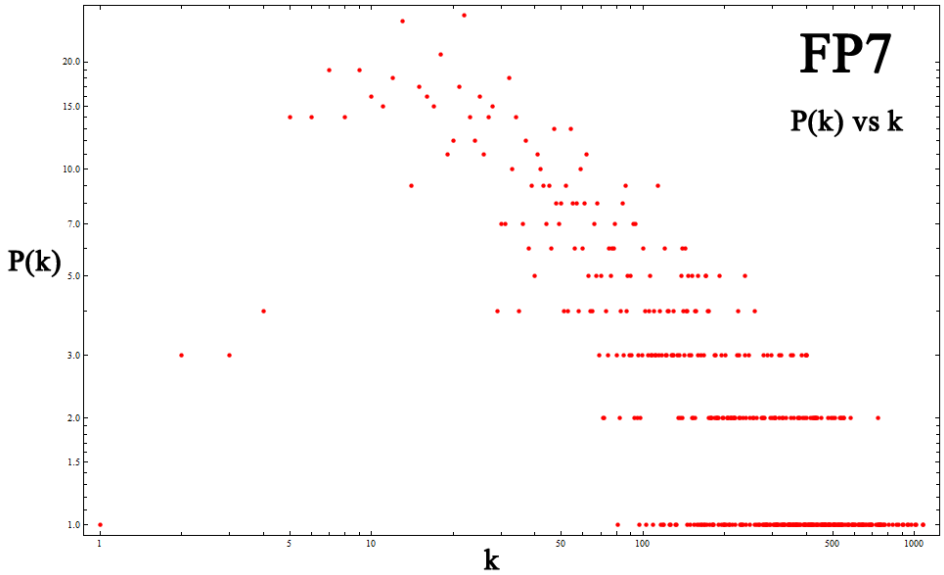
For the FP6, FP7 projects dataset:

- ✓ No NUTS field available for the proposals
- ✓ Construct the network of cities involved in the projects dataset
- ✓ Examine the probability distribution of the network
- ✓ FP6: 1318 NUTS
- ✓ FP7: 1436 NUTS

*Nomenclature of Territorial Units for Statistics, level 3

FP7-projects, NUTS (level 3)
Probability distribution of the NUTS network ():

- ✓ The probability distribution is skewed
- ✓ No power law behaviour



*Nomenclature of Territorial Units for Statistics, level 3

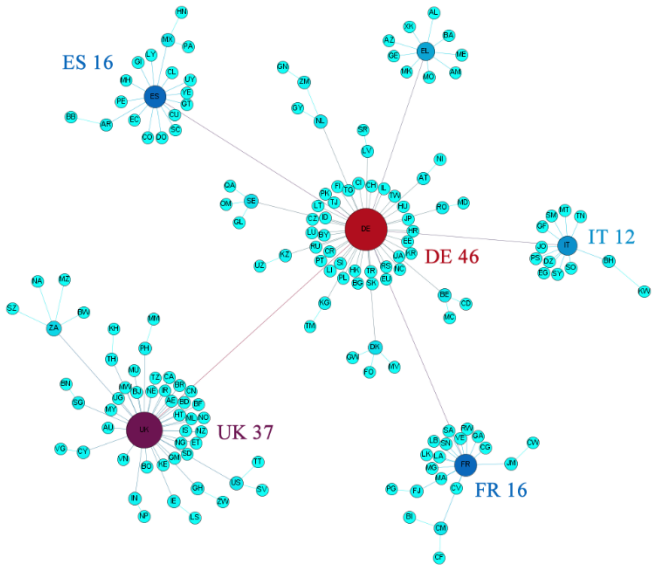
Level 4, countries:

For the FP6, FP7 projects and proposals datasets:

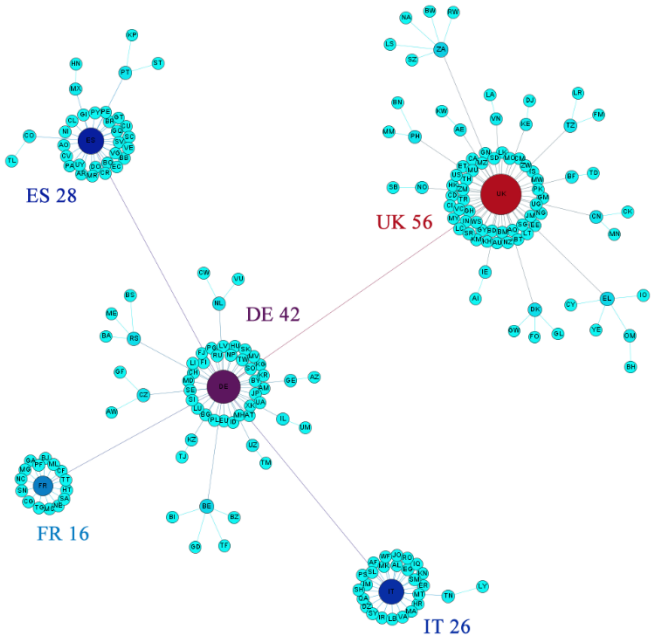
- ✓ Construct the network of countries involved in both datasets
- ✓ Find and compare the MSTs for both datasets
- ✓ FP6: 153 countries in projects, 212 countries in proposals
- ✓ FP7: 170 countries in projects, 210 countries in proposals

FP7, Minimum Spanning Tree for the country level network:

Projects



Proposals

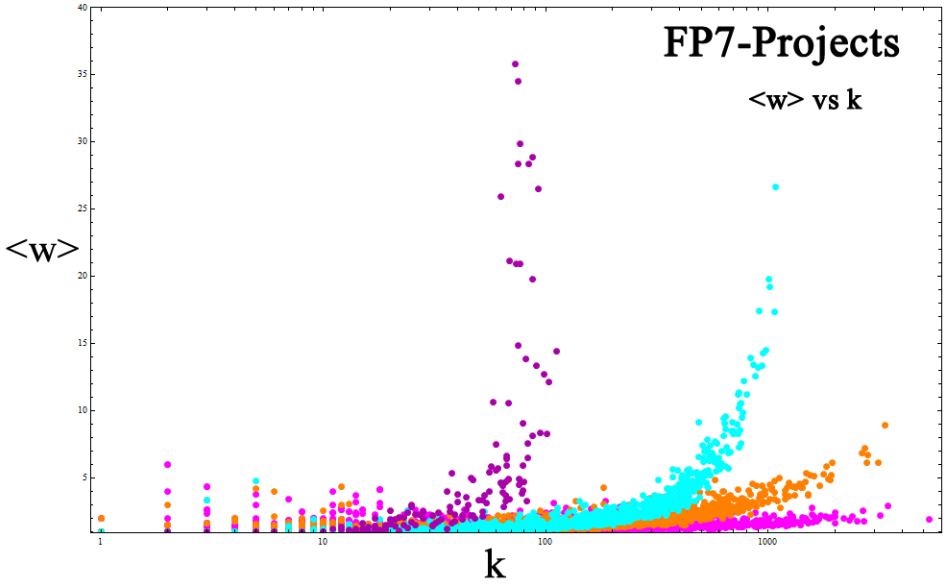


Some remarks:

- ✓ Differences in the rankings of the countries
- ✓ Strength of the collaborations between countries in the applicants network differ from those in the participants one

FP7 - projects <w> vs k for the networks of all available levels:

- ✓ The dependence of the mean node weight, <w>, on the node degree k varies for the 4 levels
- ✓ The denser the network, the bigger the dependence



Final remarks:

- ✓ The analysis of the data up to this point displays some interesting results
- ✓ Further study is needed
- ✓ Focus on the country level
- ✓ Ranking estimation with other methods, such as k-shell decomposition, centralities, etc.
- ✓ PhD work of Maria Tsouchnika, supported by Multiplex