The Rise & Fall of an Online Project. Is Bureaucracy Killing Efficiency in Open Knowledge Production?

Nicolas Jullien, iSchool, ICI-M@rsouin, Télécom Bretagne, Nicolas.Jullien@telecom-bretagne.eu

Kevin Crowston, School of Information Studies, Syracuse University, crowston@syr.edu

Felipe Ortega, University Rey Juan Carlos, jfelipe@libresoft.es

OpenSym'15, San Francisco

Research motivation

- Open online communities' goals
  - To produce valuable content from volunteers' contribution
- Recurrent but also growing concern about decreasing efficiency in doing so
  - Wikipedia: Halfaker et al. (2013), Ortega (2009)
  - FLOSS: Koch (2008)
- Decrease in recruitment...
  - See before & Crowston, Jullien & Ortega (2013)

Research questions

- Why do we see a decrease in turning the effort of volunteers into pieces of knowledge ?
  - Normal, project entering in a mature phase (Koch, 2008, FLOSS; Marwell & Oliver, 1993, any collective action)
  - Or over-administration making contribution less rewarding (Ransbotham & Kane, 2011, Wikipedia)?

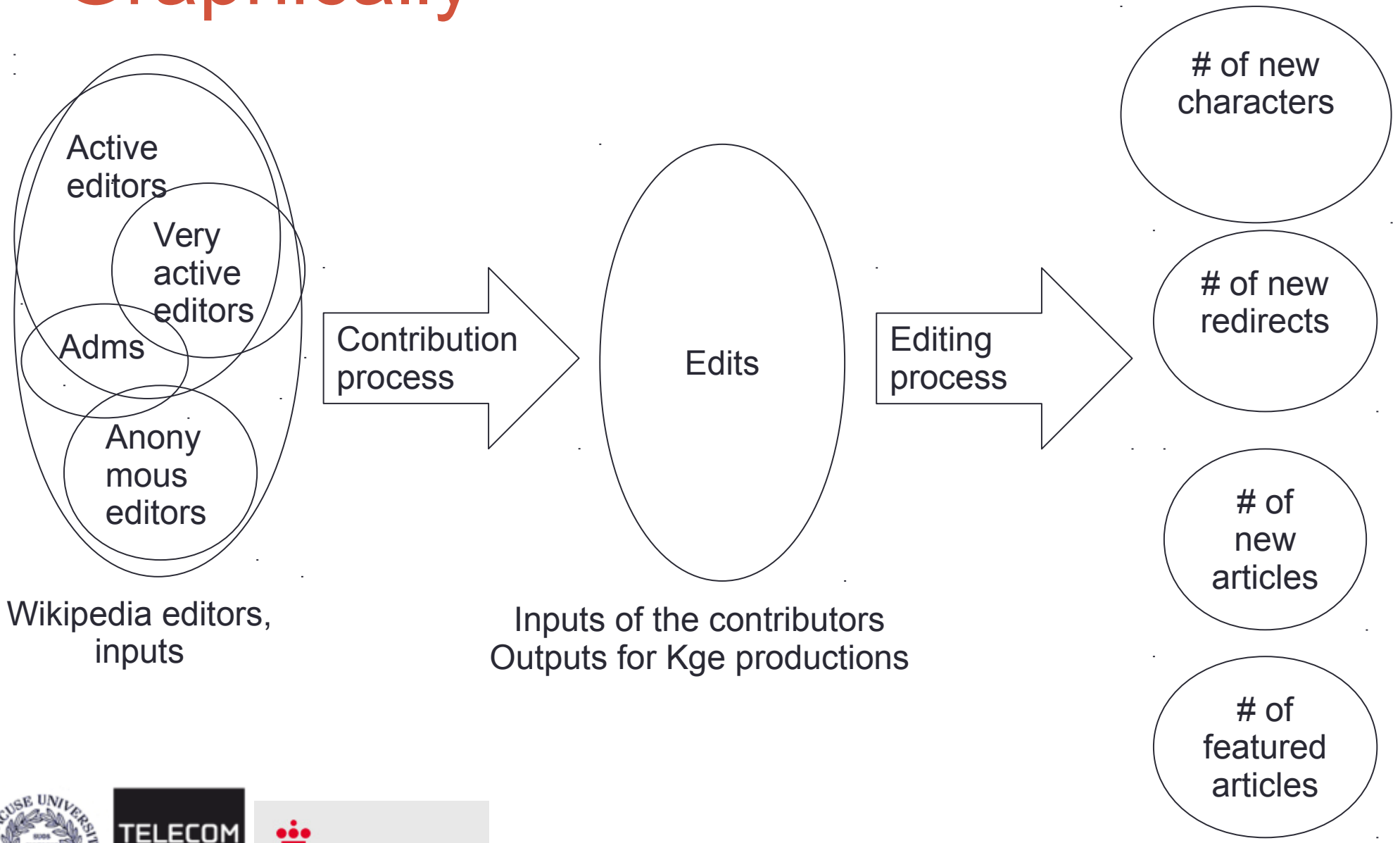
Main concepts and goal of the article

- Measurement of the efficiency of the projects:
  - Production function, as a link between inputs and outputs
  - Form and coefficients of this function unknown
  - We do not want to characterize the function but to compare different projects/organization
- Testing hypotheses to explain the decrease in efficiency, beyond the size
  - Comparison between (39) Wikipedia language projects
  - Same tools, same goal (writing articles)

Two questions, two sets of variables

- The turning of editors into edits, and edits into articles and articles of quality
- Inputs:
  - First model : the number of very active Wikipedians, active Wikipedians and other contributors + the number of existing articles and the number of existing links (size control variables) ;
  - Second model : the number of edits
- Outputs:
  - First model: the number of edits per month;
  - Second model: the number of new articles along with the number of new FA.

Graphically



Wikipedia editors, inputs

Inputs of the contributors
Outputs for Kge productions

Outputs of the project

Hypotheses

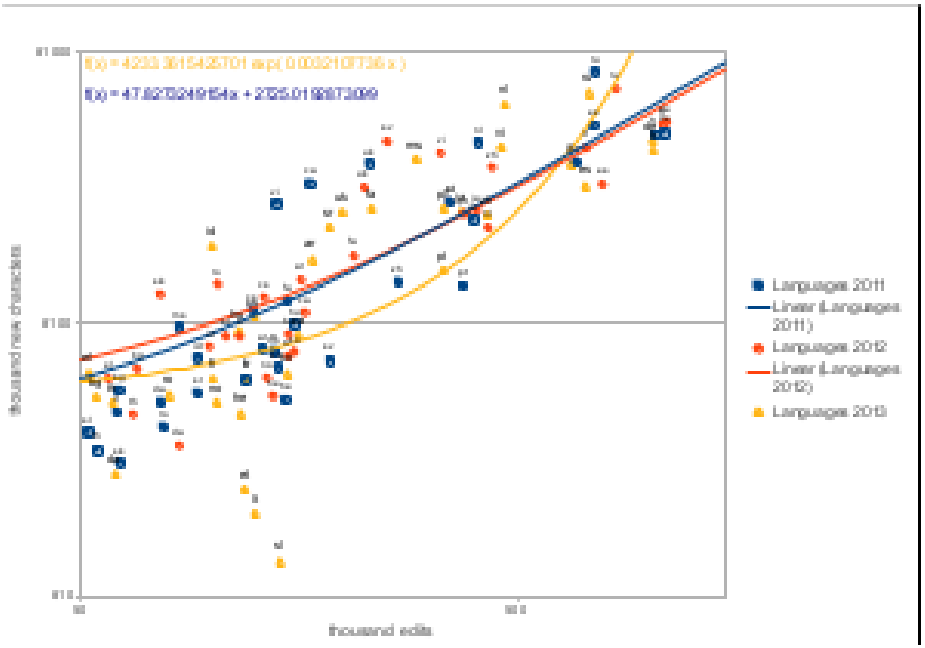
- H1. Big projects are less productive than small ones (i.e. projects exhibit decreasing return to scale), Marwell & Olliver
- H2. Structure of the team matters
  - H2.1. Following Uzzi, the efficient projects are heterogeneous, but not too much, regarding the variety of the participants, between big and small contributors,
  - H2.2 Following Hannan & Freeman (1984) on the tendency for any structure to become over-bureaucratic, we hypothesize that the efficient projects have neither a too heavy, nor too light an administrative structure.

Data collection

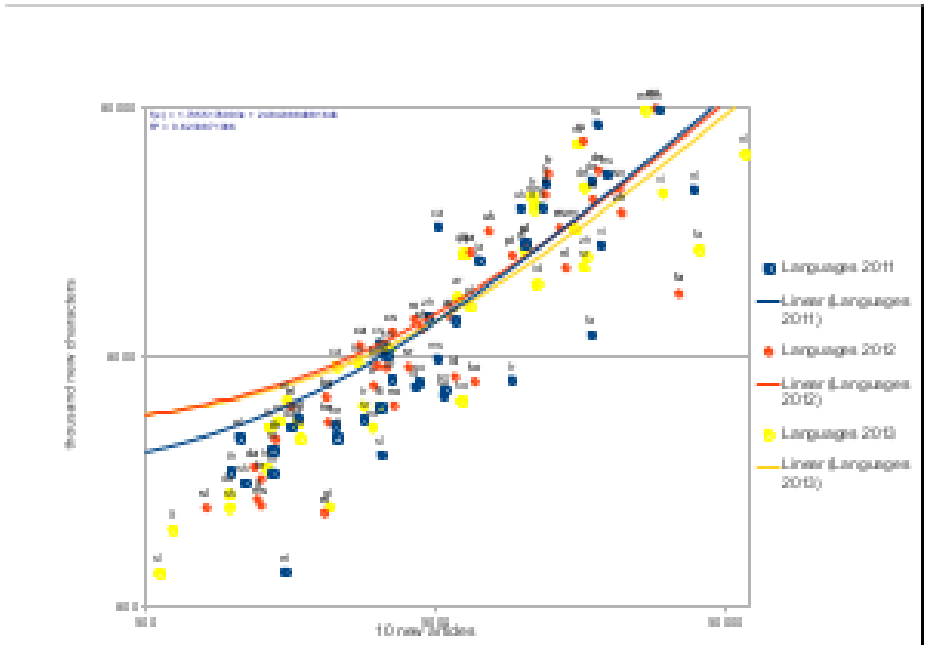
- Complete database dump with all edits performed in 39 Wikipedias in different languages
  - 3 years (2011 to 2013)
  - Cleaned
  - Via a software program in Python, part of WikiDAT (Wikipedia Data Analysis Toolkit)
  - More accurate & precise than Wikipedia's statistics (admins, FA...)

Initial analysis, size & production

Number of new characters versus number of edits, 2011, 2012, & 2013



Number of new articles versus number of new characters, 2011, 2012, & 2013



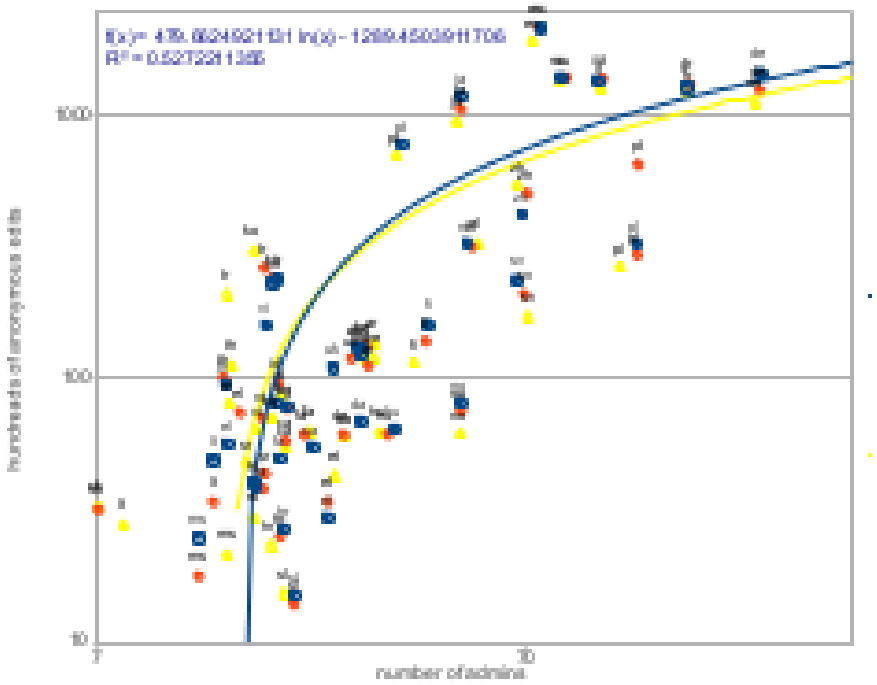
SYRACUSE UNIVERSITY FOUNDED A·D·1870

TELECOM Bretagne

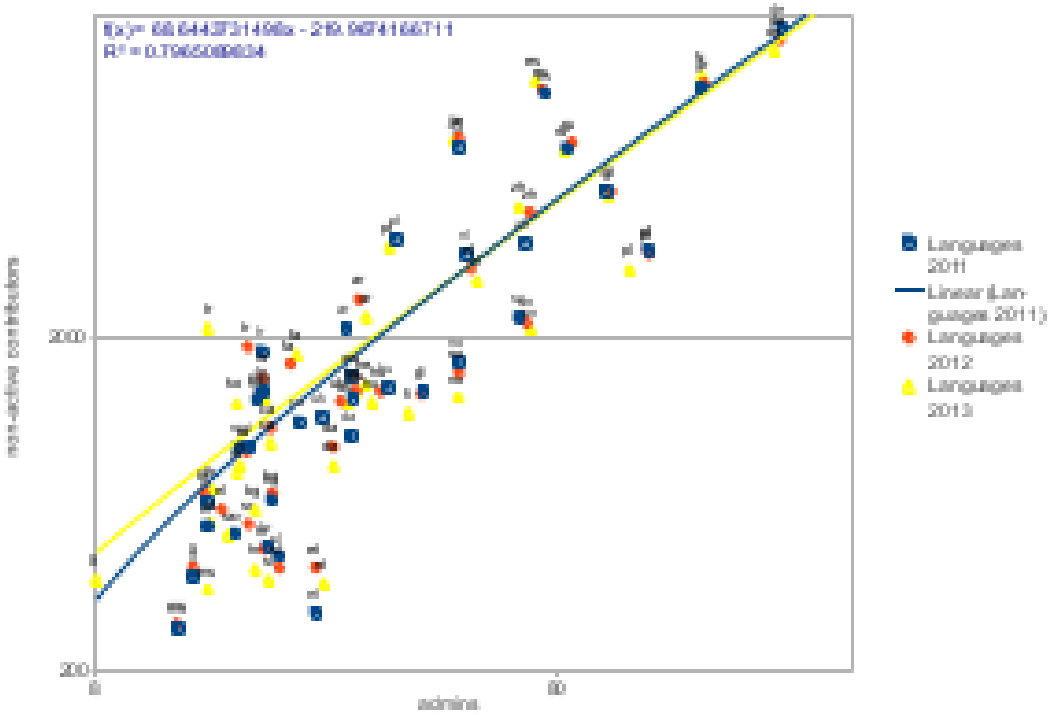
Universidad Rey Juan Carlos

Initial analysis, size & bureaucracy

Number of anonymous edits versus number of admins, 2011, 2012, & 2013



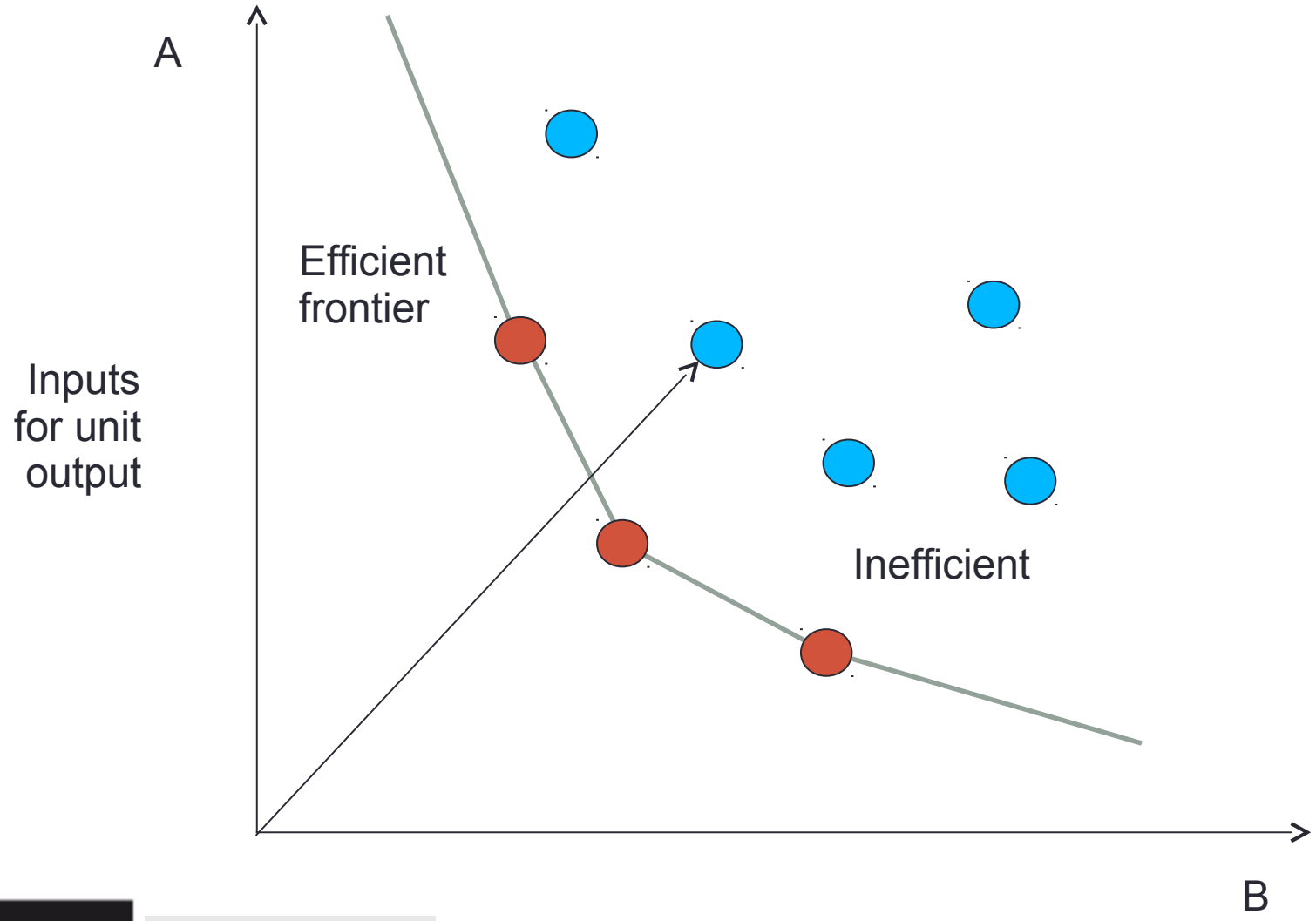
Number of low active editors versus number of admins, 2011, 2012, & 2013



Multiple inputs, multiple outputs comparison: DEA

- Conceptual tool:
  - “Frontier production function” (Farell, 1957)
  - Data Envelopment Analysis models (Charnes, Cooper & Rhodes, 1978), used by Koch 2009 for FLOSS
  - Taking into account the possible decrease of efficiency due to the size of the project (decreasing return to scale)

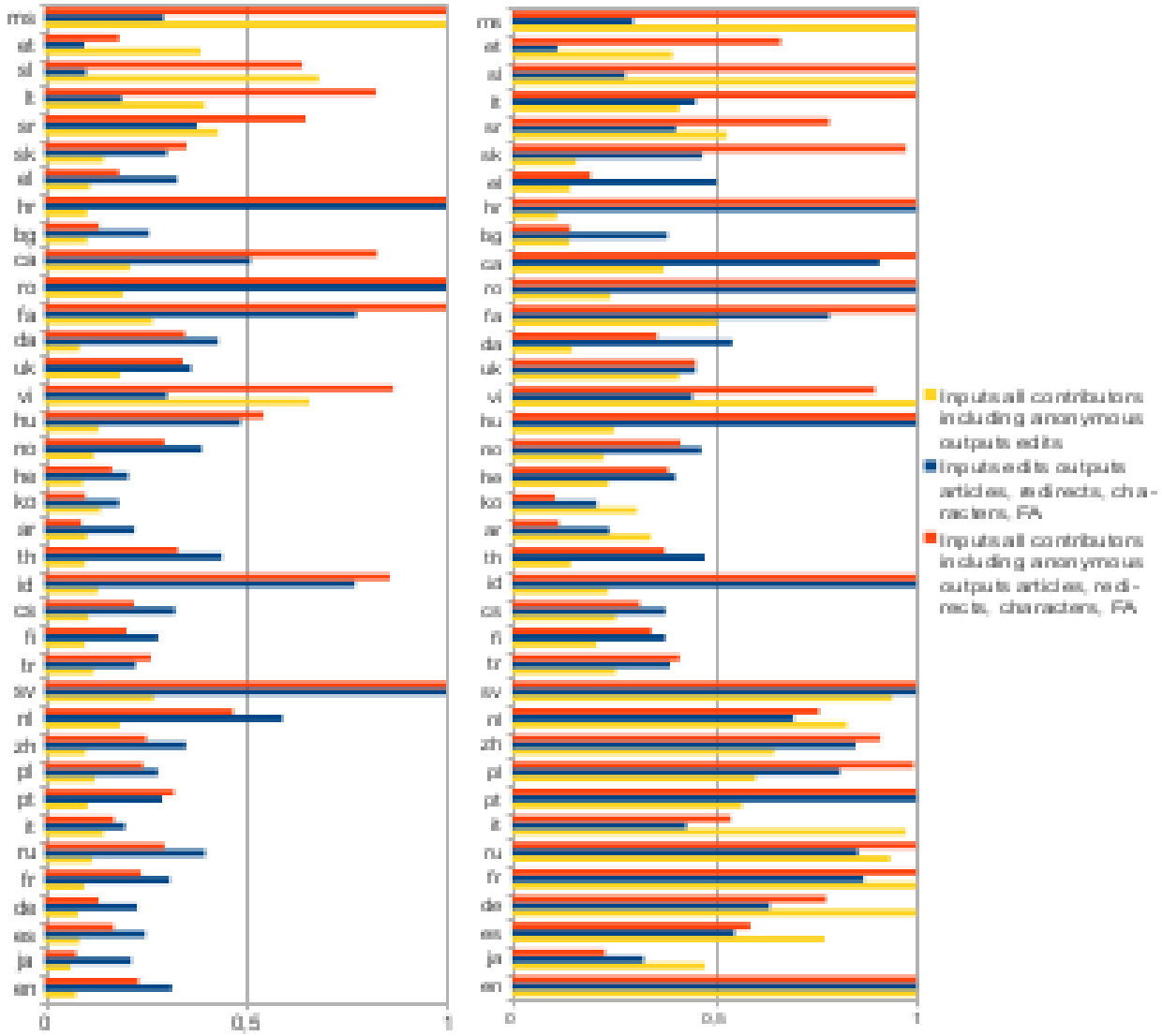
Data Envelopment Analysis graphically



H1. Size & efficiency in production of edits and new knowledge

for the 37 Wkipedia language projects, 2013

(left, without taking into account the return to scale, right taking into account the return to scale)



Results for Hypothesis 1

- Big projects are less efficient (decreasing return to scale)
  - Particularly true when looking at the conversion of contributors into edits
  - Not very sensitive to taking onto account the number of FA or the anonymous edits

H2. Structure of the projects and performance

- Linear regression on assessed efficiency in turning contributors into knowledge (measured by the DEA model)
- Explanatory variables:
  - ratio of administrators over anonymous edits, ratio of administrators over contributors (to test Hypothesis 2.2),
  - Ratio active and very active contributors over contributors (to test Hypothesis 2.1),
  - Hofstede's cultural dimensions, and whether or not the language project concerns more than one country (control variables).

Results on Hypothesis 2

- Only one statistically significant relation:
  - the link between the ratio of the number of administrators to anonymous edits and the efficiency of the projects
  - efficient projects are significantly more administrated

Discussion

- Comparison between projects different in size is possible (DEA)
- Big projects are in their decreasing return to scale phase, but quite efficient in controlling it
  - (and supposed lack of efficiency due to elements not measured? The rephrasing of an article, the adding of a picture, templates...)
- Some results are inconclusive (structure of the teams)
  - May be due to the similar structure of the teams in all the projects (Stand. Dev. Is low)

Ratios (%)	Means	Stand. Dev.	Min	Max
Active contrib. over contrib.	33	4.9	20	43
Very active contri. / contrib.	5.3	1.3	2.6	8.6

Limitations and future work

- Good data, but small data set (project x year)
  - More years are needed, especially the golden years, 2006 to 2008
- Measure of quality should be improved
  - (ideas?)
- Measure regarding the efficiency of the edits are disputable
  - We assumed that for any project the mean time to perform an edit was the same
  - (harder to perform an edit in a big project than in a small one?)
- We dropped robot contributions, is it relevant?
  - They are part of the process of production