What does it take to make a good CS conference?

Reverse-Engineering Conference Rankings

Peep Küngas, Svitlana Vakulenko, Marlon Dumas, Luciano Garcia-Banuelos, Cristhian Parra, Fabio Casati, Marju Valge, Svetlana Vorotnikova, and Karina Kisselite

LiquidPub

SEVENTH FRAMEWORK PROGRAMME

UNIVERSITY OF TRENTO - Italy

Springer

CSIC

Institut | Nicod
CNRS-EHESS-ENS

UNIVERSITAS FRIBURGENSIS

Background (1)

o Evaluating venues and research groups

- o ImpactFactor – a measure for ranking journals – one of the first ones
- o PageRank – a measure for ranking scientists
- o …
- o *ImpactFactor finds the popularity while PageRank score shows the prestige*

- o ranking research groups by their performance through bibliometric indicators
- o automated ranking of collections of articles, including conference proceedings based on analyzing citation networks (No empirical evaluation though)
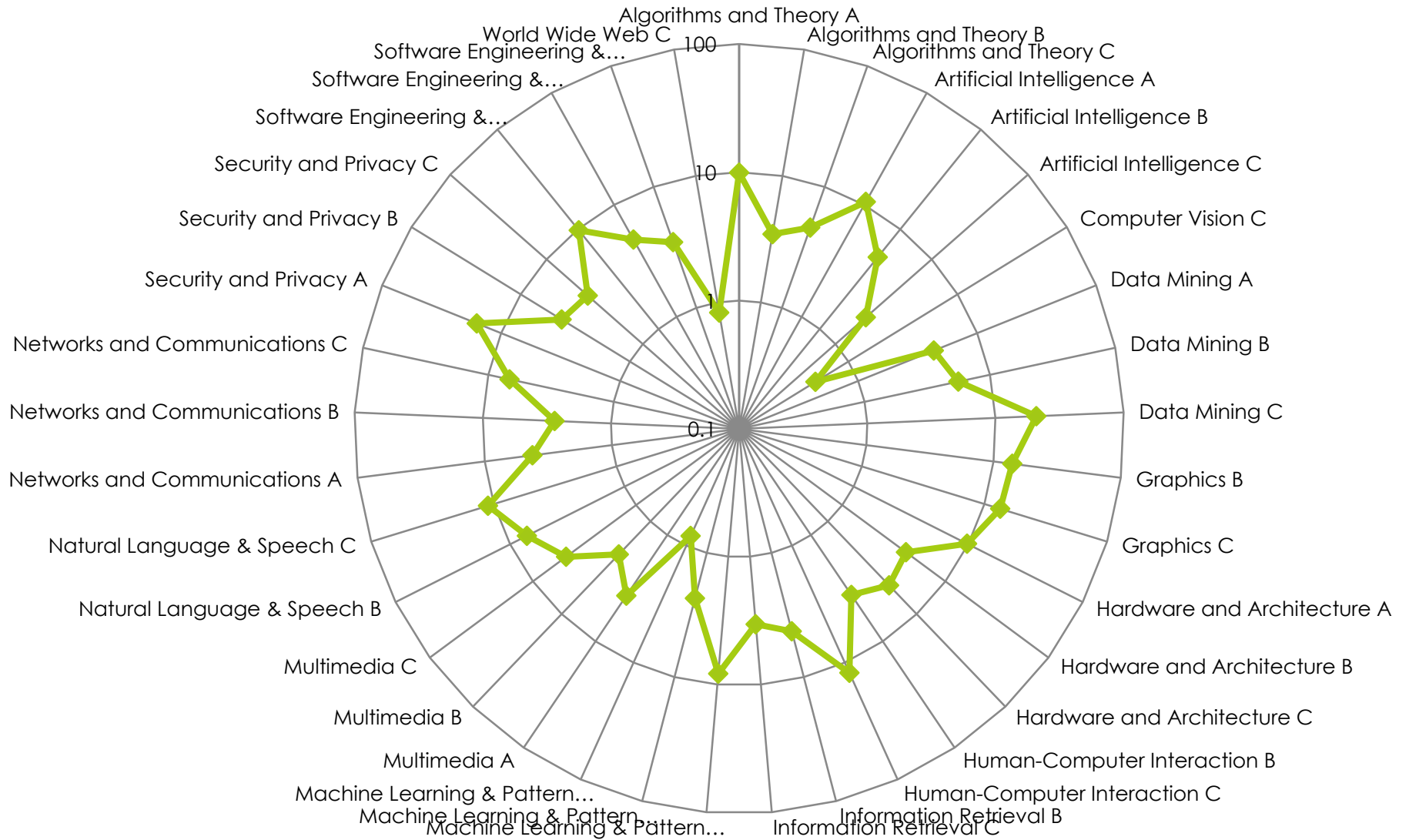
Background (2)

o Analyzing academic promotions of individuals

    o the number of published papers has generally small impact for reputation though it implies that a scholar is able to change jobs, and it also raises salaries

    o bibliometric indicators predict promotions of researchers better than random assignment (the best predictor for promotion being H-index followed by the number of published papers)

Background (3)

o Publication patterns

  o crossing-community, or bridging citation patters are high risk and high reward since such patterns are characteristic for both low and high impact papers

  o citation networks of recently published paper are trending toward more bridging and interdisciplinary forms. In the case of conferences it implies that more interdisciplinary conferences should have higher potential for high impact work

  o to maximize metrics such as H-index and G-index, the authors should focus to more mainstream research topics with respect to more revolutionary work

Research Questions

o RQ1 : Which function, composed of tangible conference indicators, correlates most with their perceived reputation?

o RQ2 : In which extent can be tangible conference indicators used to automatically determine conference rankings?

o Approach: we use a conference ranking as a metric for perceived reputation of a conference.

  o There are many rankings available ….

Radar chart showing values across categories: Algorithms and Theory A, Algorithms and Theory B, Algorithms and Theory C, Artificial Intelligence A, Artificial Intelligence B, Artificial Intelligence C, Computer Vision C, Data Mining A, Data Mining B, Data Mining C, Graphics B, Graphics C, Hardware and Architecture A, Hardware and Architecture B, Hardware and Architecture C, Human-Computer Interaction B, Human-Computer Interaction C, Information Retrieval B, Information Retrieval C, Machine Learning & Pattern…, Machine Learning & Pattern…, Machine Learning & Pattern…, Multimedia A, Multimedia B, Multimedia C, Natural Language & Speech B, Natural Language & Speech C, Networks and Communications A, Networks and Communications B, Networks and Communications C, Security and Privacy A, Security and Privacy B, Security and Privacy C, Software Engineering &…, Software Engineering &…, Software Engineering &…, World Wide Web C. Radial scale: 0.1, 1, 10, 100.

Data Sources (1)

o Acceptance ratios:

- o http://wwwhome.cs.utwente.nl/~apers/rates.html (database conferences - Peter Aper's Stats Page),
- o http://www.cs.wisc.edu/~markhill/AcceptanceRates_and_PCs.xls (Architecture conference stats (ISCA, Micro, HPCA, ASPLOS), see the Prichard, Scopel, Hill, Sohi, and Wood Excel File)
- o http://people.engr.ncsu.edu/txie/seconferences.htm (software engineering - Tao Xie's Stats Page)
- o http://www.cs.ucsb.edu/~almeroth/conf/stats/ (networking conferences - Kevin C. Almeroth's page)
- o http://web.cs.wpi.edu/~gogo/hive/AcceptanceRates/ (Graphics/Interaction/Vision conference stats - see Rob Lindeman's Stats Page)
- o http://faculty.cs.tamu.edu/guofei/sec_conf_stat.htm (Computer Security conference stats, see Guofei Gu's Computer Security Conference Ranking and Statistics Page)
- o http://www.adaptivebox.net/CILib/CICON_stat.html - (Acceptance Ratio statistics for Computational Intelligence & Related conferences)

Data Sources (2)

o **Bibliometric data:**

  o Microsoft Academic Search for 2511 Computer Science conferences

    o The number of papers published at a conference

    o The overall number of citations to conference papers

o **CS Conference rankings:**

  o ERA 2010 by an Australian national agency

  o Crank (Rank X) by Sourav S Bhowmick (?)(Singapore-MIT)

    o http://www3.ntu.edu.sg/home/assourav/crank.htm

    o http://dsl.serc.iisc.ernet.in/publications/CS_ConfRank.htm

  o How strongly these rankings are correlated?

    o Pearson 0,55 for correlation between ERA 2010 and Crank

Dataset Size and Rank Distribution

Rankings	Rank X overall	ERA 2010 overall	Rank X with acceptance ratios	ERA 2010 with acceptance ratios
1	65	137	31	58
2	113	117	36	19
3	150	66	9	6
4	199		17	
Total	**527**	**320**	**93**	**83**

Experiments

- Used ML methods for learning decision trees
    - Because we wanted human-interpretable models
    - Used the following features of conferences:
        - Average number of submissions over time,
        - Average number of accepted paper over time,
        - Average acceptance ratio over time
        - Published rankings (both ERA 2010 and Rank X)
        - Bibliomentric indicators (#papers, #citations, citation per paper)

- Data and findings available at http://math.ut.ee/~svitlanv/

Machine Learning Methods

o ZeroR

o IB1

o **J48**

o LADTree

o BFTree

o NaiveBayes

o NaiveBayesMultinominal

o NaiveBayesUpdateable

o OneR

o RandomForest

o **RandomTree**

Machine Learning Results

Dataset size	ERA 2010	Rank X
Acceptance rates	83	93
Bibliometrics	262	353
Combined	82	91

Weighted average f-measure	ERA 2010	Rank X
Acceptance rates	0,72 (random tree)	0,48 (random tree)
Bibliometrics	0,56 (J48)	0,48 (random tree)
Combined	0,75 (random tree)	0,55 (random tree)

Extracted Rules (1)

Rank A conferences:

- **Number of citations>=33710 and citations per article>=10.23**

- **Number of citations>4088 and citations per article>10.68**

- **Number of citations>5869 and citations per article>10.23**

- **Average conference acceptance ratio<0.23**

- **Citations per article>=0.76 and average conference acceptance ratio<0.25**

- **Number of citations>=5814.5**

- **Number of citations>=8338 and average conference acceptance ratio<0.32**

Extracted Rules (2)

Rank A conferences:

o **Number of citations>=5814.5 and number of articles<3161**

o **Number of citations>5760 and number of articles<=5048**

Rank B conferences:

o **Number of citations>=5890 and citations per article<10.23**
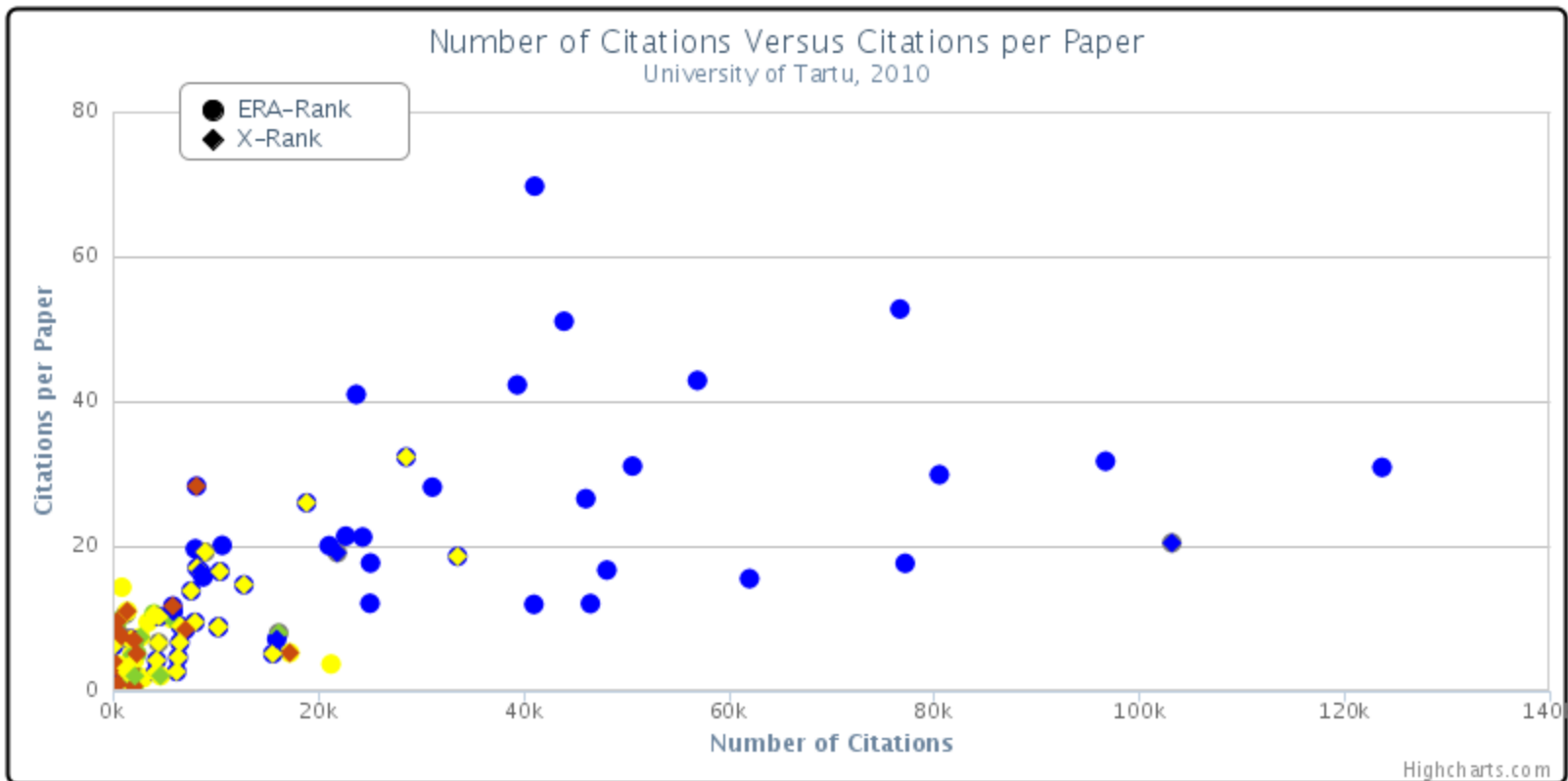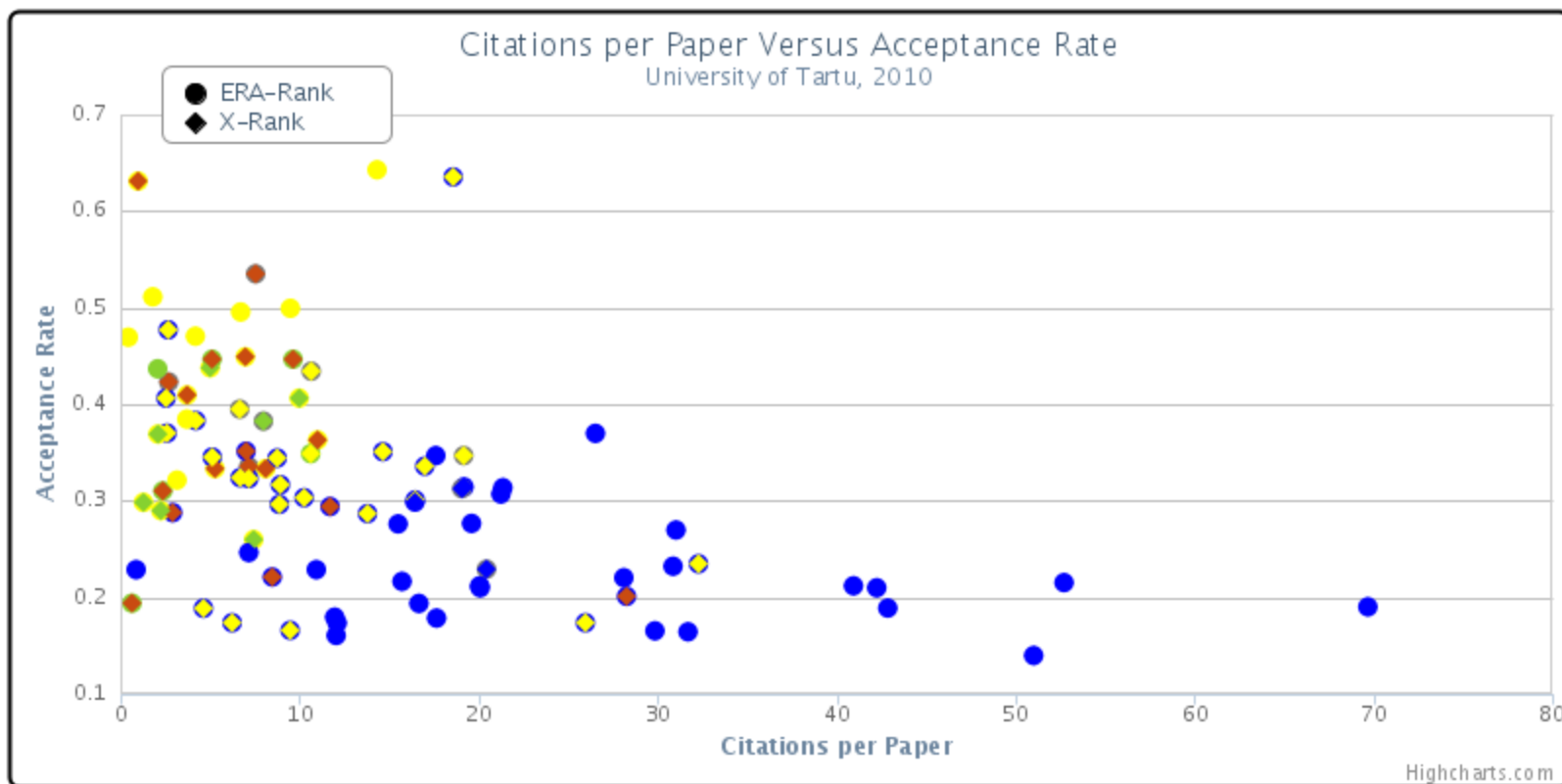
Legend of Figures

Blue – rank A
Yellow – rank B
Green – rank C
Red – rank D/unranked

Citations vs Citations per Paper

Citations per Paper vs Acceptance Ratio

Citations per Paper Versus Acceptance Rate
University of Tartu, 2010

Citations vs Acceptance Ratio

Number of Citations Versus Acceptance Rate
University of Tartu, 2010
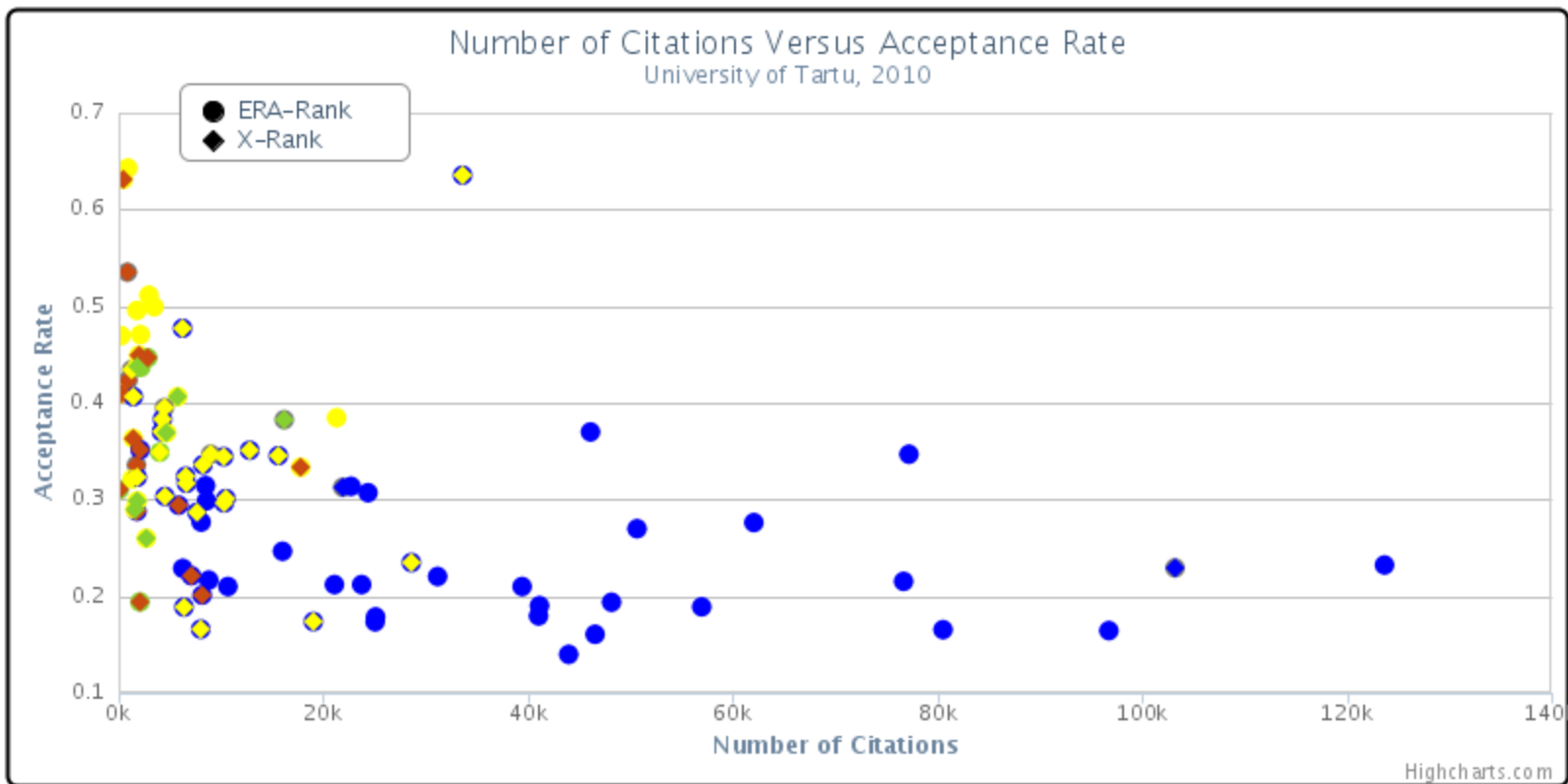
Conclusions

o   Accepance rate is generally the best predictor of conference reputation

o   However, combination of acceptance rates and bibliometric indicators (#citations, citations per paper) gives even better results

o   Our findings can be used to distinguish rank A conferences from conferences ranked B and C

o   There is no clear rule for distinguishing rank B conferences from rank C conferences when considering only acceptance ratios and bibliometric indicators

Thank You!