

SCIENTIFIC & FINANCIAL DATA: EXPLORING PID-BASED BRIDGES

LUC BORUTA & DAMIEN VANNSON
THUNKEN, INC.

PIDAPALOOZA 2018, GIRONA





AUTOMATED PATENT VALUATION

- ▶ Task: predict any patent's current market value
- ▶ Indicators: citations, number of claims, age, family size, etc.
- ▶ Requires cross-references with:
 - ▶ Publications (∞ sources)
 - ▶ Patent applications, grants, and assignments (USPTO, EPO)
 - ▶ Clinical trials (NIH, EMA)
 - ▶ Drug applications (FDA, EMA/EMCDDA)
 - ▶ Financial statements (IRS/SEC, ESMA/AMF/BaFin/...)



AUTOMATED PATENT VALUATION

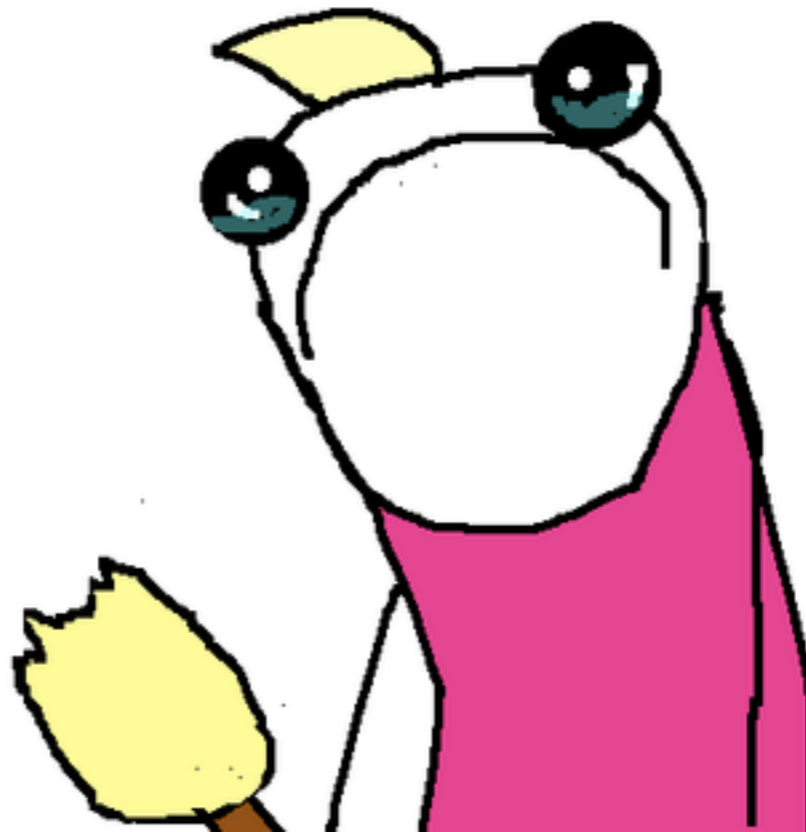
CRAWL ALL THE
X-REFS!





AUTOMATED PATENT VALUATION

No PIDS ?





When an inventor applies for a patent, the USPTO does not require that he or she record a unique identifier. As a result, searching for all the patents associated with a specific inventor can be difficult.

Official USPTO FAQ on [PatentsView.org](https://patentsview.org)



NO PIDS
NO PROJECT





#ALTMETRICSFORALL

- ▶ Automated valuation is ~~very~~ too ambitious
- ▶ Can we answer easier questions about IP portfolios?
 - ▶ Can we track innovations from research to market?
 - ▶ Who's talking about your (competitor's) IP?
- ▶ No PIDs, but tons of URLs and URNs
 - ▶ F. Lester, "Backlinks: Alternatives to the Citation Index for Determining Impact". doi:10.3998/3336451.0010.207



#ALTMETRICSFORALL

- ▶ <1% of the world population speak English as their first language
- ▶ Only two hypotheses for English-only projects:
 - ▶ All conversations about scholarly or legal content happen in English
 - ▶ English data is a perfectly-balanced sample
- ▶ Both are wrong



COBALTMETRICS.COM

- ▶ Monitor trusted sources to index citations, PIDs, and URLs
 - ▶ All IP-related documents, all fields, all languages
 - ▶ Started with data from Wikimedia and StackExchange
 - ▶ 45M documents and 63M citations/backlinks
- ▶ Goal: deliver normalized altmetrics data
 - ▶ We are **not** building a new metric
 - ▶ We keep track of usage stats when available



PID EXTRACTION/NORMALIZATION

- ▶ **doi:10.1351/GOLDBOOK.A00046**
- ▶ **https://doi.org/10.1351/goldbook.A00046**
- ▶ **{{Citation|doi=10.1351/goldbook.A00046}}**
- ▶ **{{GoldBookRef|file=A00046}}**
- ▶ **http://bit.ly/2DZINQT** 🤔



luc@thunken.com

www.thunken.com

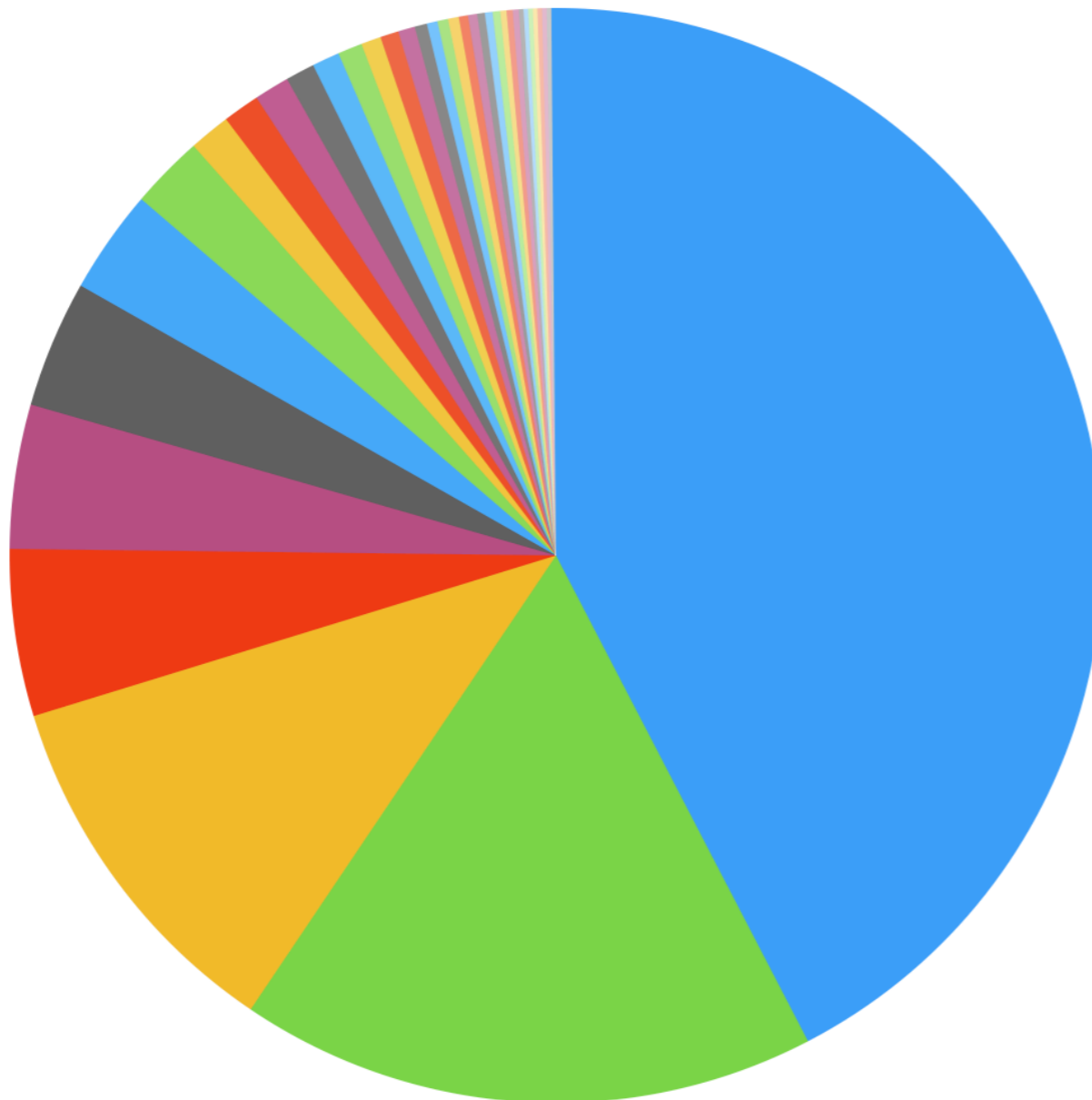
www.cobaltmetrics.com



AMBIGUITY

Which game(s) have you played the most?

3,994 responses



- Zelda
- The Legend of Zelda: Breath of the Wild
- Breath of the Wild
- BOTW
- Botw
- Breath of the wild
- BotW
- zelda
- Legend of Zelda: Breath of the Wild
- Legend of Zelda
- Zelda BOTW
- BoTW
- botw
- Zelda: Breath of the Wild
- Zelda BotW
- Zelda Breath of the Wild
- The Legend of Zelda
- Breath of The Wild
- The Legend of Zelda Breath of the Wild
- Zelda: BOTW
- Zelda: BotW
- Breath of the Wild
- Zelda breath of the wild
- Breath Of The Wild
- Legend of Zelda Breath of the Wild
- LoZ
- LoZ: BotW
- Zelda botw
- zelda botw
- breath of the wild
- Legend of zelda
- legend of zelda
- LoZ BOTW
- The Legend of Zelda: Breath of The Wild
- The legend of Zelda: breath of the wild
- ZELDA
- Zelda: BoTW



AMBIGUITY

Guidelines for the use and interpretation of assays for monitoring autophagy (3rd edition). *Autophagy*. 2016;12(1):1-222.

...Wang B, Wang C, Wang CY, Wang C, Wang C, Wang C, Wang D, Wang F, Wang F, Wang G, Wang HJ, Wang H, Wang HG, Wang H, Wang HD, Wang J, Wang J, Wang M, Wang MQ, Wang PY, Wang P, Wang RC, Wang S, Wang TF, Wang X, Wang XJ, Wang XW, Wang X, Wang X, Wang Y, Wang Y, Wang Y, Wang YJ, Wang Y, Wang Y, Wang YT, Wang Y, Wang ZN, ...

doi:10.1080/15548627.2015.1100356

pmid:26799652