

Distributed data collection

Matthew J. Salganik
Department of Sociology
Princeton University

Summer Institute in Computational Social Science
June 23, 2018

The Summer Institute in Computational Social Science is supported by grants from the Russell Sage Foundation and the Alfred P. Sloan Foundation.



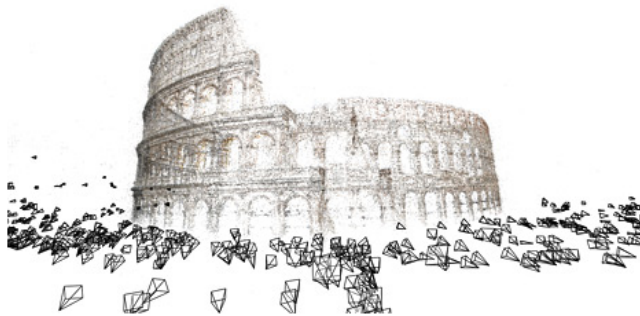
- ▶ people can be where the researchers can't

- ▶ people can be where the researchers can't
- ▶ scale that researcher cannot match

- ▶ people can be where the researchers can't
- ▶ scale that researcher cannot match
- ▶ sometimes hard to separate from human computation



Tuite et al. (2011) "PhotoCity: Training Experts at Large-scale Image Acquisition Through a Competitive Game" *CHI*: <http://dx.doi.org/10.1145/1978942.1979146>



Rome in a Day (Agarwal et al., 2009)



Two campuses: University of Washington and Cornell University

Over 2 months, 100,000 photos submitted by 45 players



(a) Lewis Hall (UW)



(b) Sage Chapel (Cornell)



(c) Uris Library (Cornell)

Beautiful design solves lots of problems

Beautiful design solves lots of problems

- ▶ data collection is standardized because of cameras

Beautiful design solves lots of problems

- ▶ data collection is standardized because of cameras
- ▶ verification is automatic by comparison with nearby images

Beautiful design solves lots of problems

- ▶ data collection is standardized because of cameras
- ▶ verification is automatic by comparison with nearby images
- ▶ game points are assigned based on the value of data, trains people to collect more valuable data

Questions?