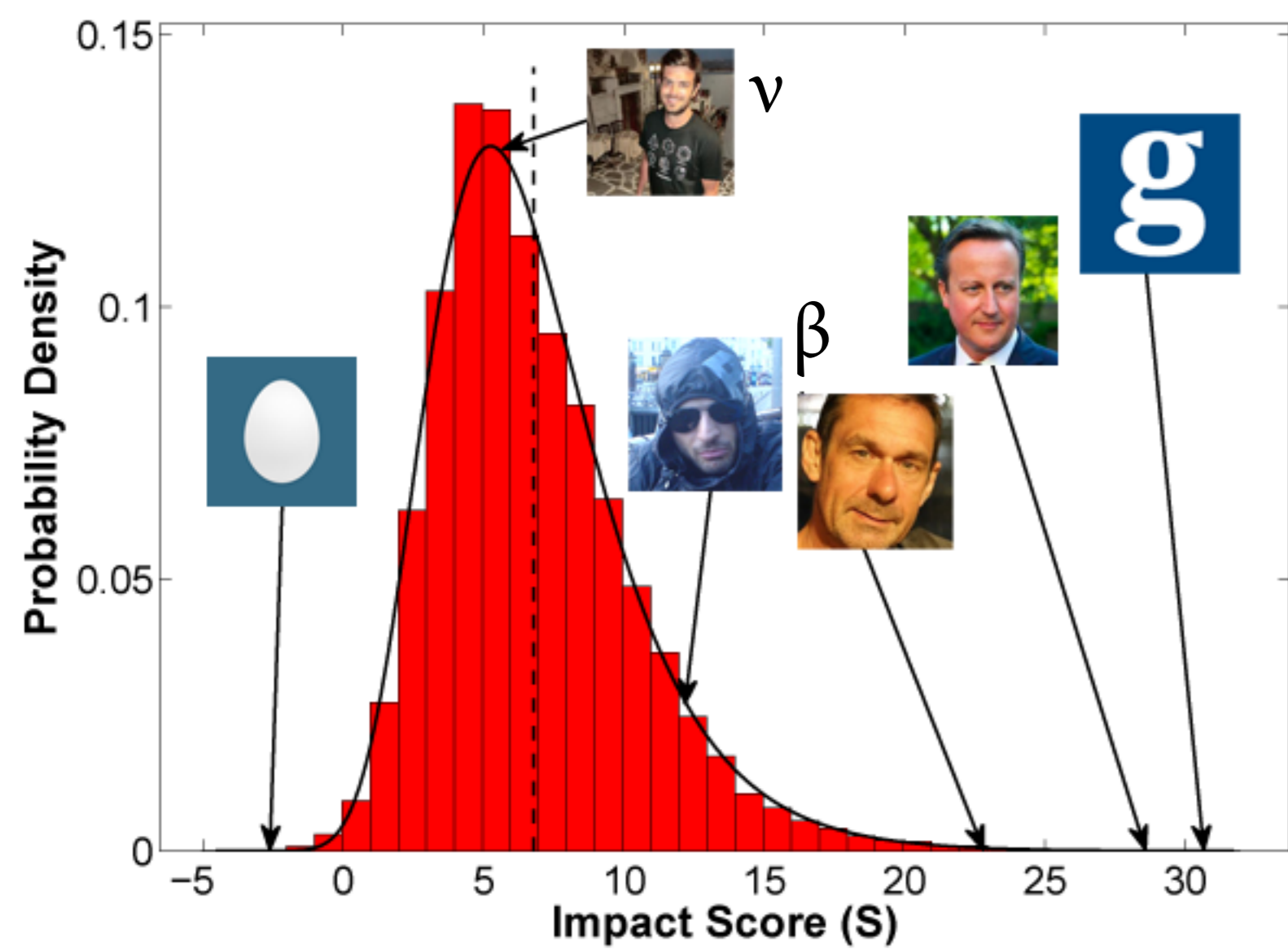


VASILEIOS LAMPOS^β, NIKOLAOS ALETRAS^ν, DANIEL PREOȚIUC-PIETRO^δ, TREVOR COHN^τ
 V.LAMPOS@UCL.AC.UK, {N.ALETRAS,DANIEL}@DCS.SHEF.AC.UK, TREVOR.COHN@GMAIL.COM

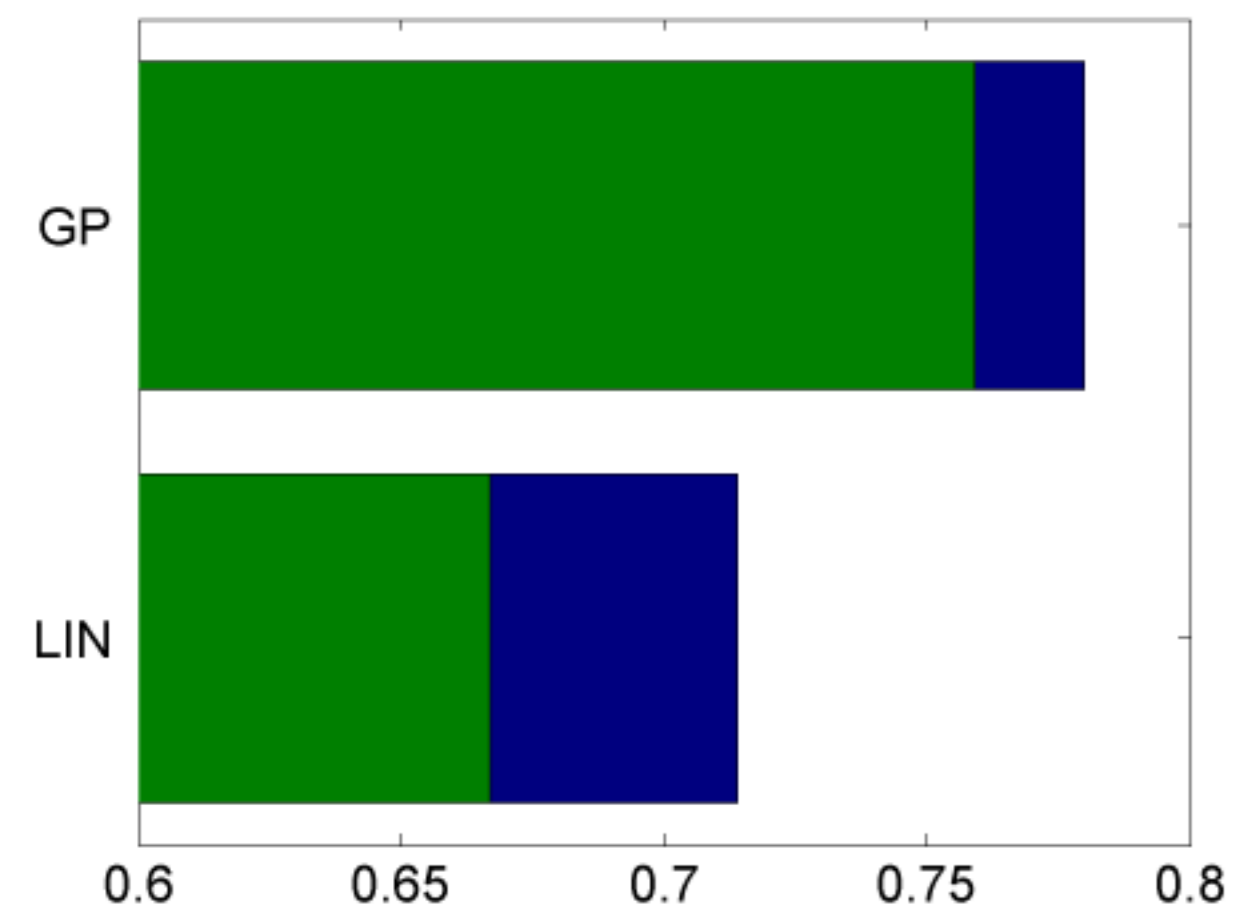
IMPACT DEFINITION

$$\ln \left(\frac{\text{listings} \cdot \text{followers}^2}{\text{followees}} \right)$$



IMPACT PREDICTION

Task: Predict user impact based on features under the user's control
Ridge regression (LIN)
Non-linear methods (GP)
 Gaussian Processes with ARD kernel

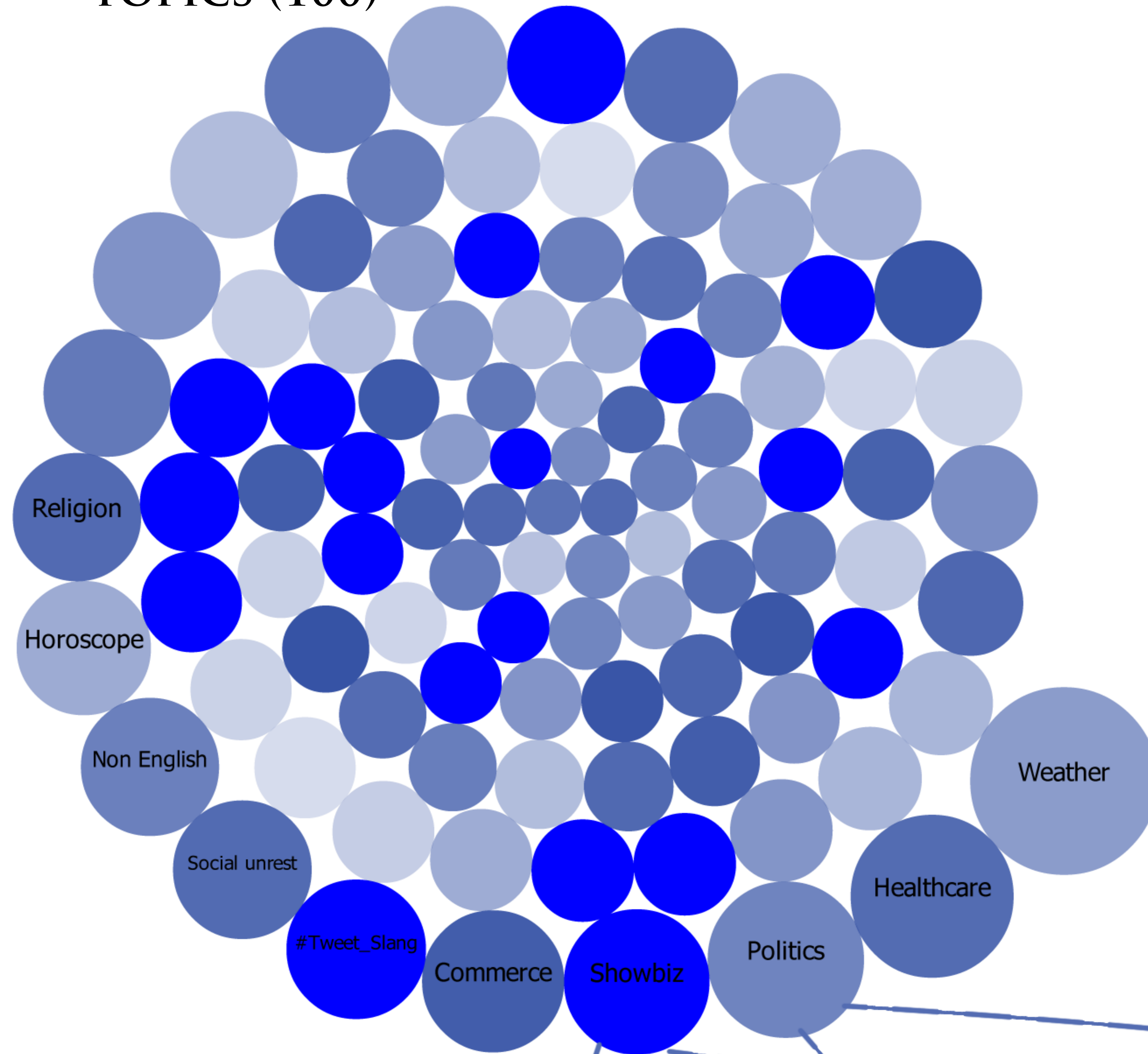


Green represents correlation using profile features, blue is the gain adding topics

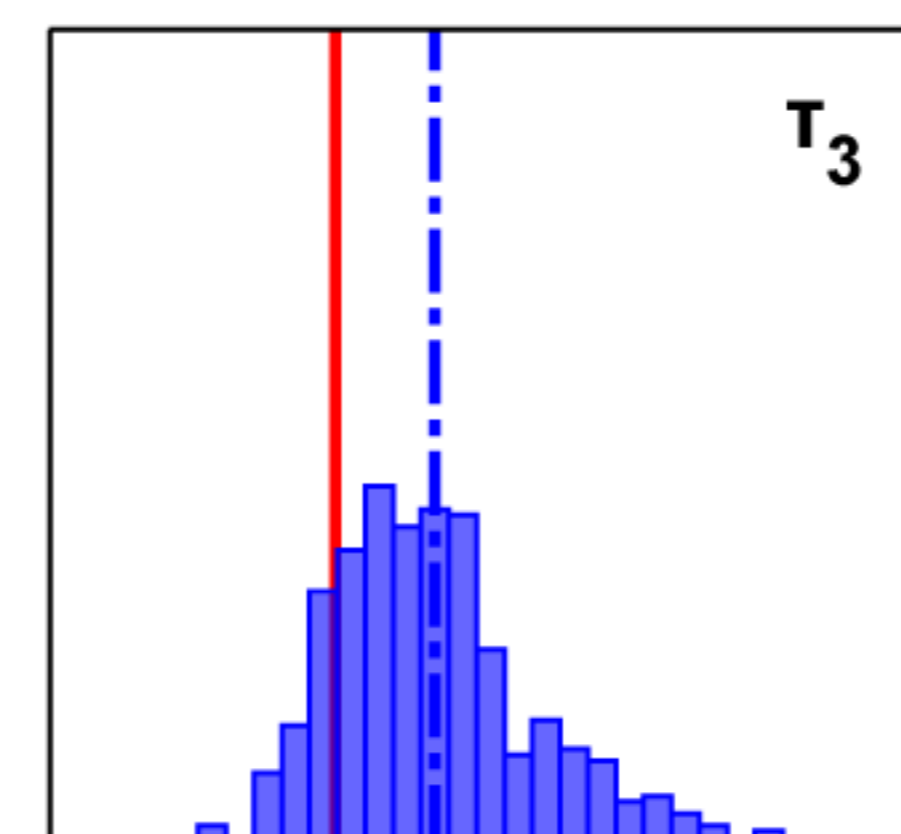
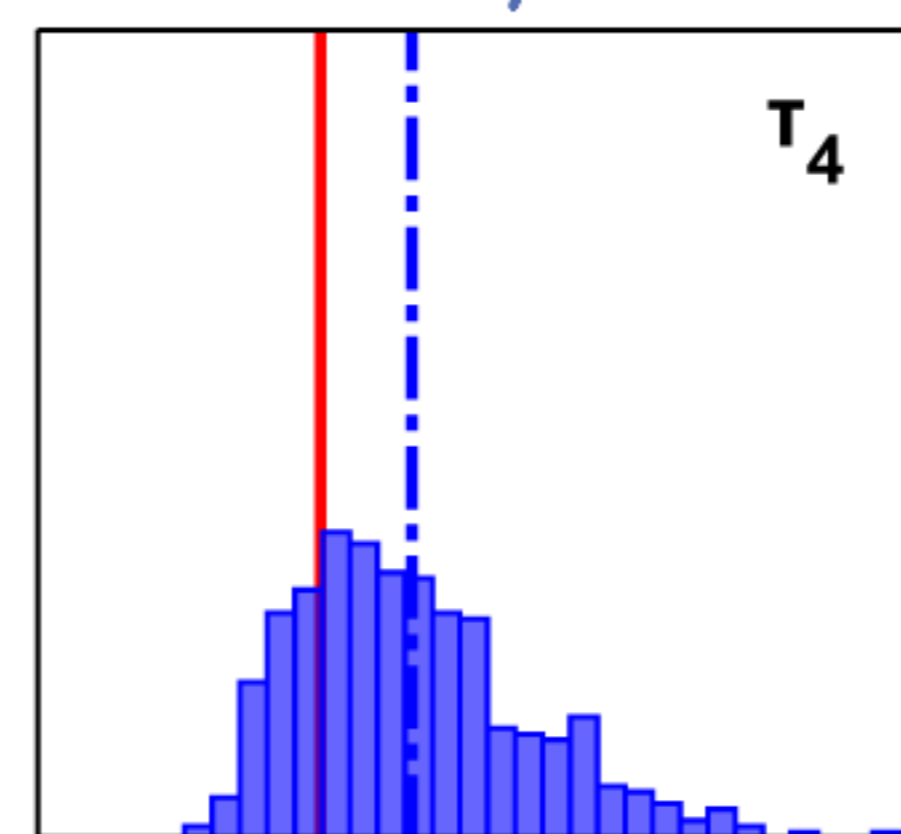
DATA

38,020 UK located Twitter users from 14.04.2011-12.04.2012
 ~48 mil deduplicated tweets

TOPICS (100)*



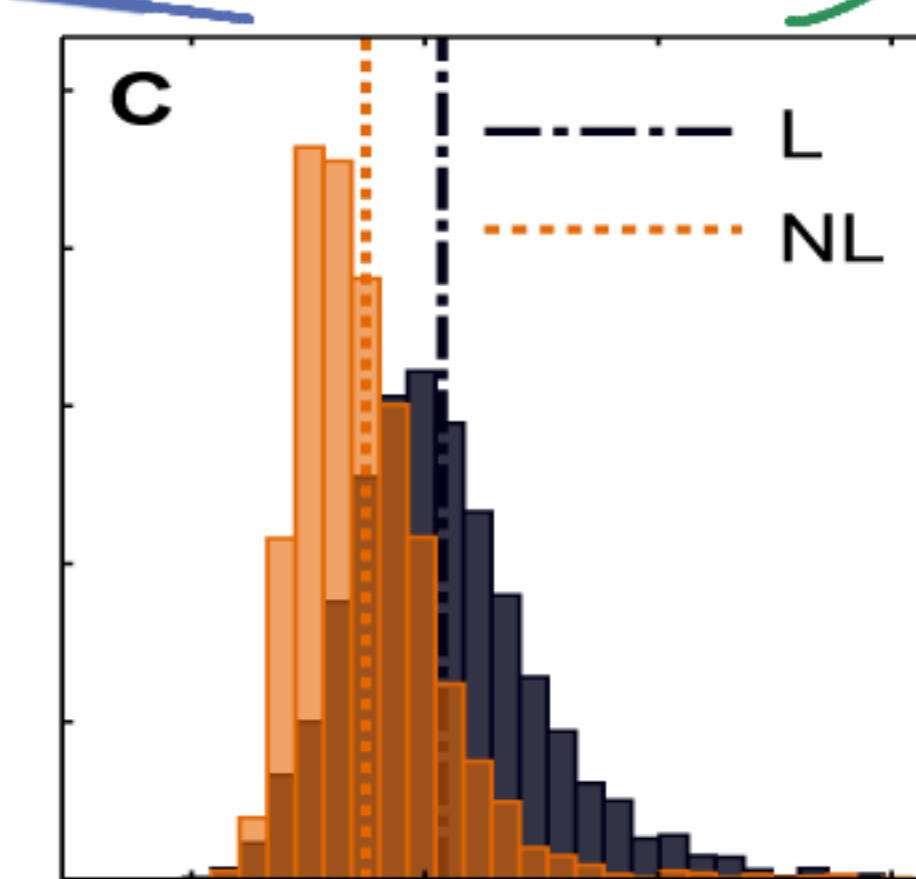
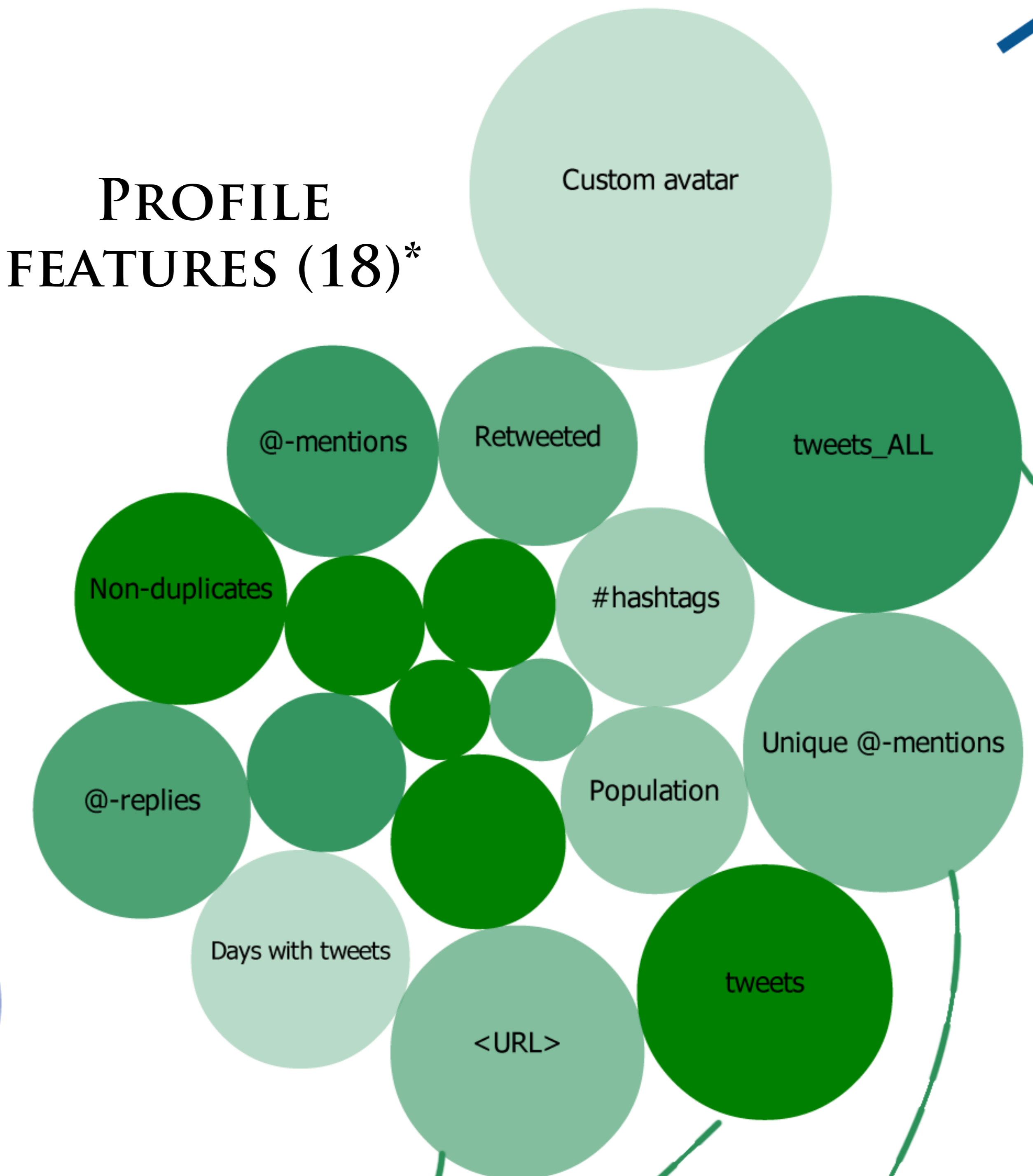
damon, potter
 #tvd, harry
 elena, kate
 portman, pattinson
 hermione, jennifer



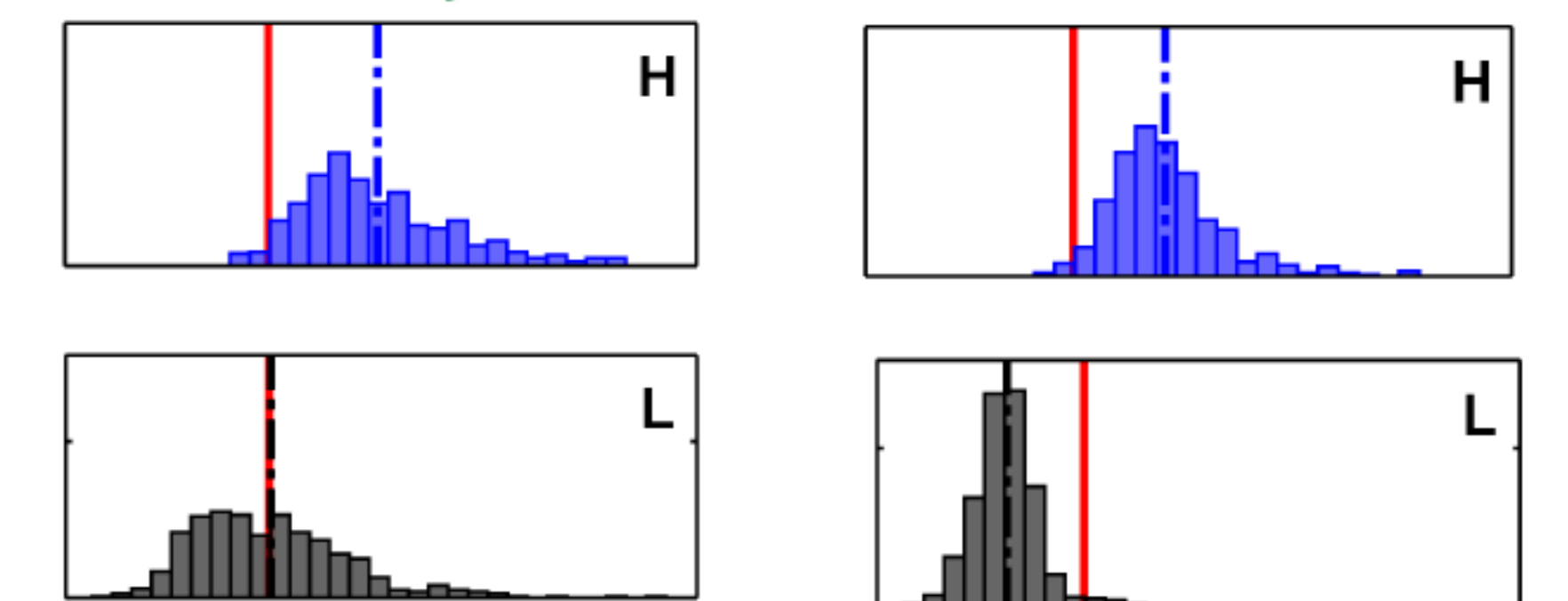
Topics computed over reference Twitter corpus
 Spectral clustering with NPMI as similarity measure

* bubble size inverse proportional to learned GP ARD kernel lengthscales and represent predictive relevance; colours split between the types of features; nuances of the same colour are only for visual effect

PROFILE FEATURES (18)*



Tweet a lot about Showbiz and Politics, with (L) or without (NL) using <URL>'s



Impact distribution for users with **high (H)** values of this feature as opposed to **low (L)**. Red line is the **mean impact** score.

