The Wisdom of Social Multimedia:
Using Flickr For Prediction and Forecast

Liangliang Cao,
Andrew Gallagher,
Jiawei Han, Xin Jin,
Jiebo Luo

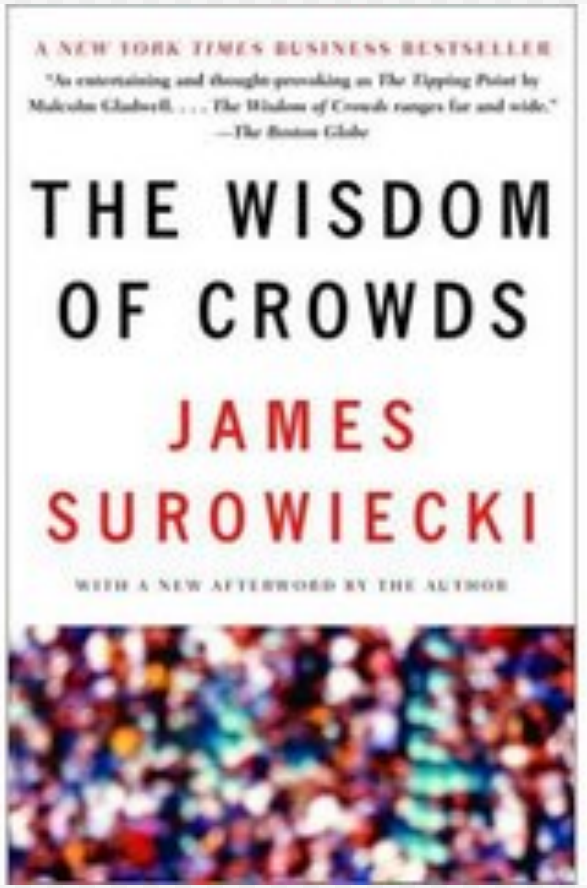
Slides by Tony Gaskell

Outline

- Criteria
- Flickr: Meta-information
- Querying Flickr
- Prediction models
- Experiments
- Evaluation

Criteria of a ‘wise’ crowd

- Diversity
- Independence
- Decentralization
- Aggregation



Criteria of a ‘wise’ crowd

- ✓ Diversity
- ✓ Independence
- ✓ Decentralization
- ✓ Aggregation

Flickr: Meta-Information

- T, set of all terms
- U, set of all users
- I, set of all images
- D, set of dates (days)
- M, set of dates (month)
- Q, set of dates (quarter)
- Y, set of dates (year)

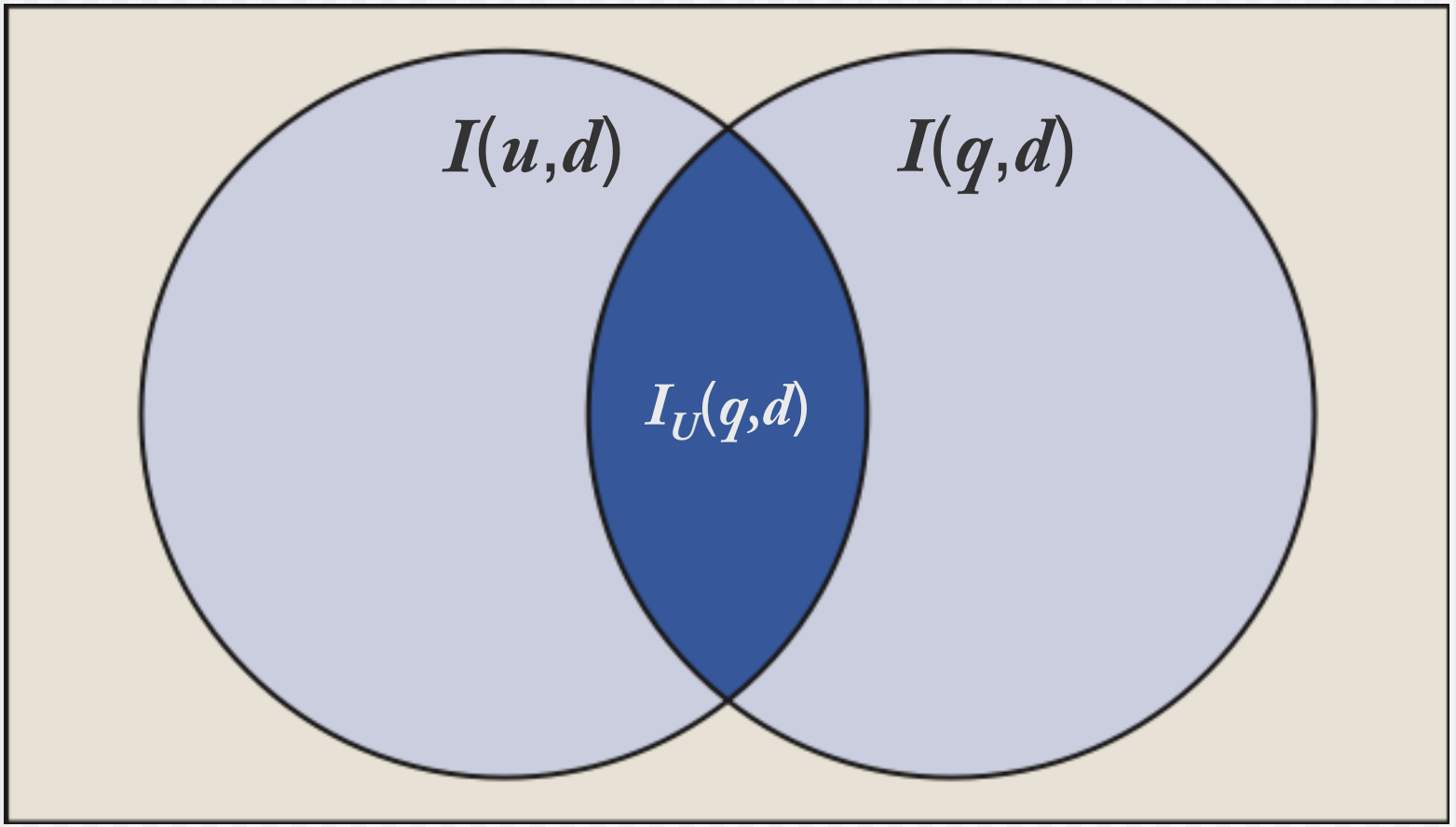
Flickr: Meta-Information (cont.)

- $T(i)$, the set of tag annotations for image i.
- $D(i)$, the day image i was taken.
- $M(i)$, the month image i was taken.
- $Q(i)$, the quarter image i was taken.
- $Y(i)$, the year image i was taken.
- $I(u)$, the set of images uploaded by user u.
- $I(q)$, the set of relevant images to query q.

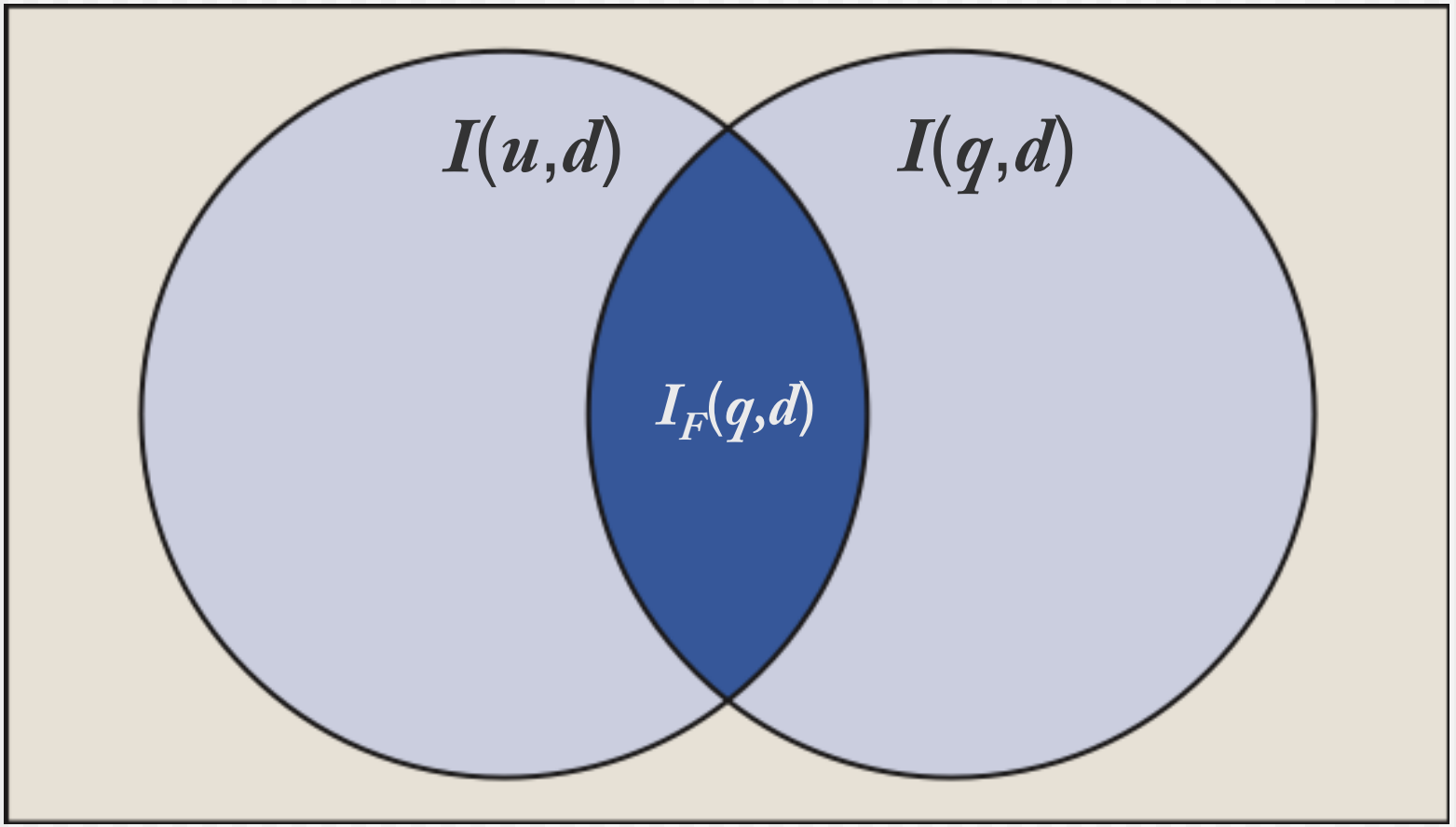
IPD and TIPD

- Images per day (IPD)
  - Looks at the number of relevant images that match a query on a certain day day.
- Tagged images per day (TIPD)
  - Images may or may not be tagged, we need to differentiate the two.

Unique users per day (UPD)



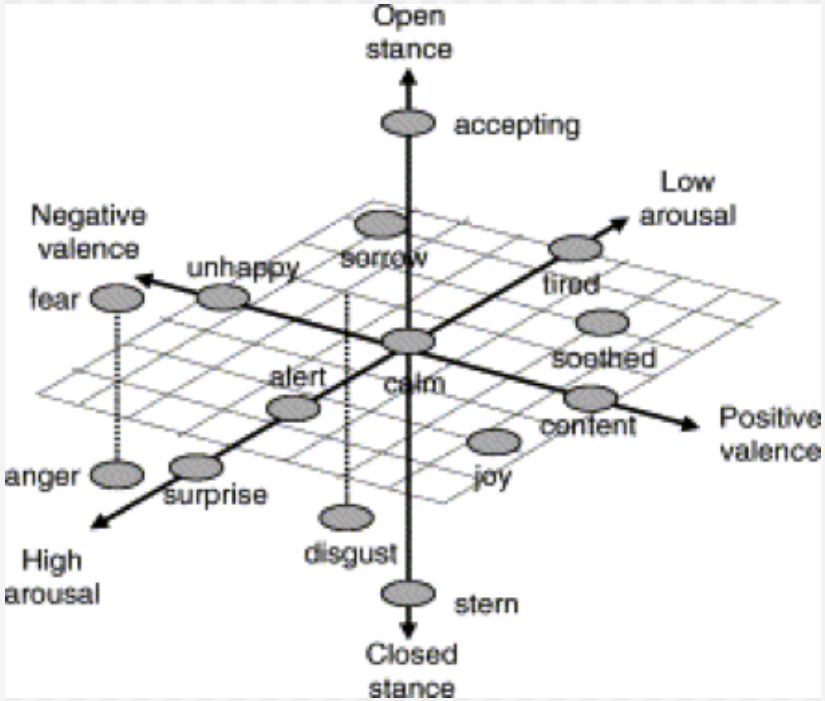
First- time Unique users per day (FUPD)



$D_{first}(u,q)$ represents the first day a user uploaded an image that matched a certain query.

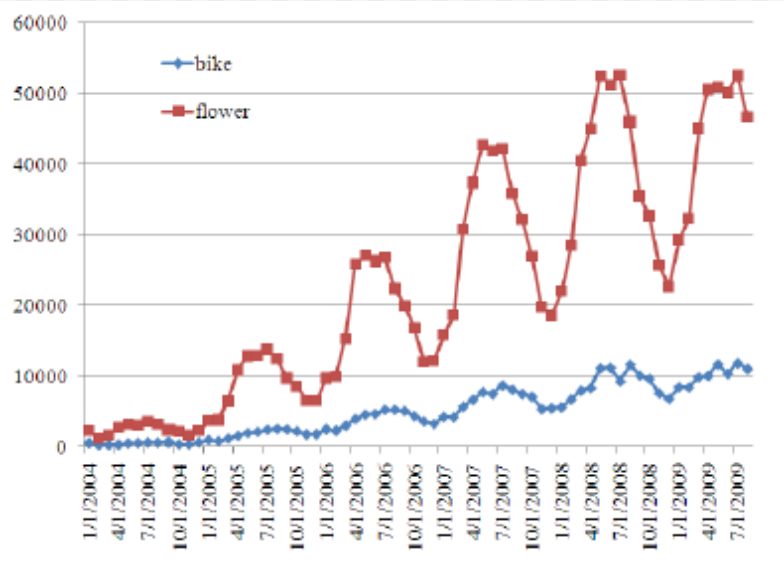
Image Visual Relevance

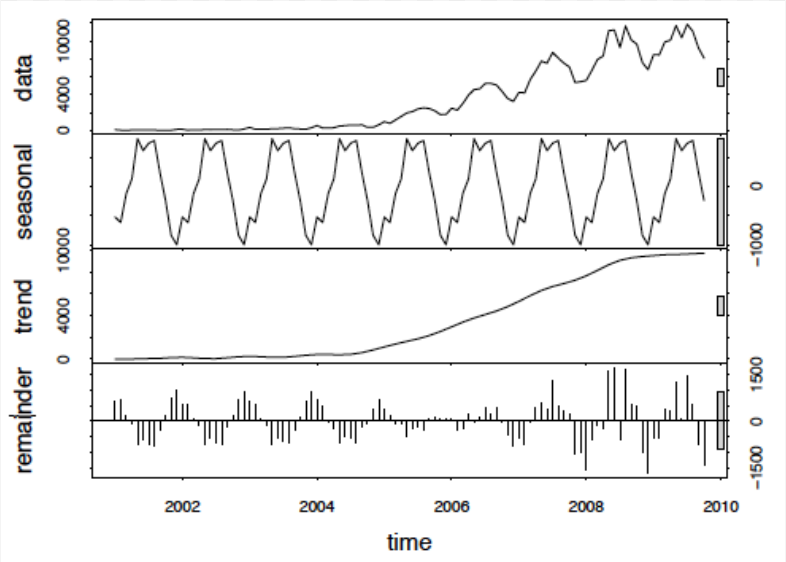
- Color / edge histograms
- Texture
- Color Correlogram
- Shape
- N-dimensional feature space



Quantifying emotions

Flickr Background Model





- STL decomposition

  Separates raw data into seasonal, trend, and remaining data.

F_B = Average trend of all general Flickr queries

Prediction Models

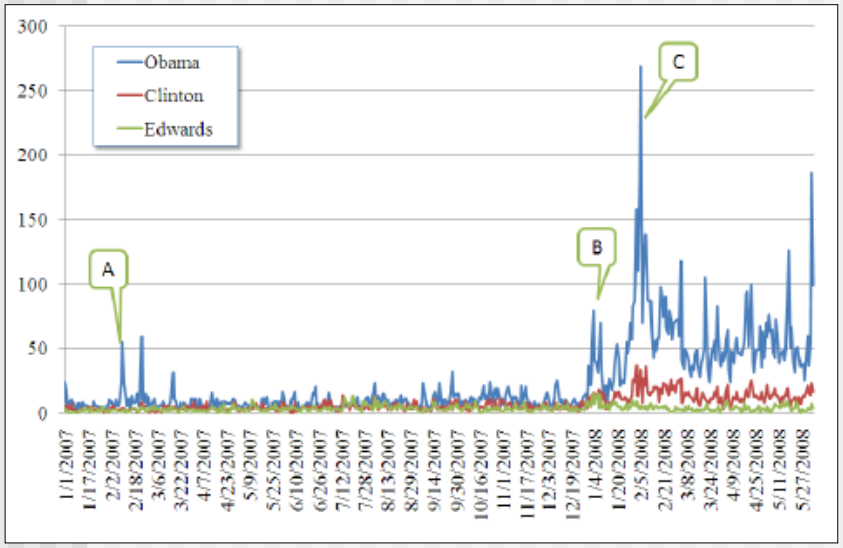
- Autoregressive (AR)
  - Attempts to predict the outcome of a system based on previous inputs.
- Seasonal Autoregressive (SAR)
  - Like AR, but has a seasonal factor to consider.
- Bass Diffusion
  - Describes the process of how products get adopted as an interaction between users and potential users.

Flickr Index

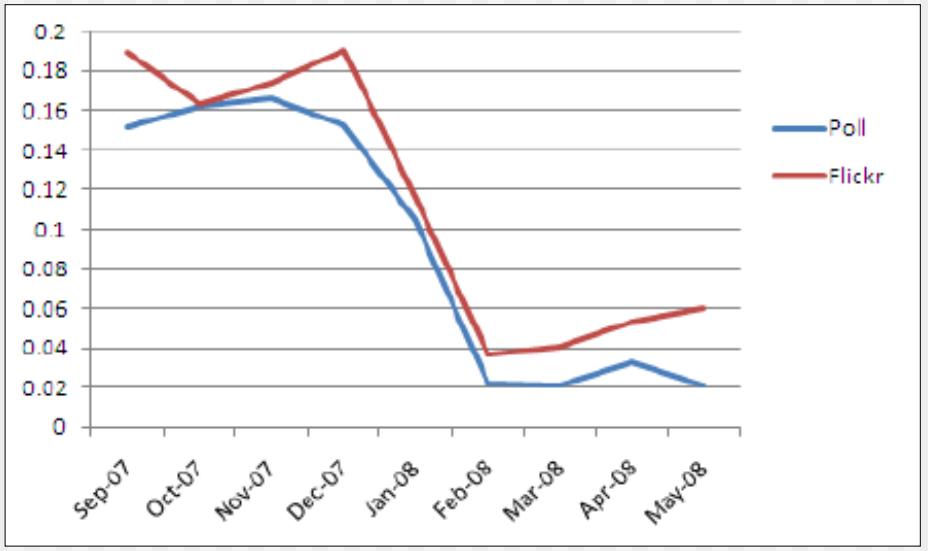
$\lambda \Theta F_t$

- Scaling value
- Makes a correlation ratio between ‘product sales’ and Flickr feature value.

2008 Elections and Poll Results

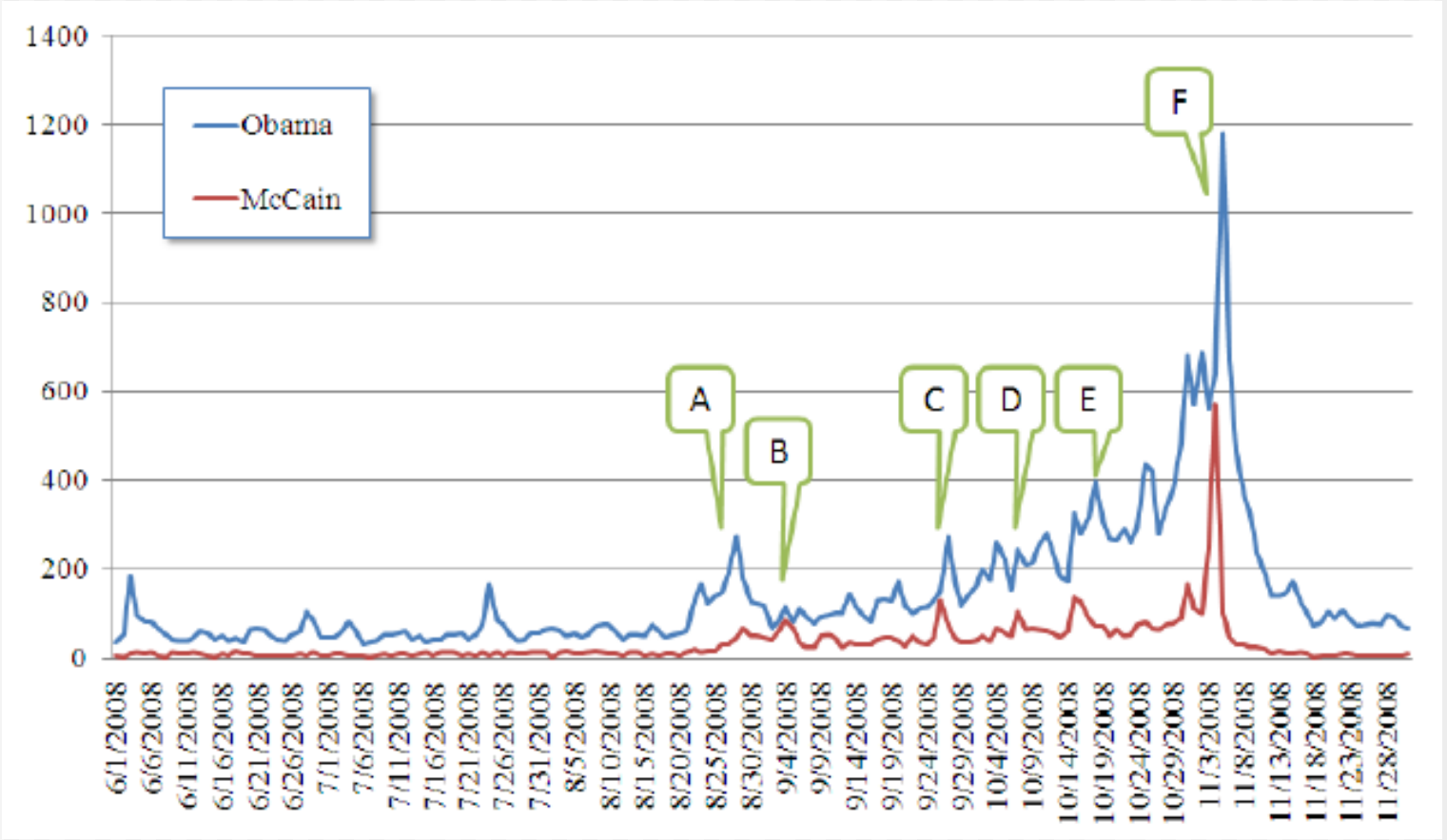


2008 Democratic Primaries

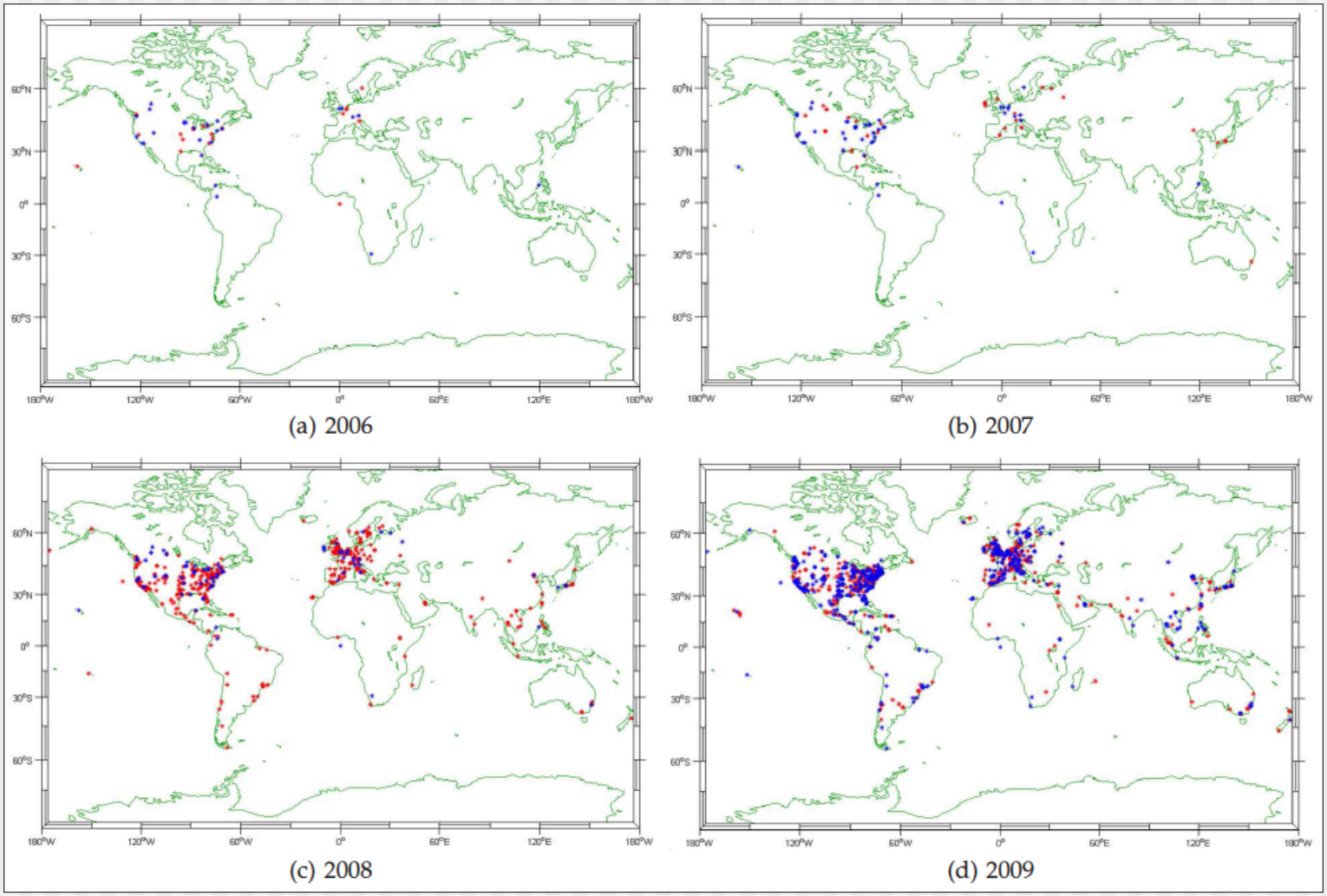


John Edwards popularity poll vs Flickr index

2008 General Election



Product Distribution

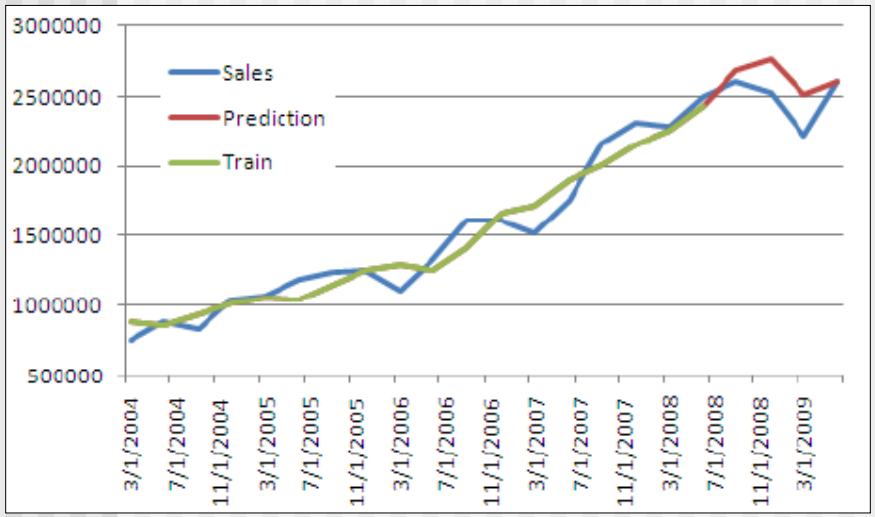


(a) 2006 (b) 2007

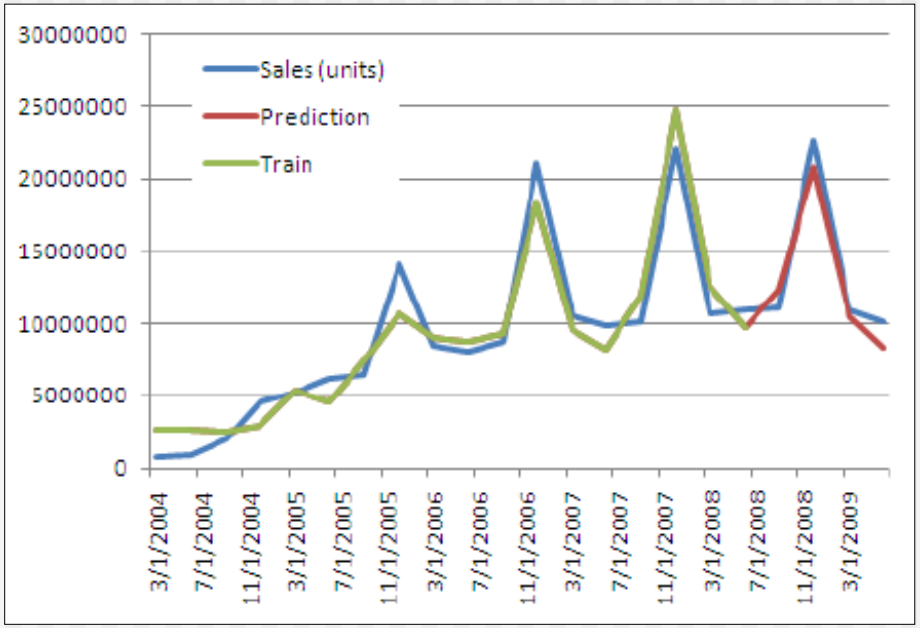
(c) 2008 (d) 2009

Sales Prediction

iPod

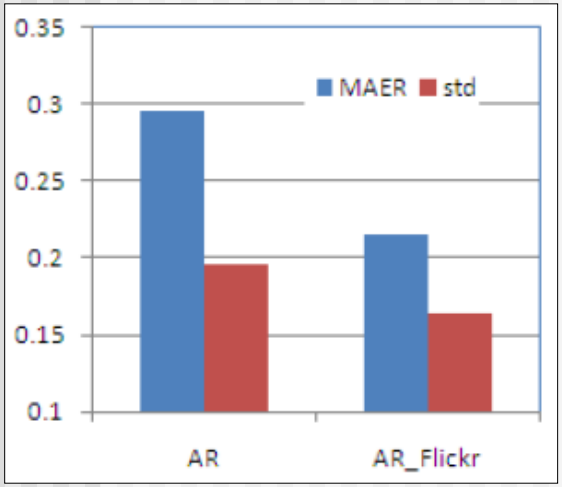


Mac

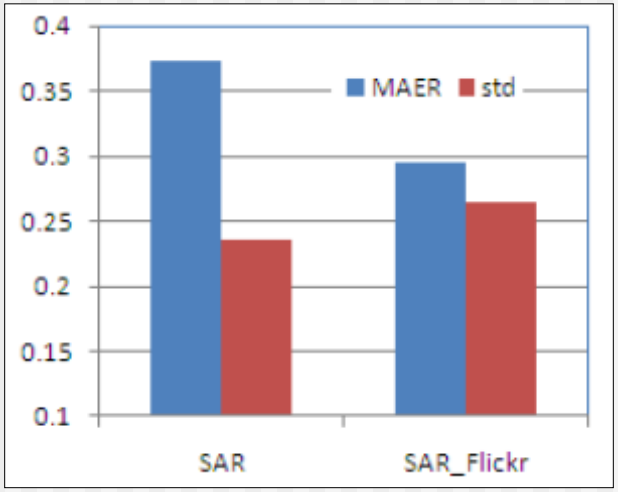


Evaluation

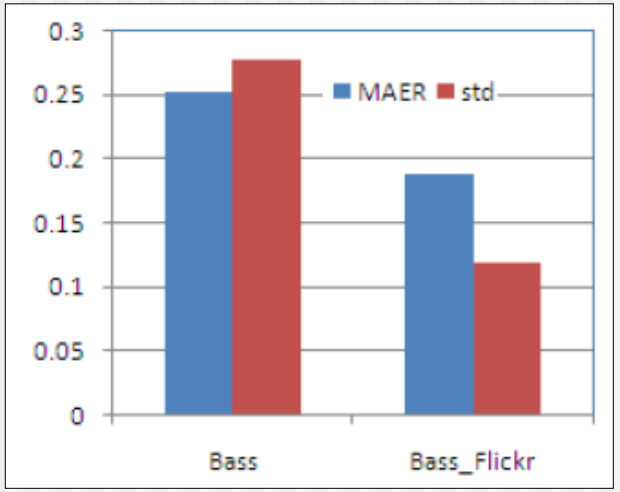
AR



SAR



Bass



- Comparison results showed that the extended Flickr models were more accurate!
- Why do you think that it was?

Questions

- How do you feel about using social multimedia for scientific purposes?
- What other online resources could be mined, and for what applications?