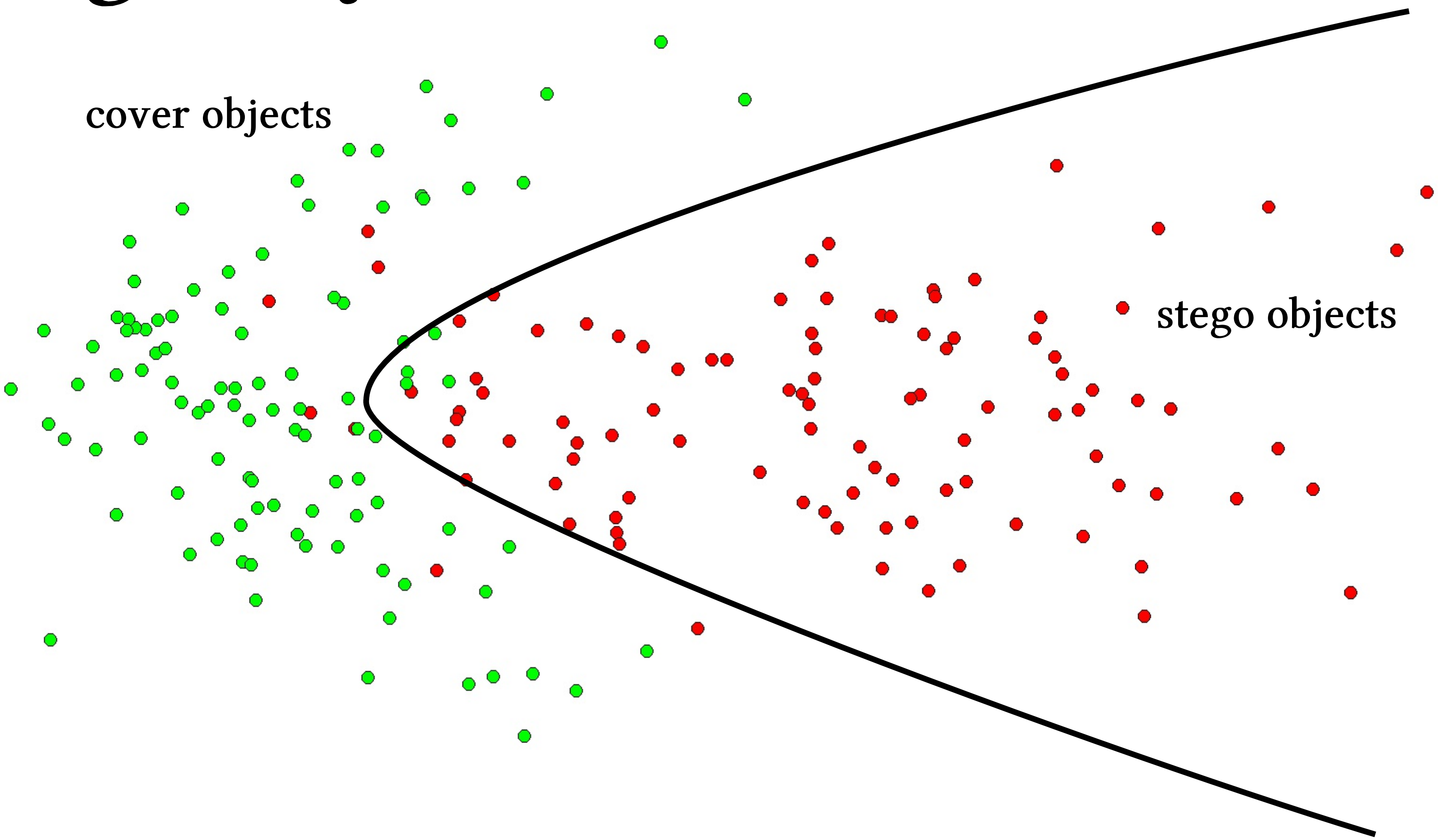
Identifying a steganographer
in realistic and heterogeneous data sets

Andrew Ker
adk@cs.ox.ac.uk
Department of Computer Science, Oxford University

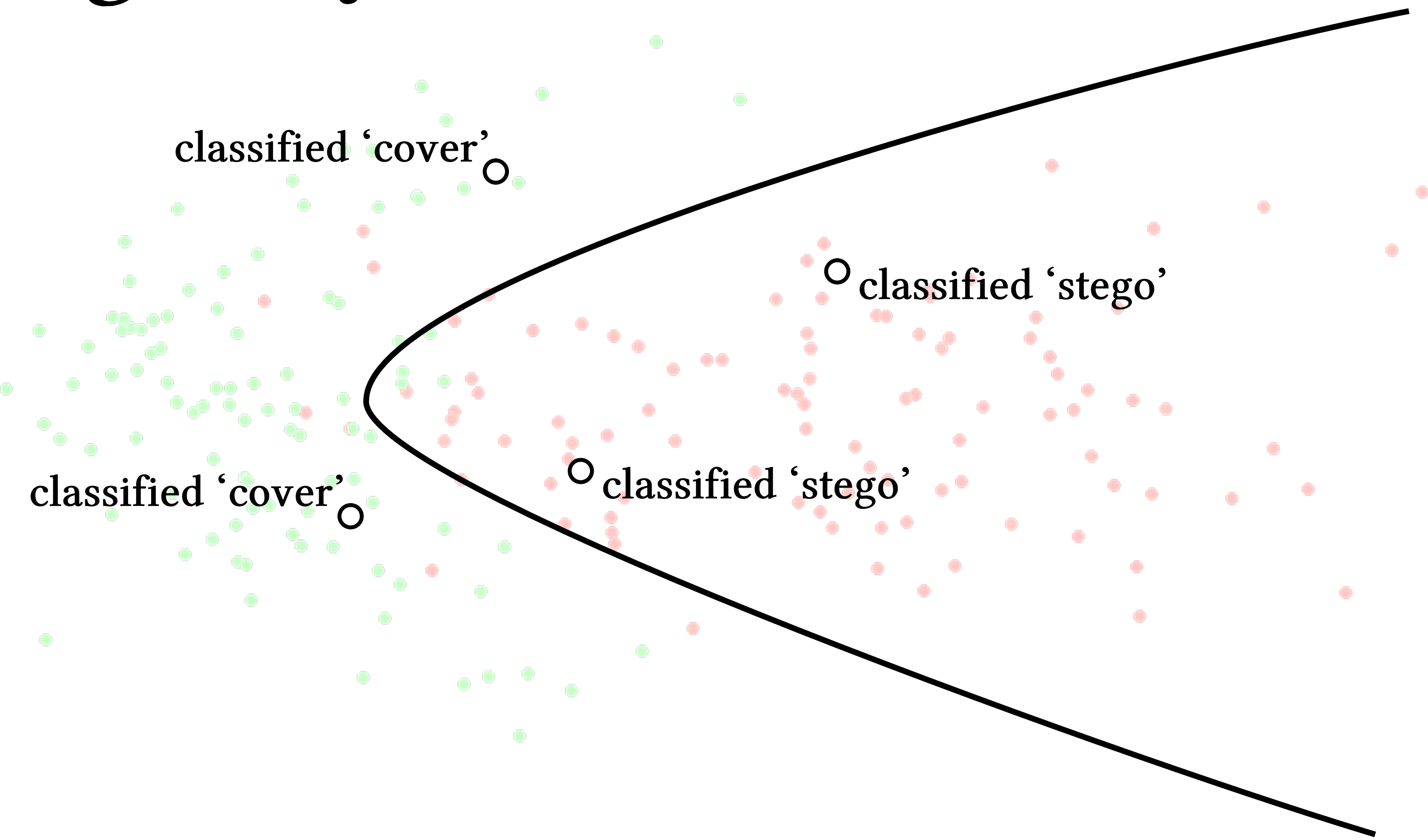
Tomáš Pevný
pevnak@gmail.com
Agent Technology Center, Czech Technical University in Prague

SPIE/IS&T Electronic Imaging, San Francisco, 25 January 2012

Steganalysis

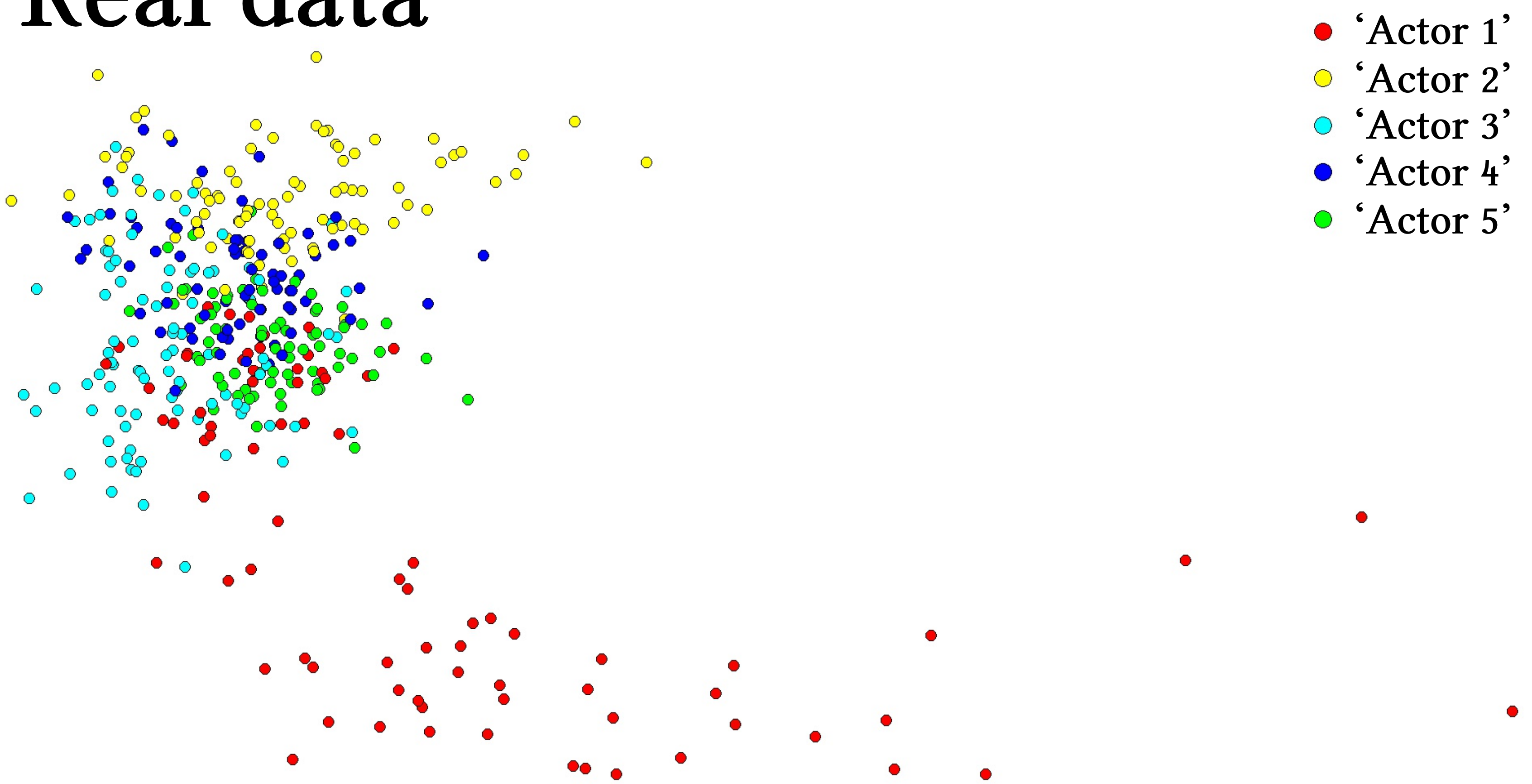


Steganalysis



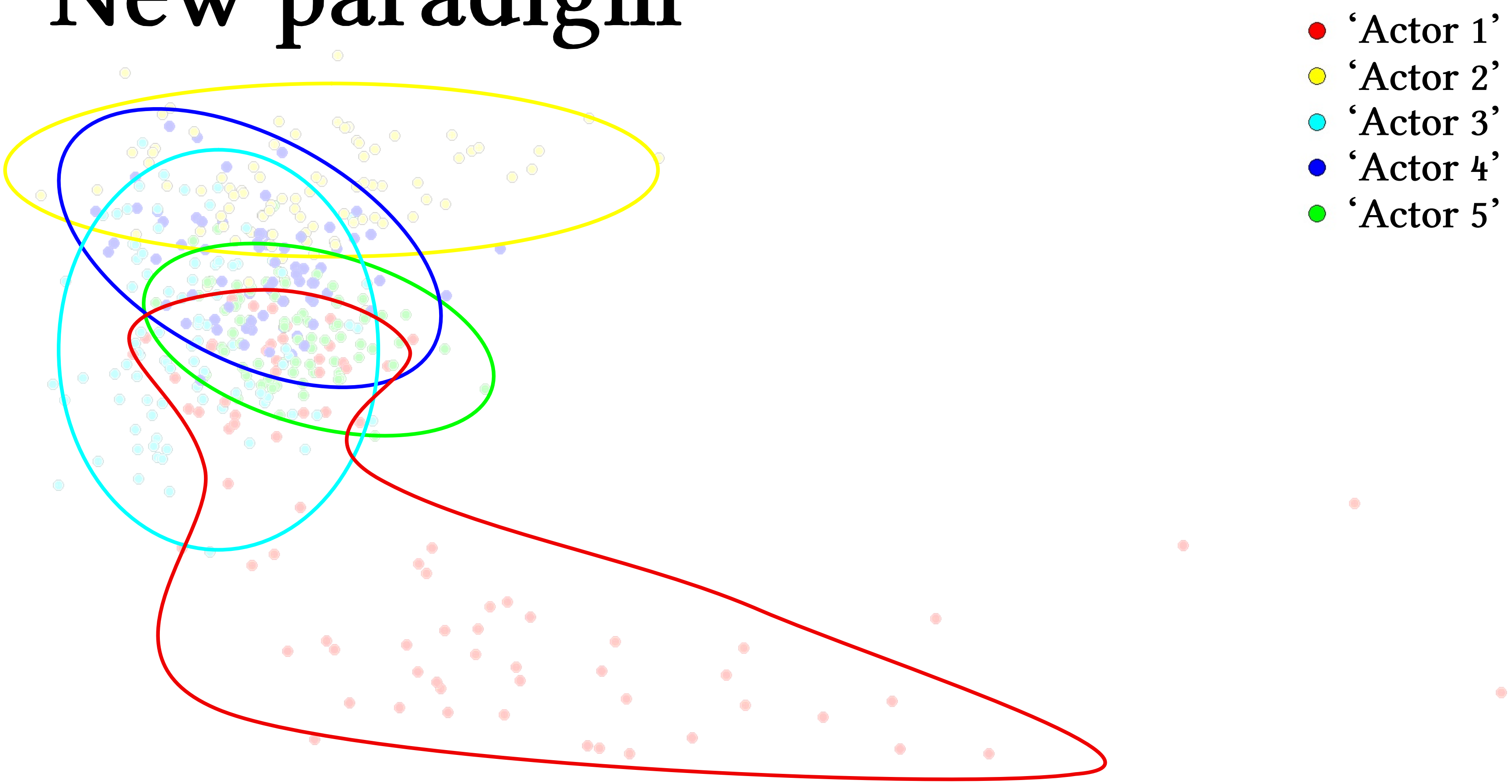
- Stego training set requires knowledge of embedding algorithm.
- Does not ‘identify a steganographer’.

Real data



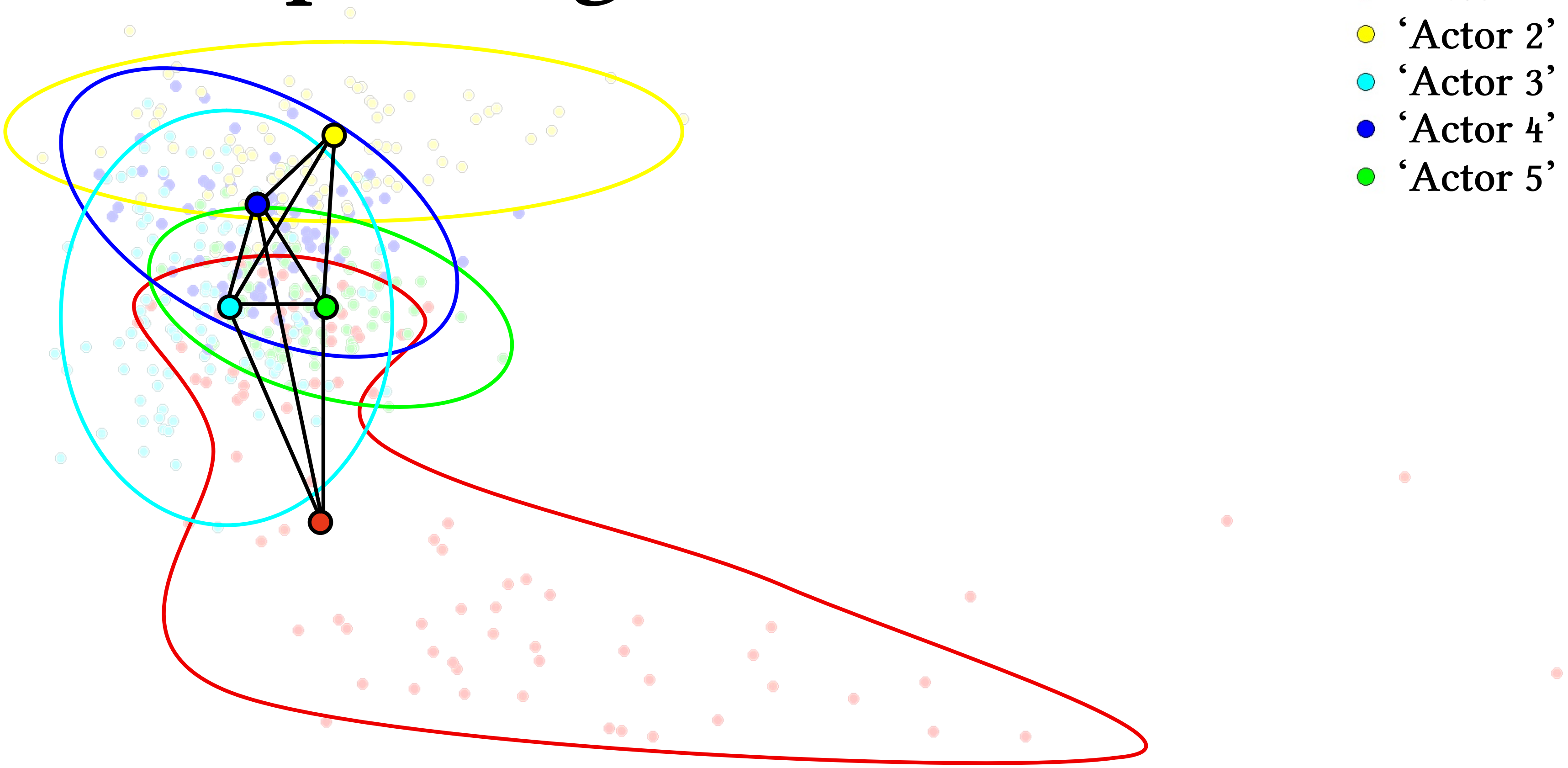
- Many actors, transmitting many objects each.
- Different actors’ sources have different characteristics:
  model mismatch is guaranteed!

New paradigm



1. Extract features.
   Use each actor's output to estimate their overall distribution.

New paradigm



1. Extract features.
   Use each actor's output to estimate their overall distribution.
2. Compute a **distance** between each pair of actors.
3. Identify the steganographer(s).

New paradigm

Features

- 'PF274' features: 274-dimensional features for JPEGs.

Same method should work with any stego-sensitive features.

Distance between actors

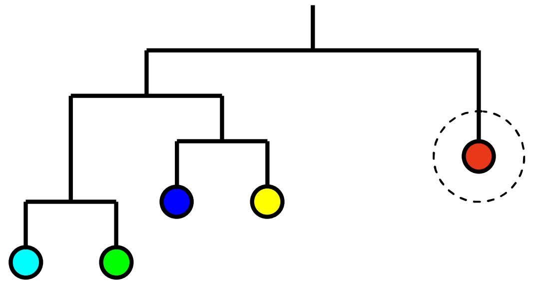
- Maximum Mean Discrepancy: $D(X, Y) = \sup_f E[f(X)] - E[f(Y)]$.

Has simple consistent estimator.

Identification of steganographer(s)

- Previous work: hierarchical clustering.

Worked well with 13 cameras as actors.
Does not work well on real-world data.



New paradigm

Features

- 'PF274' features: 274-dimensional features for JPEGs.

Same method should work with any stego-sensitive features.

Distance between actors

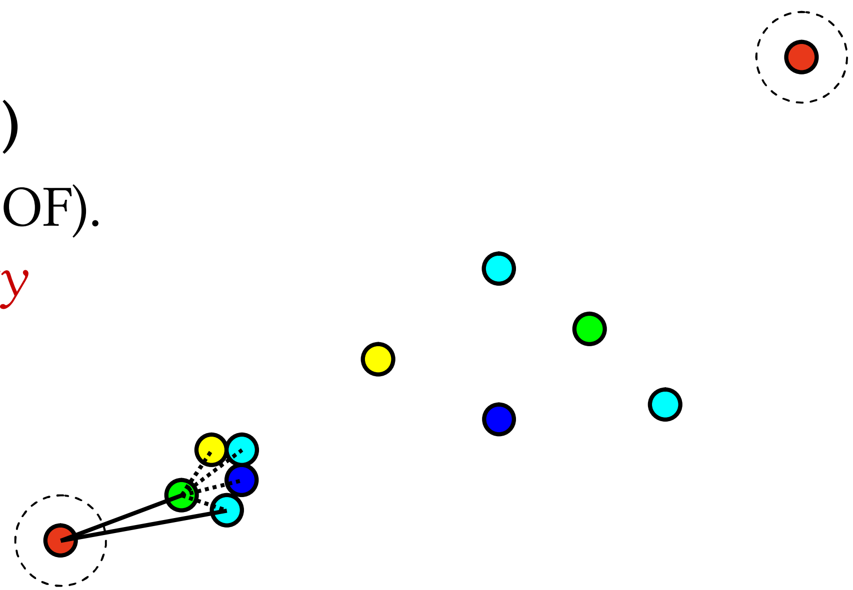
- Maximum Mean Discrepancy: $D(X,Y) = \sup_f E[f(X)] - E[f(Y)]$.

Has simple consistent estimator.

Identification of steganographer(s)

- New work: local outlier factor (LOF).

Compares local density with density around k-nearest neighbours.



Local outlier factor

Advantages

- Scale invariant.
- Has only one parameter (k), to which it is relatively insensitive.
- Allows actors to be ranked by ‘degree of being an outlier’.

Even if you cannot identify the guilty actor uniquely, can hope to include them in a ‘top n most-suspicious’ list.

Realistic, heterogeneous data set

On a leading social networking site...

- some users permit global access to images they appear in;
- we can click next image or see more of user (if user permits).

Automated process of following links, restricted to 'Oxford University' users, resulted in 4,051,928 images from 78,107 uploaders.

Ethics

- All data anonymized.
- Kept only images, grouped by 'owner', no personal information.
- All images globally visible at the time of download.

Realistic, heterogeneous data set

On a leading social networking site...

- some users permit global access to images they appear in;
- we can click next image or see more of user (if user permits).

Automated process of following links, restricted to 'Oxford University' users, resulted in 4,051,928 images from 78,107 uploaders.

Data set

- Selected 200 images from each of 4000 uploaders (actors).
- Filtered only for triviality and standard JPEG quality factor.

+ Same quality factor, similar size (around 1 Mpix).

- Mixture of sources, even within actors.
- Resampling & double-compression artefacts.
- Some already-tampered images; hopefully no stego images.

Experiments

- Select $\{50, 100, 200\}$ images from $\{100, 200, 400, 800, 1600, 3200\}$ actors at random.

- One is the **guilty** steganographer: embed using nsF5 in some of their images.

- Perform identification:

  1. *Compute PF274 features of every image,*
  2. *MMD distance between each actor,*
  3. *Rank actors by LOF ($k = 10$).*

- Measure how often **guilty** actor appears in top 10 or top 1% most suspicious.

Feature pre-processing

PF274 features have different scales, and must be pre-processed:

- **scaling** (zero mean, unit variance)
- **whitening** (PCA)

Also allows comparison with methods from literature, projecting features onto 1 dimension (distance to separating hyperplane):

- **1-class SVM** **[Farid, Pevný et al.]**

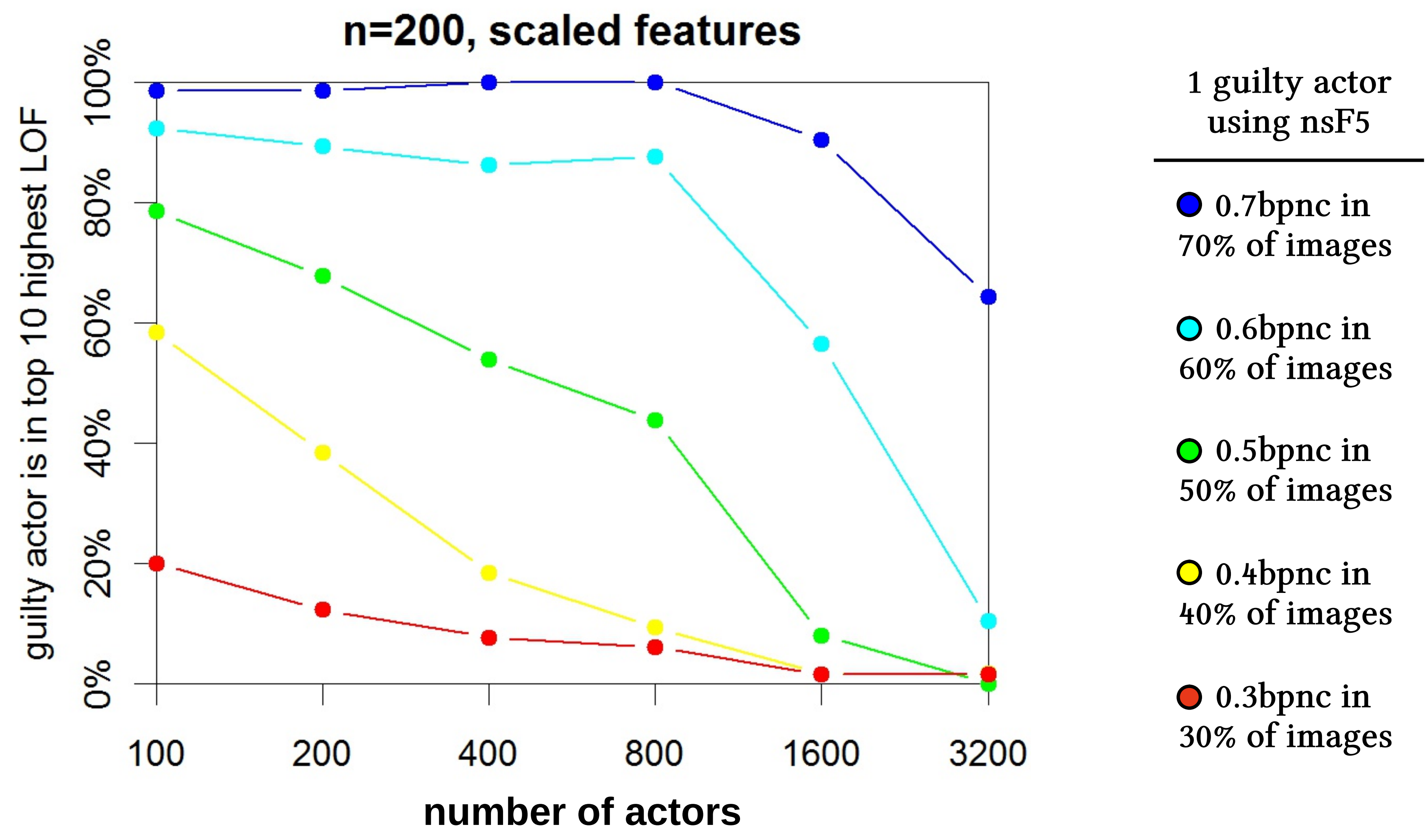
ν-SVM (ν=0.01) trained on 6000 randomly-selected observed features.

- **2-class SVM** **[many authors!]**

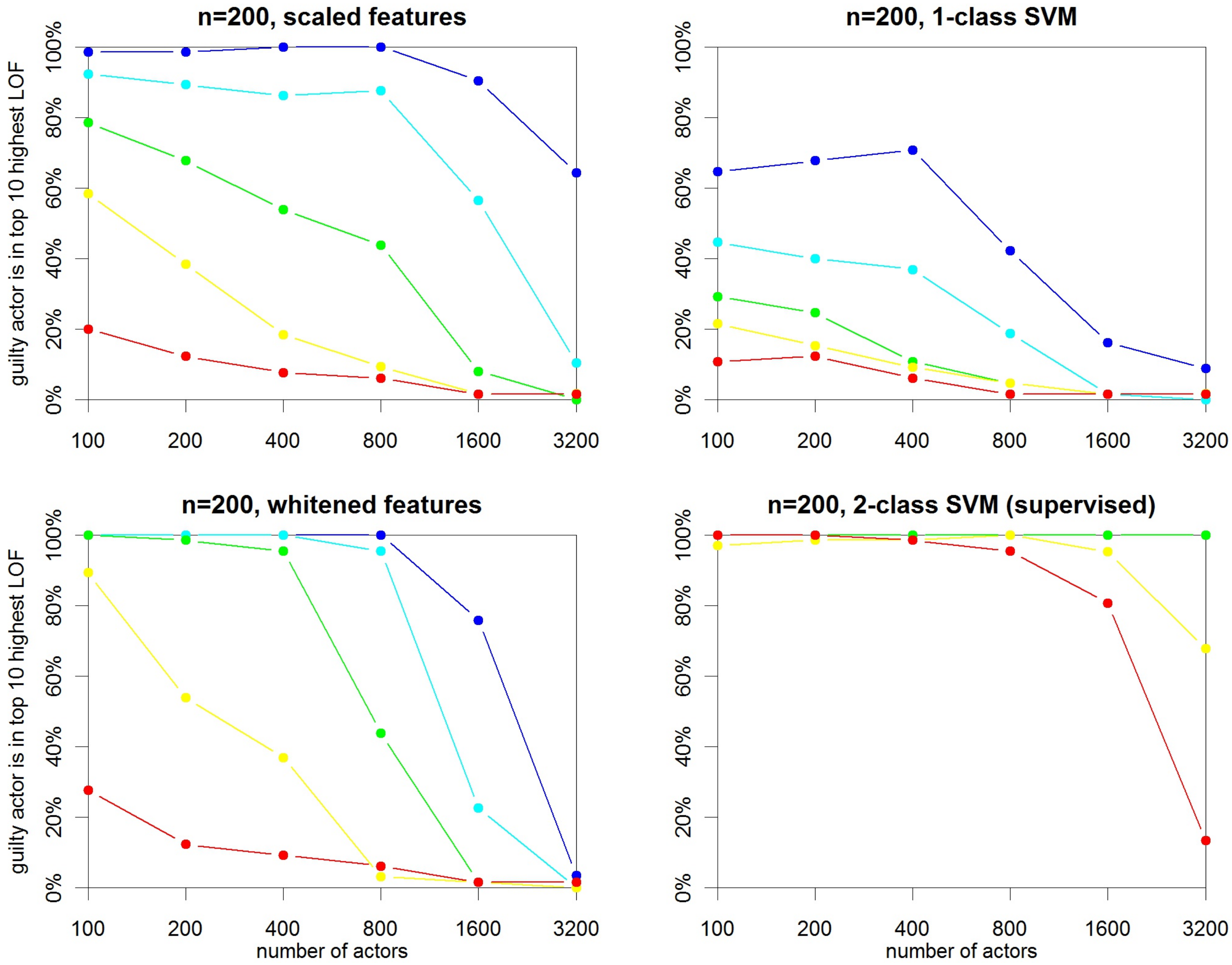
20 C-SVMs trained on 6000 randomly-selected cover/stego features from a disjoint training set…

… not a fair comparison because requires knowledge of embedding algorithm, as well as expensive training.

Results


n=200, scaled features
guilty actor is in top 10 highest LOF
0%
20%
40%
60%
80%
100%
100
200
400
800
1600
3200
number of actors
1 guilty actor using nsF5
0.7bpnc in 70% of images
0.6bpnc in 60% of images
0.5bpnc in 50% of images
0.4bpnc in 40% of images
0.3bpnc in 30% of images

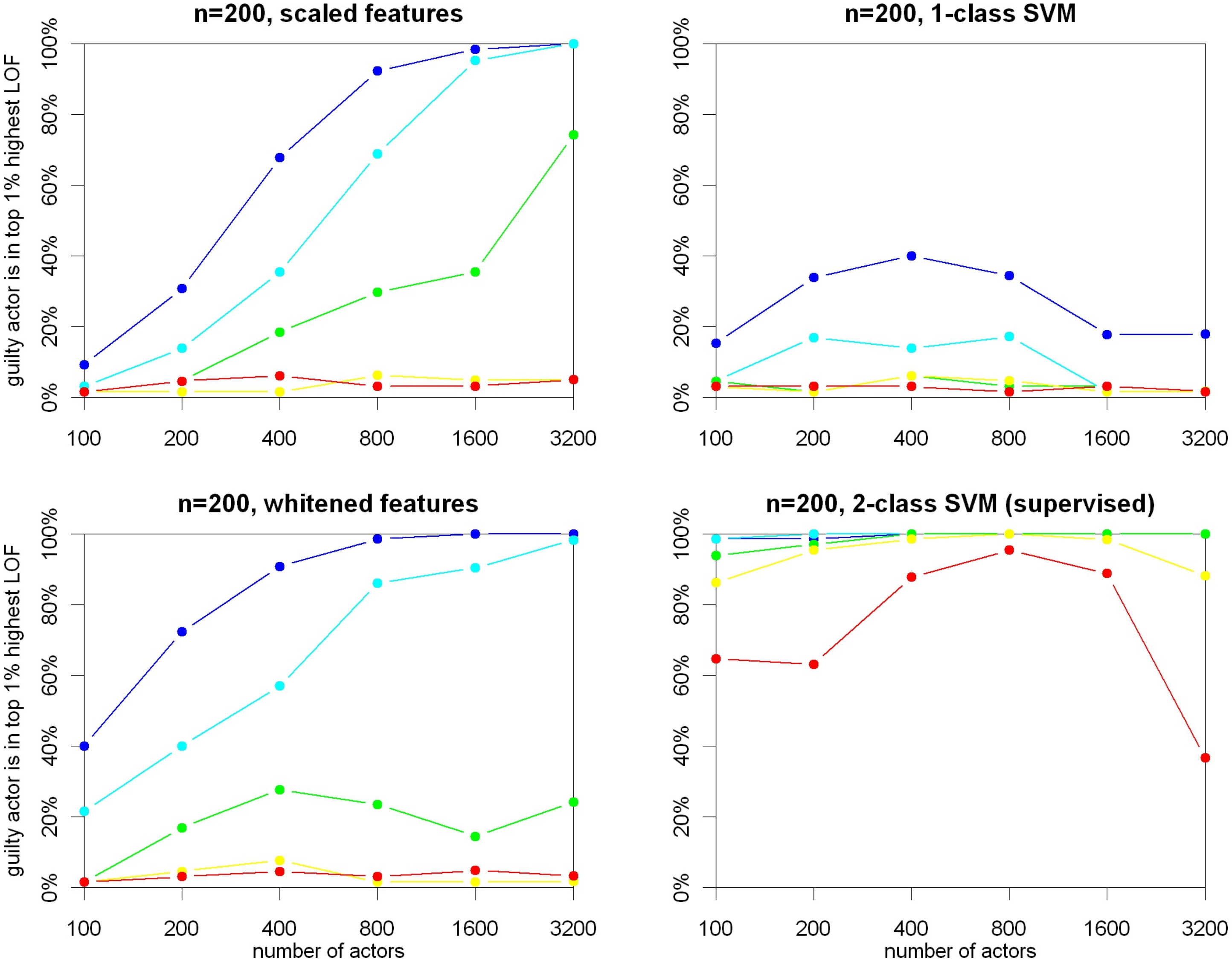
Results



1 guilty actor using nsF5

- 0.7bpnc in 70% of images
- 0.6bpnc in 60% of images
- 0.5bpnc in 50% of images
- 0.4bpnc in 40% of images
- 0.3bpnc in 30% of images

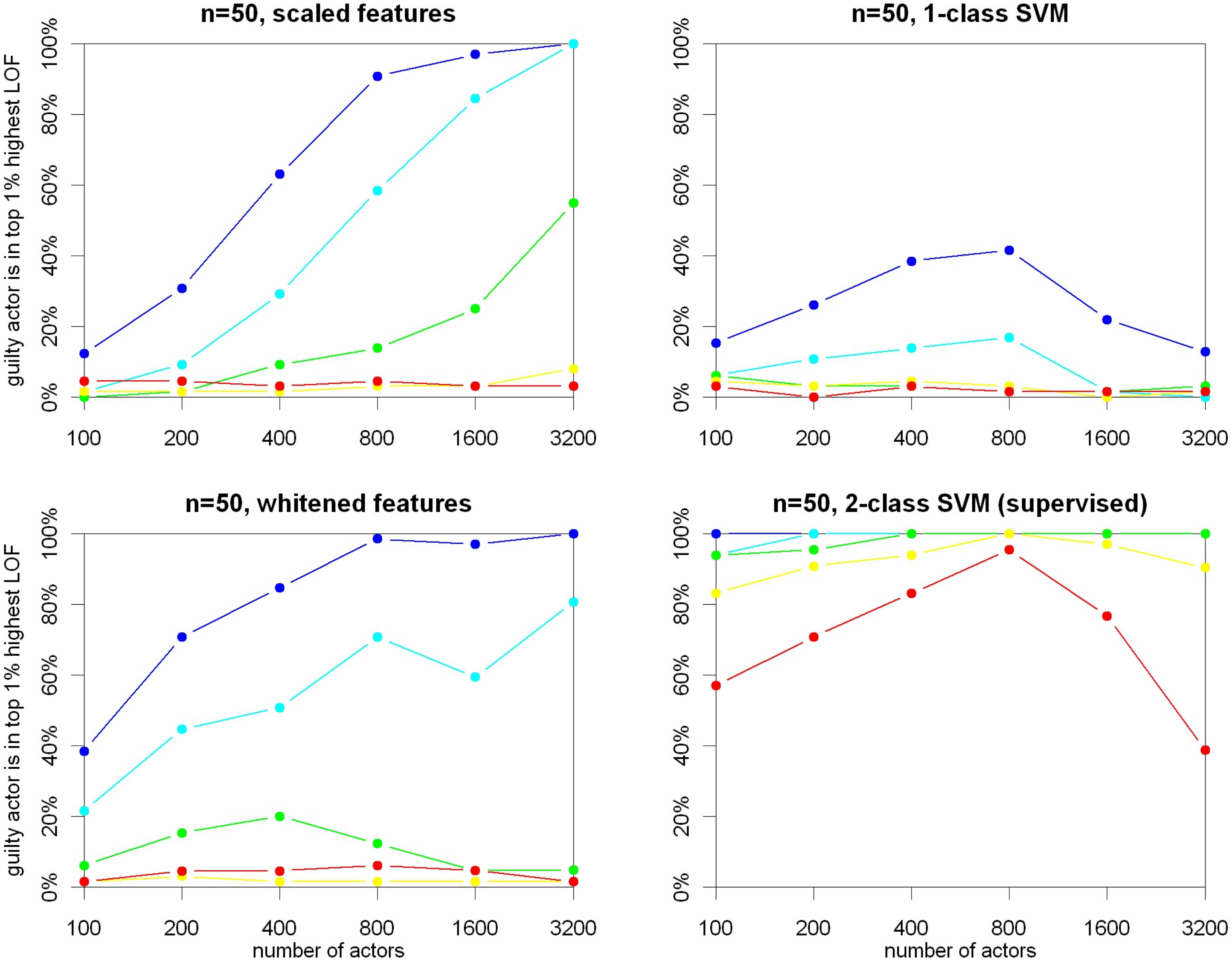
Results



1 guilty actor using nsF5

- 0.7bpnc in 70% of images
- 0.6bpnc in 60% of images
- 0.5bpnc in 50% of images
- 0.4bpnc in 40% of images
- 0.3bpnc in 30% of images

Results



1 guilty actor using nsF5

- 0.7bpnc in 70% of images
- 0.6bpnc in 60% of images
- 0.5bpnc in 50% of images
- 0.4bpnc in 40% of images
- 0.3bpnc in 30% of images

Conclusions

- Identifying steganographer(s) means working on the level of actors, not individual images.

  Allows us to identify a 'typical' level of model mismatch.
  Make heterogeneous data work for us, not against us.

- Outlier analysis is more flexible than clustering.

  Can rank by suspicion.
  We are only identifying steganographer(s) if the feature set is stego-sensitive.

- This method works on real-world data.

- Potential for universal unsupervised steganalysis...

  ...effectively self-training as long as the majority of actors are innocent.

Interesting questions

- Is linear MMD really the best distance metric?
  Surely not.

- How to deal with multiple guilty actors?
  LOF has some problems if k too small.

- How few images (per actor) is sufficient?
  Fewer than 50?

- How to refine features to make them more stego-sensitive?
  Massive feature sets are probably no good for this application.