Computer Forensics Application

- **eBay-UAB Collaborative Research**: Product Image Analysis for Authorship Identification

Project Overview

- A new framework that provides additional clues extracted from images and free-form texts for:
  - ***Authorship identification***
  - ***Multi-account linking***

- Three main modules:
  - ***Picture Quality Assessment:*** classifying product listing images into professional/amateur/stock images
  - ***Image Editing Style Summarization:*** summarize the common patterns shared by images under the same seller
  - ***Image Embedded Text Analysis***: extract and summarize embedded text from listing images

Module I: *Picture Quality Assessment*

- **Assessment scheme I**: classify images into {professional, amateur} classes
  - Quality-based classification:
    - **Professional**: clean (or human edited) background, object in focus, good lighting, *etc.*
    - **Amateur**: otherwise.

Module I: *Picture Quality Assessment*

- Algorithm:
  - Visual features that encode background clutter, color, lighting, sharpness, *etc.* are computed for all images.
  - An SVM model is trained on the visual features and used to label the testing data.
  - Five-fold cross validation is performed on the training data to select the optimal parameters for SVM (achieving ***90.1%*** accuracy).
  - The optimal parameters are used to re-train the model on all training data.
  - The re-trained model is applied to label the testing data (achieving *89.9%* accuracy).
- http://students.cis.uab.edu/galabing/ebay/images/quality

Module I: *Picture Quality Assessment*

- **Assessment scheme II**: classify images into {stock, non-stock} classes
  - A different way to categorize images: whether they are unique to one user, or shared by multiple users (**stock**).
- Algorithm:
  - Match near-duplicate images across users: given a testing image, we retrieve all of its near-duplicates, and count the number of users to determine whether the testing image is a stock image.
  - http://students.cis.uab.edu/galabing/ebay/images/stock/stock-table.html

Module II – Image Editing Style Summarization

- **SIFT** (*Scale-invariant feature transform*) feature based summarization

Module II – Image Editing Style Summarization ---SIFT feature based

- Interest point features capture image points with distinctive texture so that the same points can be robustly matched across multiple images.
- We assume that many eBay users have developed distinctive editing styles over time to embed to their product images (logos, background, etc.)
- Such editing styles are highly repetitive within one user's images, but mostly distinctive among different users'.
- We detect and encode such editing styles for each user using interest point features, and in turn use the encoded editing styles to automatically predict the ownership for unlabeled images.

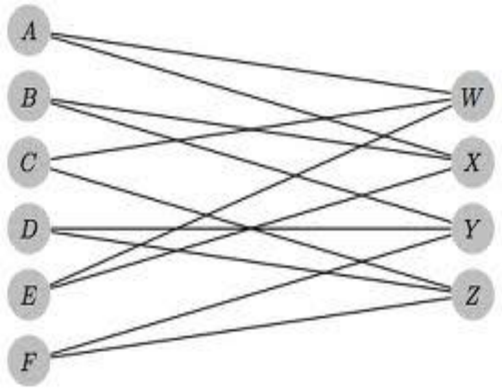
Module II – Image Editing Style Summarization ---SIFT feature based

- We collected 917 images from 10 eBay users, and applied a 10-fold cross validation.
- Give a set of (training) images for a user:
  - Extract PCA-SIFT features from all images.
  - Match PCA-SIFT features among all pairs of images.
    - Naïve method: O(N^2)
    - Our method: O(NlogN)
    - Both yield almost identical results
- In total, we achieved **96.6% of accuracy** in ownership prediction.
  - http://students.cis.uab.edu/galabing/ebay/images/signature2/

- Lin Yang, Wei-Bang Chen, Chengcui Zhang, John Johnstone, Gary Warner, and Song Gao, “Profiling Online Auction Sellers by Image Editing Styles,” **Accepted** for publication, ***IEEE Multimedia***.

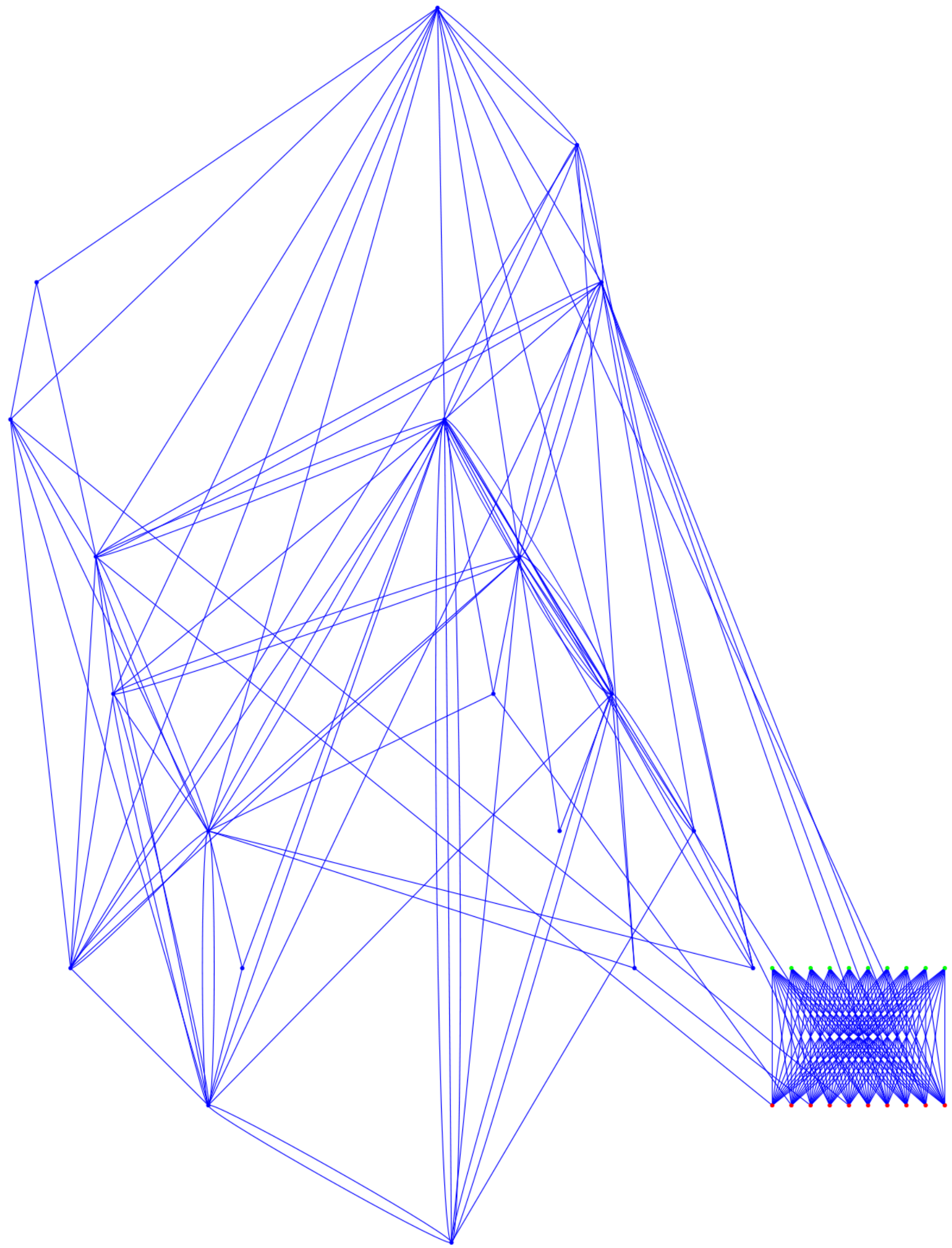
Computer Forensics Applications (cont.)

- Uncovering auction fraud from eBay transaction graph - Initial study



- Data set: eBay transaction feedbacks
  - **A total of 220,000 (two-hundred and twenty thousand) users are crawled.**
- Idea of belief propagation:
  - Fraudsters create two types of identities - *fraud* and *accomplice*, where fraud identities are the ones used eventually to carry out the actual fraud, and the accomplice identities are the ones used to help build the reputation for the fraud identities. This pattern forms a near bipartite core in the transaction graph.

- Algorithm:
  - Each vertex in the transaction graph is labeled by one of {fraud, accomplice, honest} based on their pattern of interaction with other vertexes.
  - Belief propagation (BP) is used to optimize the labeling across the entire graph by maximizing the joint probabilities of all the vertexes.
  - Honest user model: Barabasi-Albert model



- Evaluation results on the sparse eBay transaction dataset
  - 20% accomplice
  - 50% fraud???
- What can be improved:
  - Network too sparse (average degree is ~5, ideally >=10)
  - Initial probabilities (1/3, 1/3, 1/3) may not make sense.
  - **BP seems not to scale well with large graphs.**