A Graph Based Approach for Finding People in News

Derya Özkan

Bilkent University, Department of Computer Engineering
RETINA Vision and Learning Group

http://retina.cs.bilkent.edu.tr

Motivation

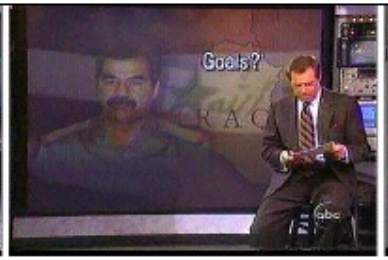
- News photographs and news videos are rich sources of information.
- People are usually the main subject of the stories in the news.

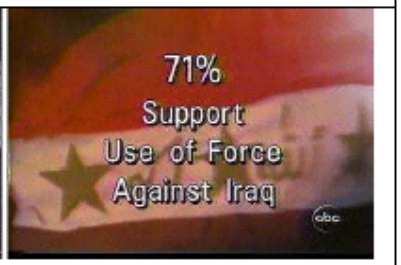
News photographs

Author J.K. Rowling (L) and her husband Neil Murray arrive at the British premiere of "Harry Potter and the Order of the Phoenix" in London July 3, 2007. REUTERS/James Boardman (BRITAIN)

Paris Hilton sits down for an interview with host Larry King (L) on CNN's 'Larry King Live', in this publicity photo released to Reuters in Los Angeles, California June 27, 2007. Hilton gave

News videos





... military power is the way to solve the Saddam Hussein problem. All of the public indications from the Clinton Administration are that because the Iraqi leader won't let the international weapons inspectors do their job. He should be taught a lesson he won't forget or won't survive. By the sound of the polls more than seventy percent support the use of force and support in congress ...

Motivation

- **Common approach:**
  - Search for the names of the people in the caption or in the speech transcript text.
- **Problem:** likely to produce incorrect results due to:
  - associated photos/shots may not include any faces/person

Democratic Presidential candidate and U.S. Senator Hillary Clinton (D-NY) and her husband, former U.S. President Bill Clinton, share a word with one another before speaking to supporters during a campaign stop in Nashua, 13, 2007.

The motorcade that carries US President George W. Bush and Vice President Dick Cheney is seen as it arrives at the US Capitol 12 June, 2007...

Motivation

 detect faces 

- Face detection algorithm may detect non-face images
- There might be other people in the picture as well






Sample detected faces that are associated with the name *President George W. Bush.*

Motivation

- Time variations between visual appearance of face and name in video.

(a)

(b)

Motivation

- Faces should be recognized in order to solve the problem.
- However, face recognition is a long standing and difficult problem.
- For larger and more realistic data sets like news photographs and/or videos, face recognition is still difficult and error-prone due to the noisy and complicated nature of these sets.

Motivation

- Faces and names appear together in news.

Author J.K. Rowling (L) and her husband Neil Murray arrive at the British premiere of "Harry Potter and the Order of the Phoenix" in London July 3, 2007. REUTERS/James Boardman (BRITAIN)

Paris Hilton sits down for an interview with host Larry King (L) on CNN's 'Larry King Live', in this publicity photo released to Reuters in Los Angeles, California June 27, 2007. Hilton gave

- If the association between names and faces is learned, then finding the images belonging to a specific person in large news achieves can be simplified.

Observations

Observation 1: **In news, a person appears around when his/her name is mentioned.**

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.

Observation 4:When a new face is encountered, it will be either in the group or not

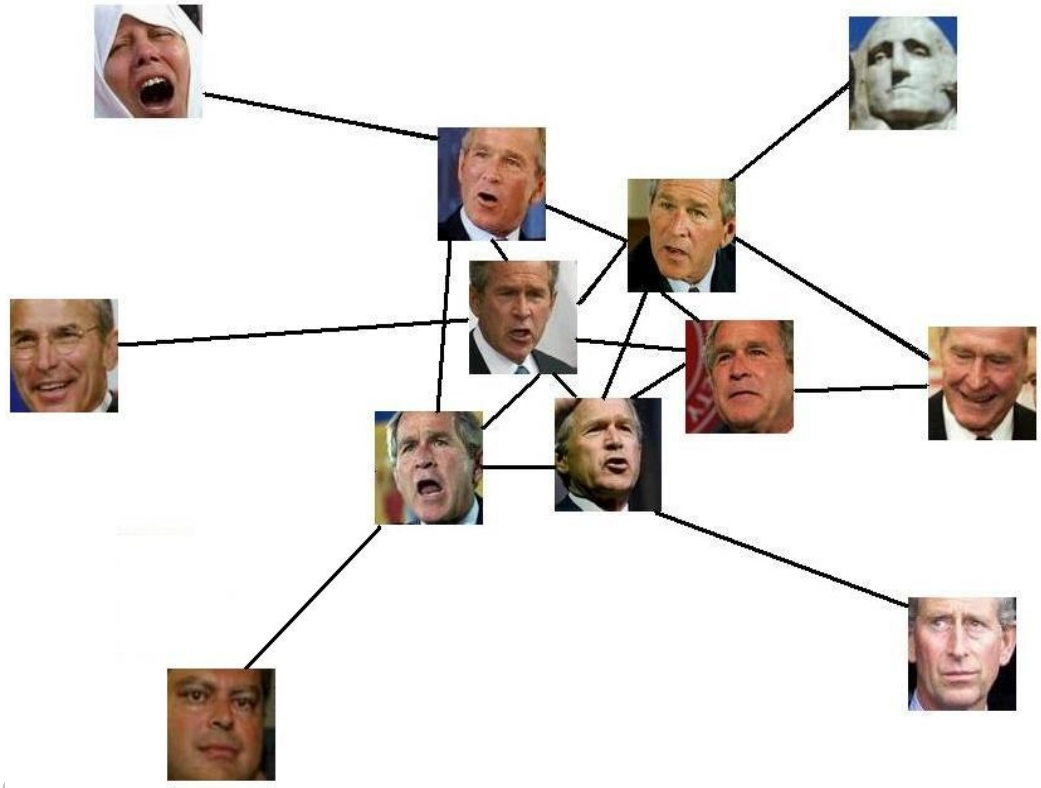
Observations

Observation 1: In news, a person appears around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.

Observation 4:When a new face is encountered, it will be either in the group or not.



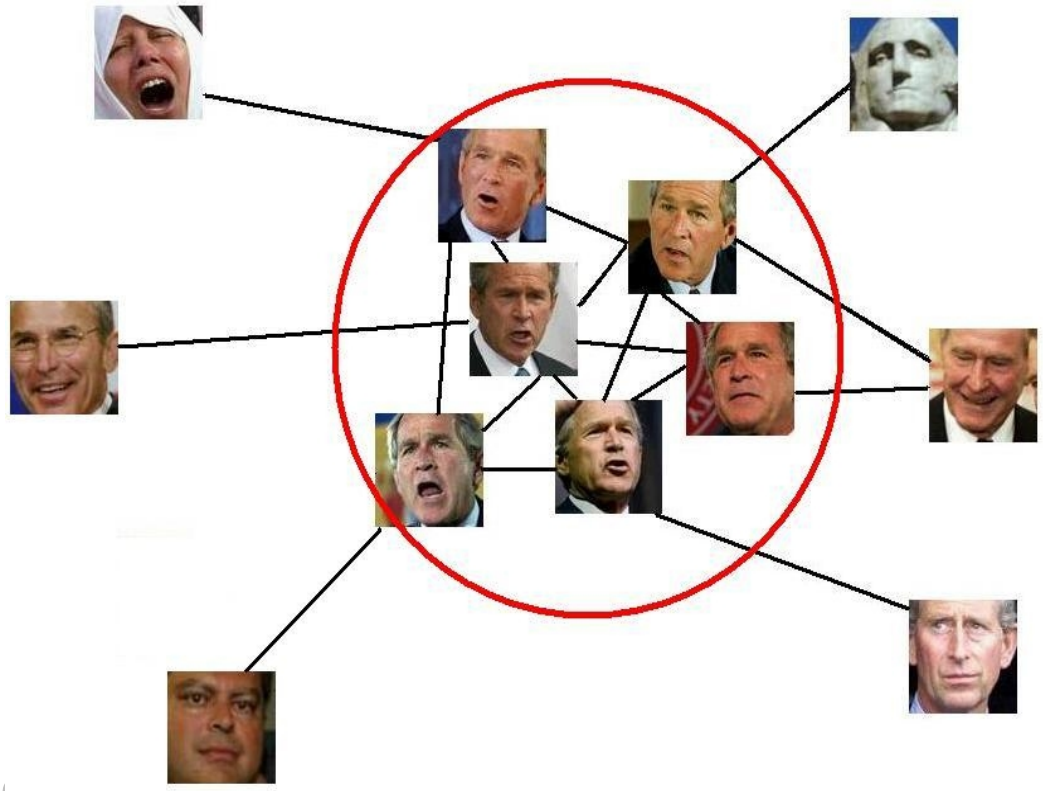
Observations

Observation 1: In news, a person appears around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces among these faces.

Observation 4: When a new face is encountered, it will be either in the group or not.

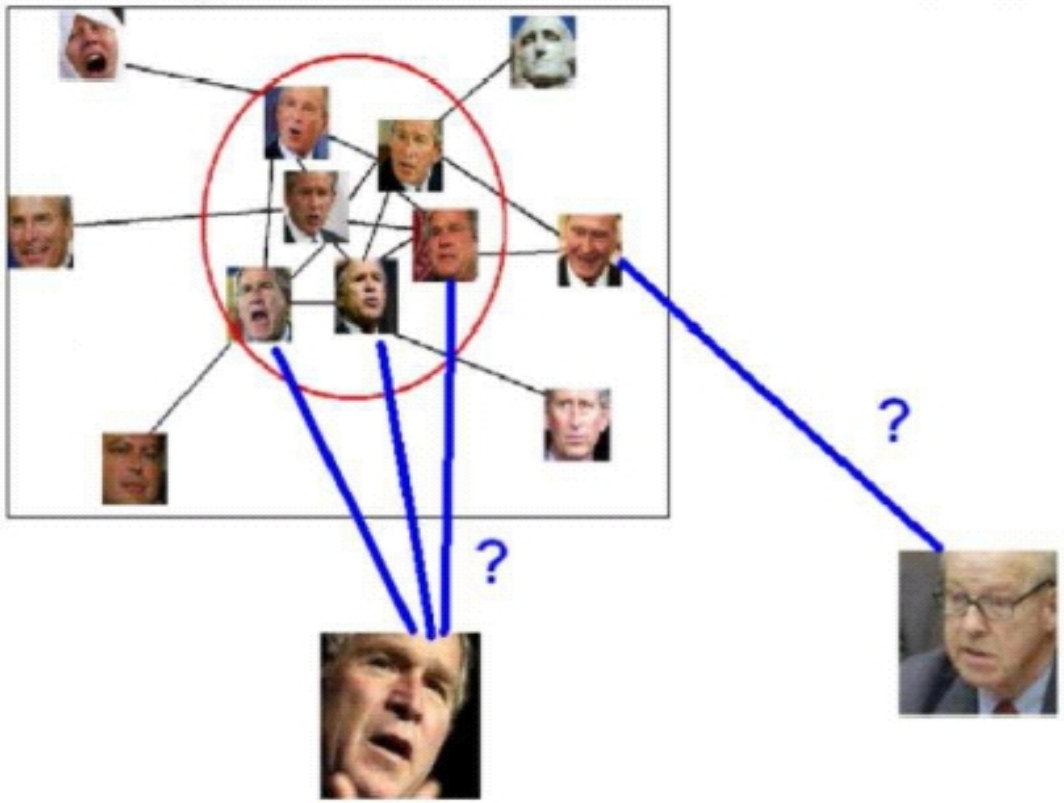


Observations

Observation 1: In news, a person appears in news around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.



Observation 4: When a new face is encountered, it will be either in the group or not.

Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.

Observation 1: In news, a person appears around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.

Observation 4:When a new face is encountered, it will be either in the group or not

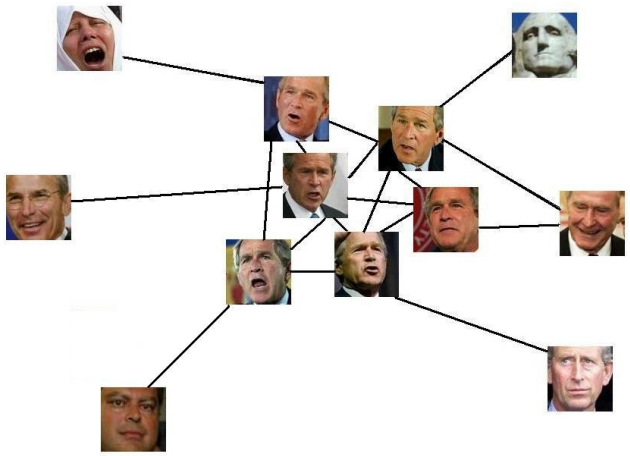
Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.



Observation 1: In news, a person appears around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.

Observation 4:When a new face is encountered, it will be either in the group or not

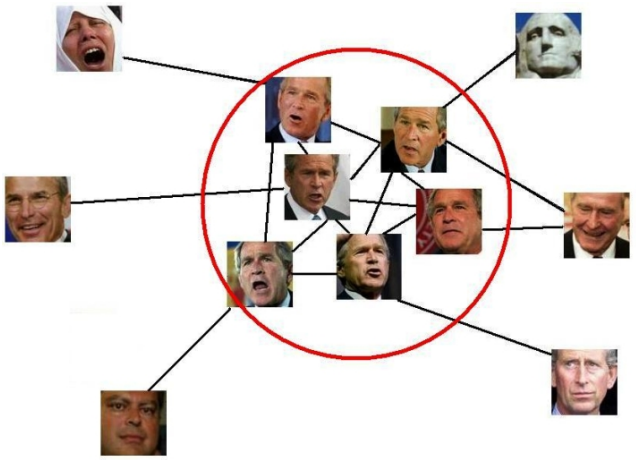
Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.



Observation 1: In news, a person appears around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.

Observation 4: When a new face is encountered, it will be either in the group or not

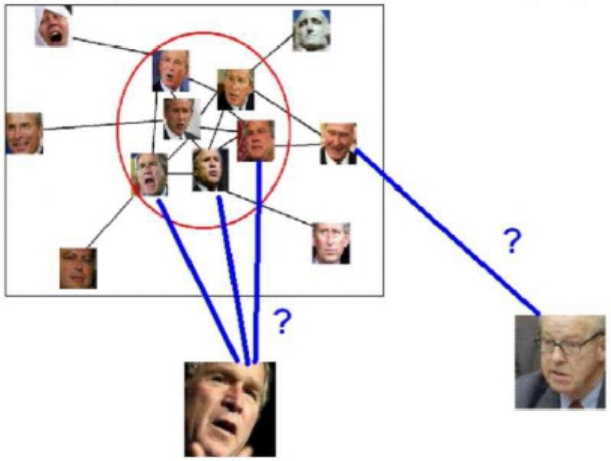
Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.



Observation 1: In news, a person appears around when his/her name is mentioned.

Observation 2: Faces of a person tend to be more similar to each other than to other faces.

Observation 3: Faces of the query person forms the largest group of similar faces in this group.

Observation 4:When a new face is encountered, it will be either in the group or not

Proposed Approach

Step 1: **Limit the search space of a query person by using name**

Step 2: Construct a similarity graph among faces in this search space.

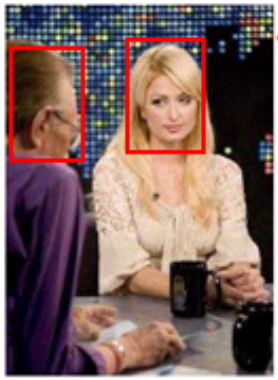
Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.

Step 1 : Integrating Names & Faces

- In news photographs, a person is likely to appear in a when his/her name is mentioned in the **caption**.

Author J.K. Rowling (L) and her husband Neil Murray arrive at the British premiere of "Harry Potter and the Order of the Phoenix" in London July 3, 2007. REUTERS/James Boardman (BRITAIN)



Paris Hilton sits down for an interview with host Larry King (L) on CNN's 'Larry King Live', in this publicity photo released to Reuters in Los Angeles, California June 27, 2007. Hilton gave

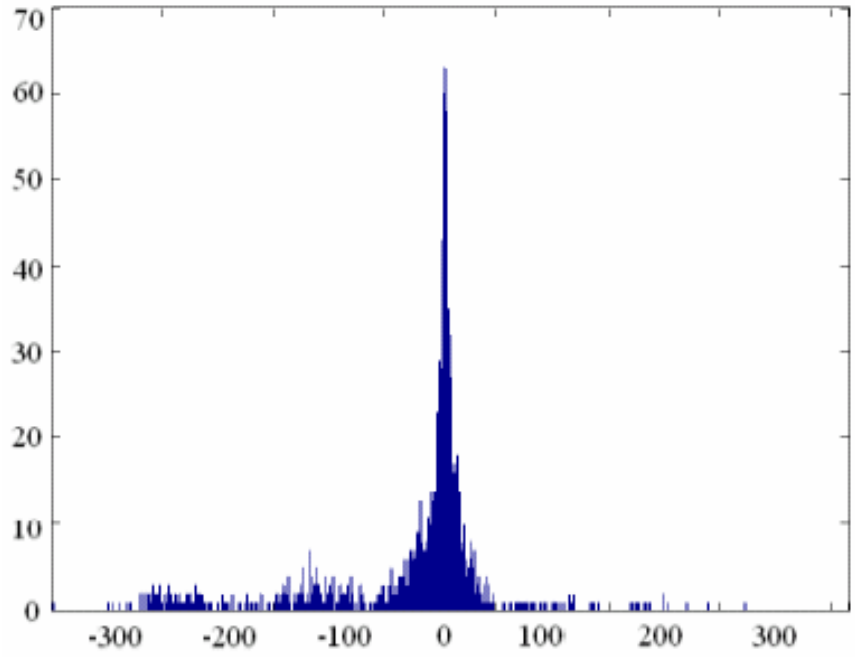
- A person's name can appear in different forms:

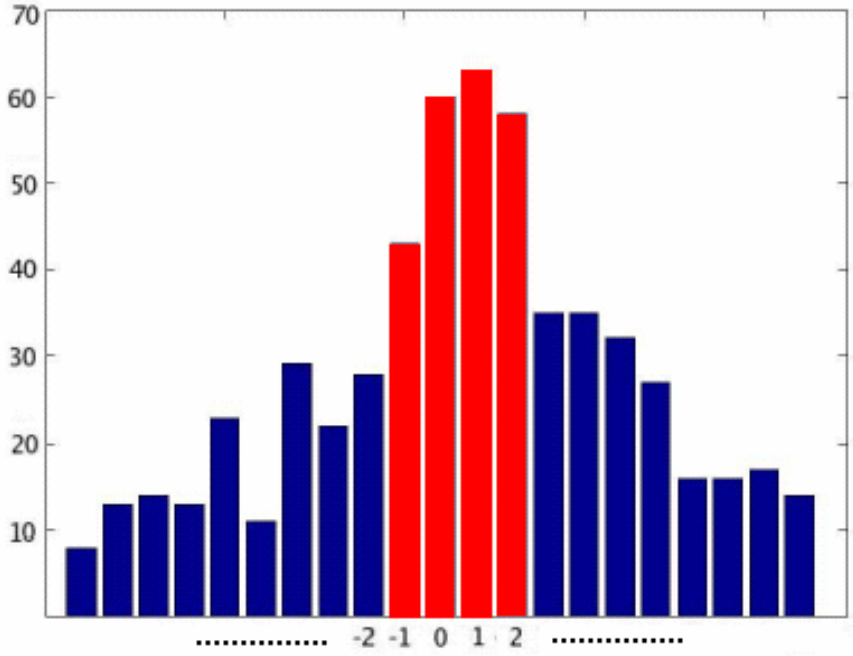
  George W. Bush, President Bush, U.S. President, and President George Bush, all correspond to the same person.

Step 1 : Integrating Names & Faces

- For the news videos, the probability of a person appearing on the screen is high when his/her name is mentioned in the **speech transcript text**.

Step 1 : Integrating Names & Faces

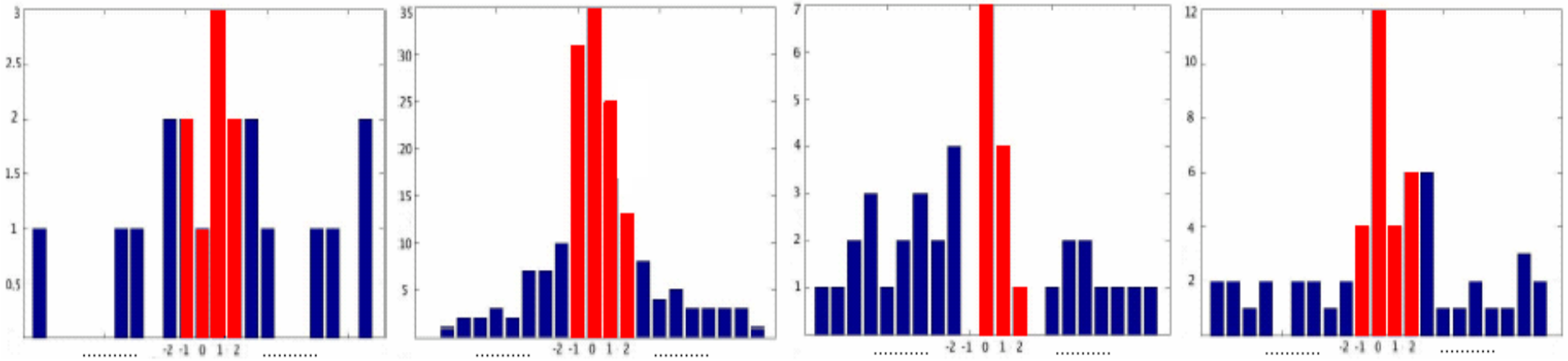




Frequency of Bill Clinton's visual appearance w.r.t the distance to the shot in which his name is mentioned.

Step 1 : Integrating Names & Faces

- The relative position of the faces to the name for Benjamin Netanyahu, Sam Donaldson, Saddam Hussein, and Boris Yeltsin respectively.



- It is seen that taking only one preceding and two following shots ([-1,2]) is also a good choice.

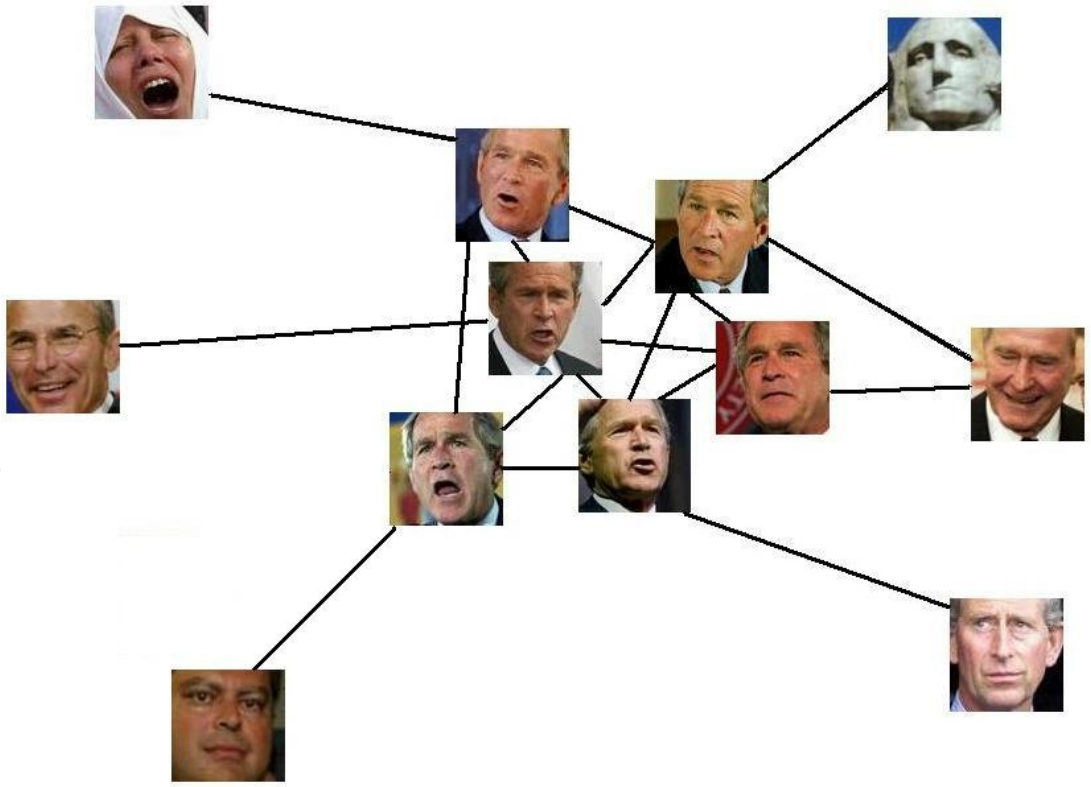
Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

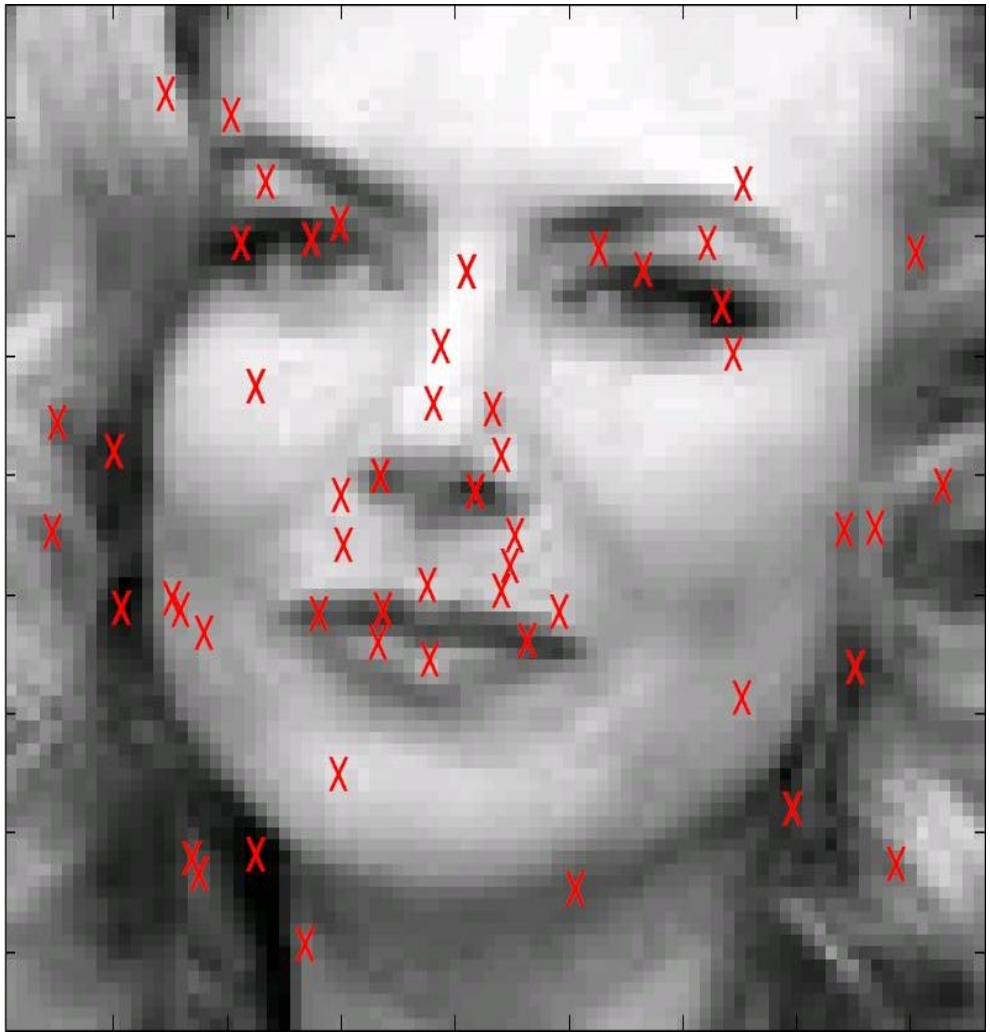
Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.



Step 2 : Construct Similarity Graph

- The similarity of faces are defined using the interest points extracted from the detected face areas.
- Each face is defined by interest points extracted by using the Lowe's SIFT operator.



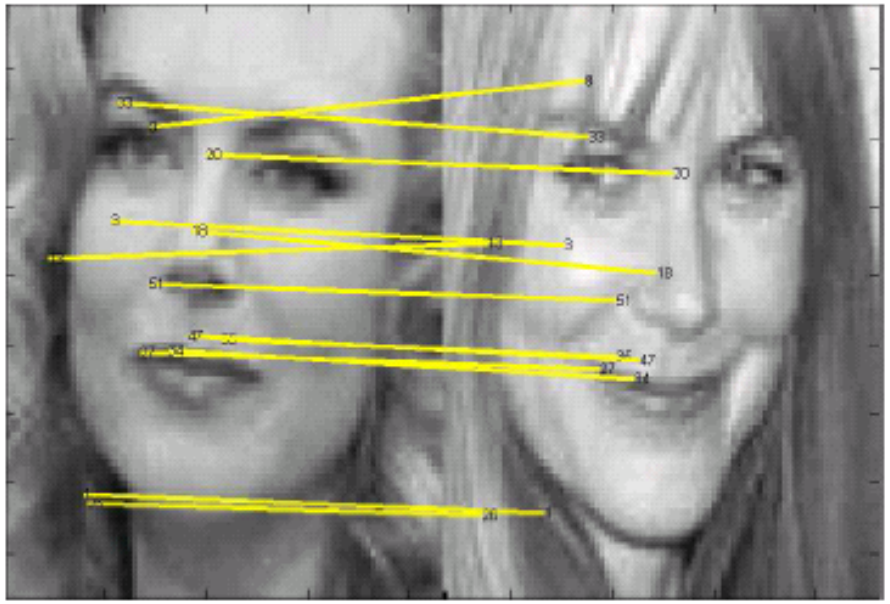
D. G. Lowe. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60(2), 2004.

Similarity Between Two Faces

- While comparing two faces, each point on one face is compared with all the points on the other face and the points with the least Euclidean distance are selected as true matches.
- However, this approach returns many false matches along with the true ones.
- Hence, we apply two constraints on those matches:
  - geometrical constraint
  - the unique match constraint.

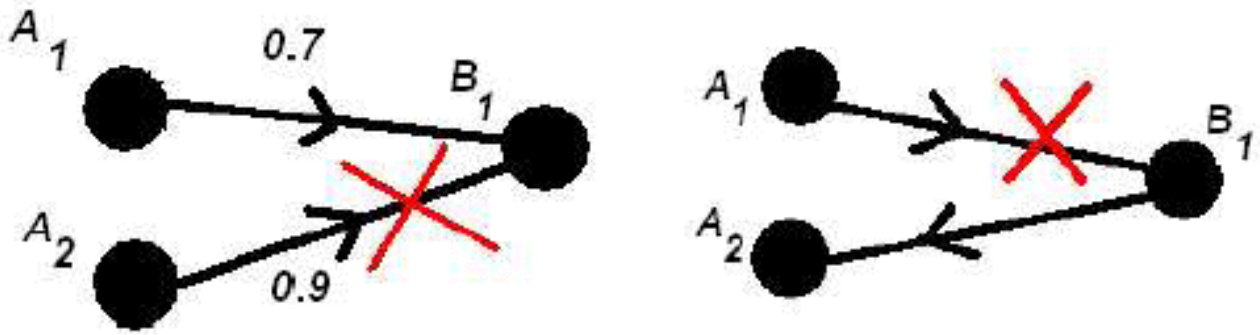
Constraints

- **Geometrical constraint** expects the matching points to appear around similar positions on the face when the normalized positions are considered.

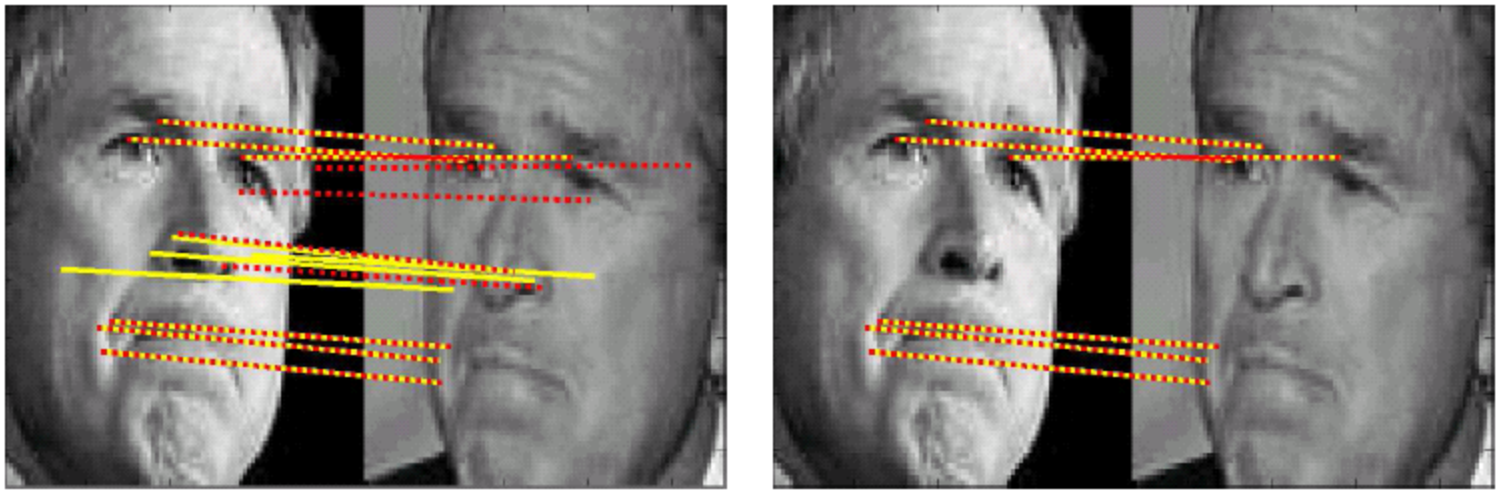


Constraints

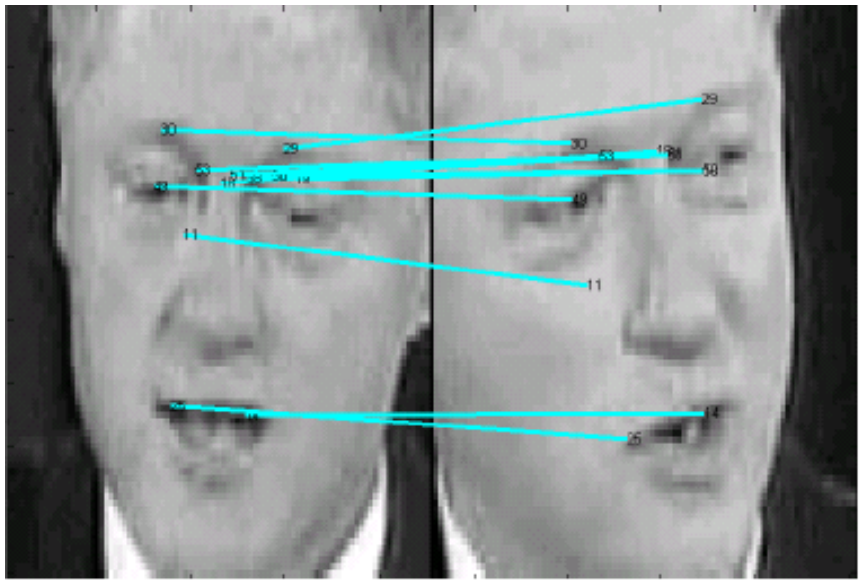
- **<u>Unique match constraint</u>** ensures that each point matches to only a single point by eliminating multiple matches to one point and also by removing one-way matches.

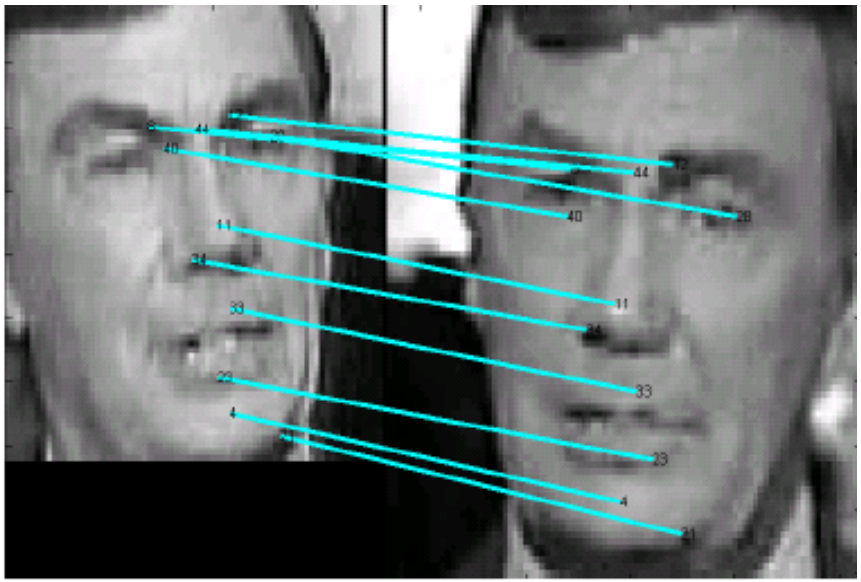


Constraints



Sample True Matches





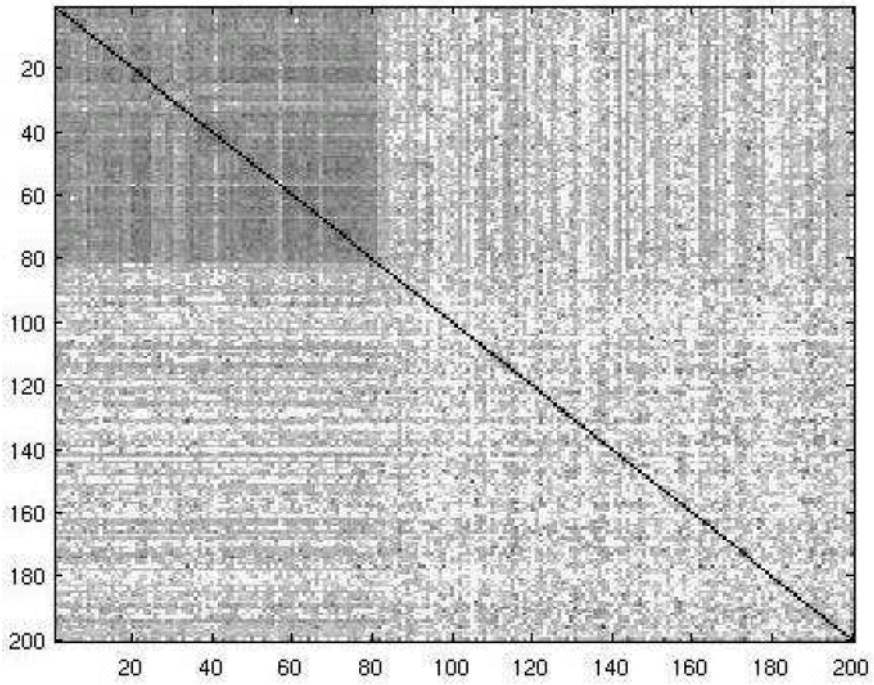
Dissimilarity Between Faces

- Distance between the two faces is defined as the average distance of all matching points between these two faces.

$$dist(A, B) = \frac{\sum_{i=1}^{N} D(i)}{N}$$

Similarity Graph

- A similarity matrix for 200 images in the search space of the name *Sam Donaldson*.
- *81 of the images are true Sam Donaldson* images, and the remaining 119 are not.



- A dissimilarity graph for all the faces in the search space is then constructed using these distances, where nodes are faces and edge weights are distances.

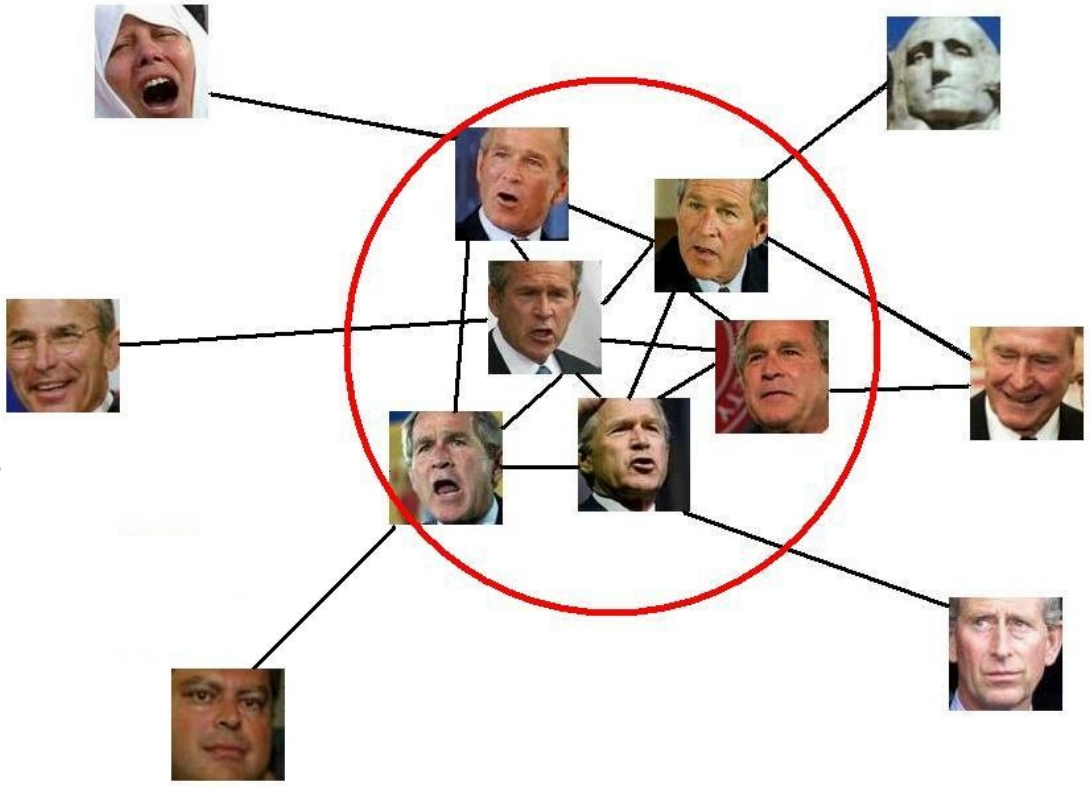
Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

Step 4: Use the result in recognizing new faces.



Step 3 : Finding Densest Component in the Graph

In [1], density of subset S of a graph G is defined as:

$$f(S) = \frac{|E(S)|}{|S|}$$

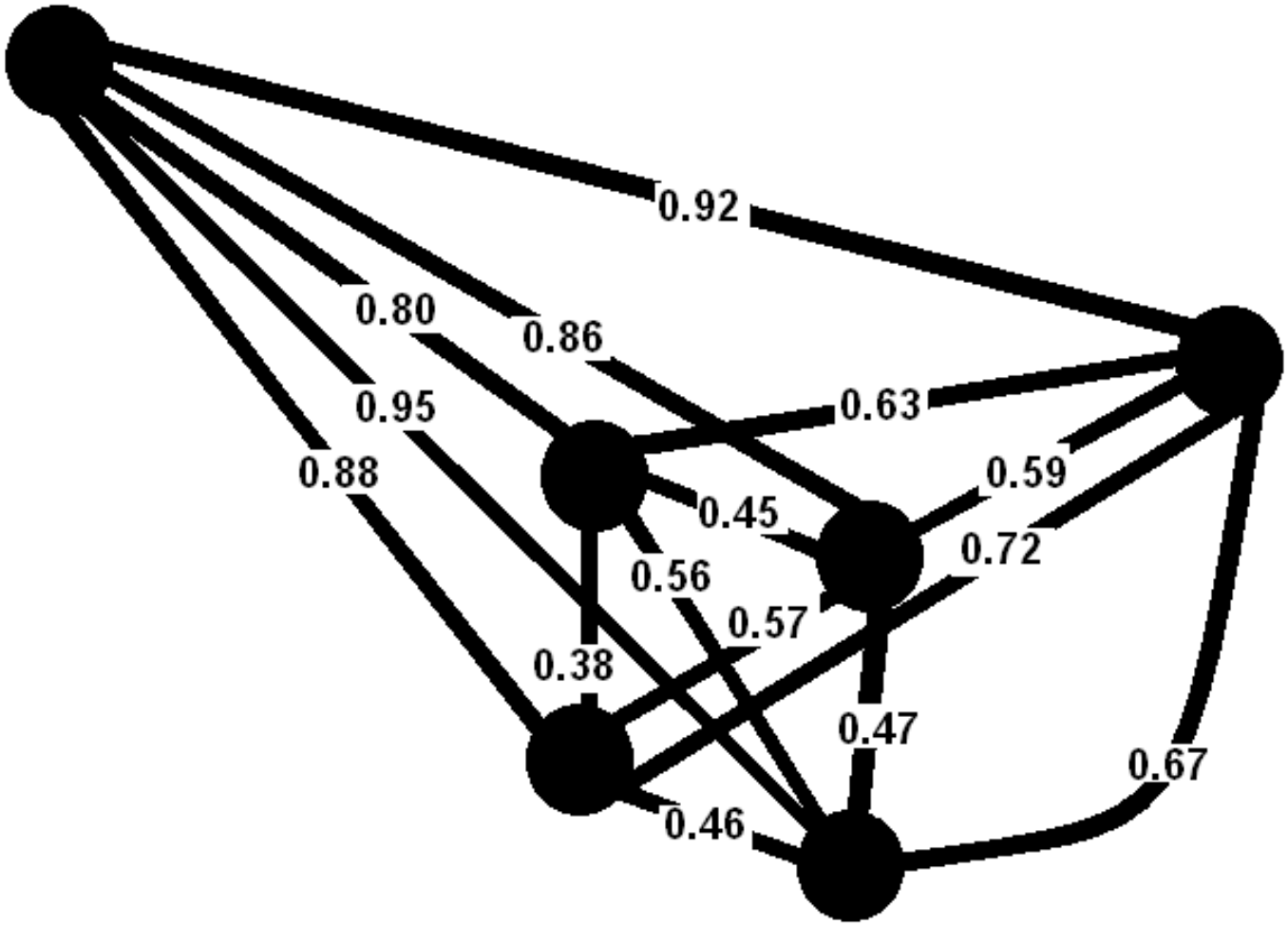
where E is the set of all edges in G and $E(S)$ is the set of edges induced by subset S.

[1] M. Charikar. 'Greedy approximation algorithms for finding dense components in a graph'. In APPROX '00: London, UK, 2000.

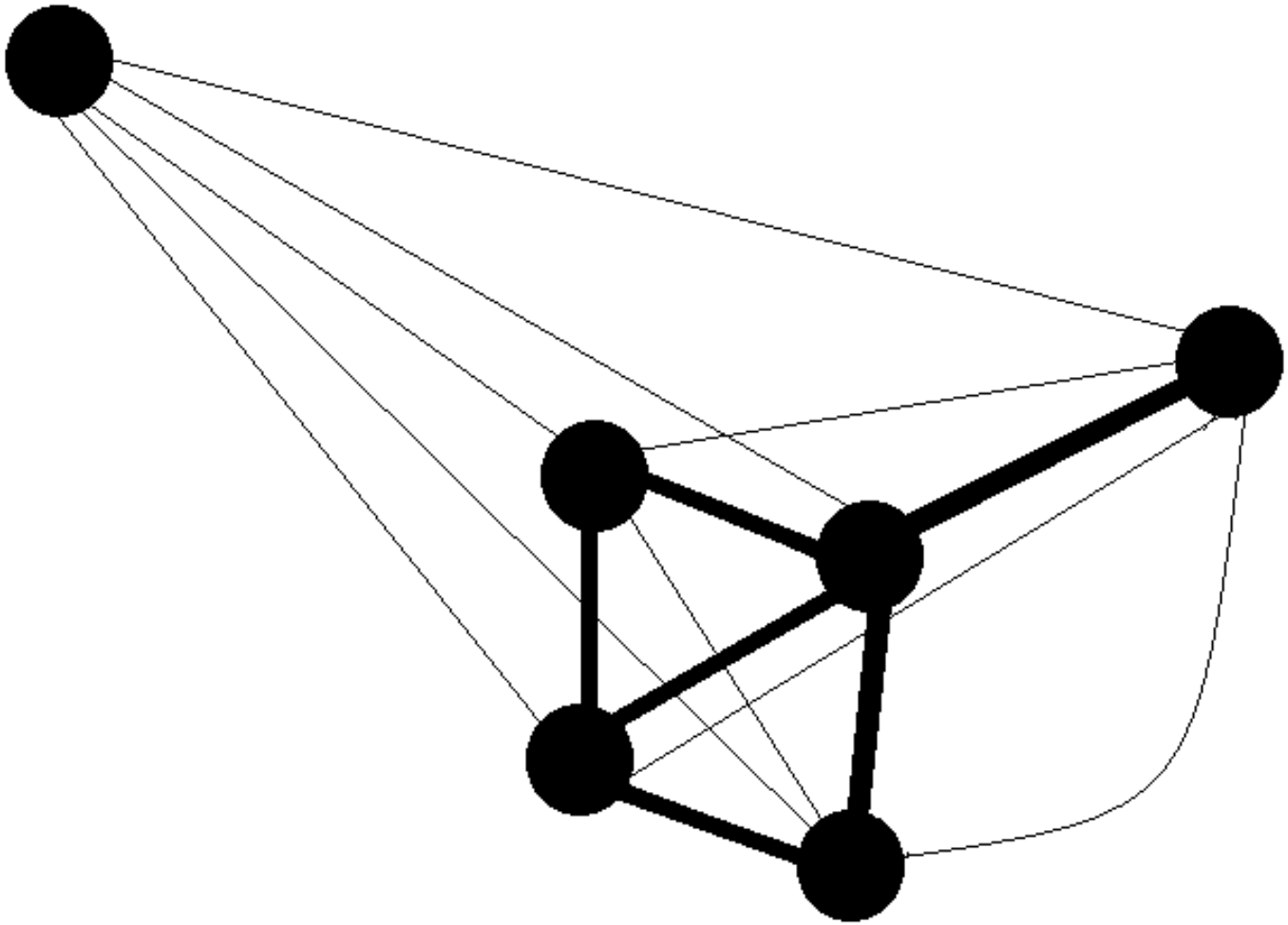
Step 3 : Finding Densest Component in the Graph

- Our goal is to find the subgraph *S* with the ***largest average degree*** that is the subgraph with the maximum density.
- Algorithm starts with the entire graph *G. (S= G)*
- In each step, the vertex with the minimum degree is removed from S and f(S) value is computed until the set S is empty.
- Finally, the set S, that has maximum f(S) value, is returned as the densest component of the graph.

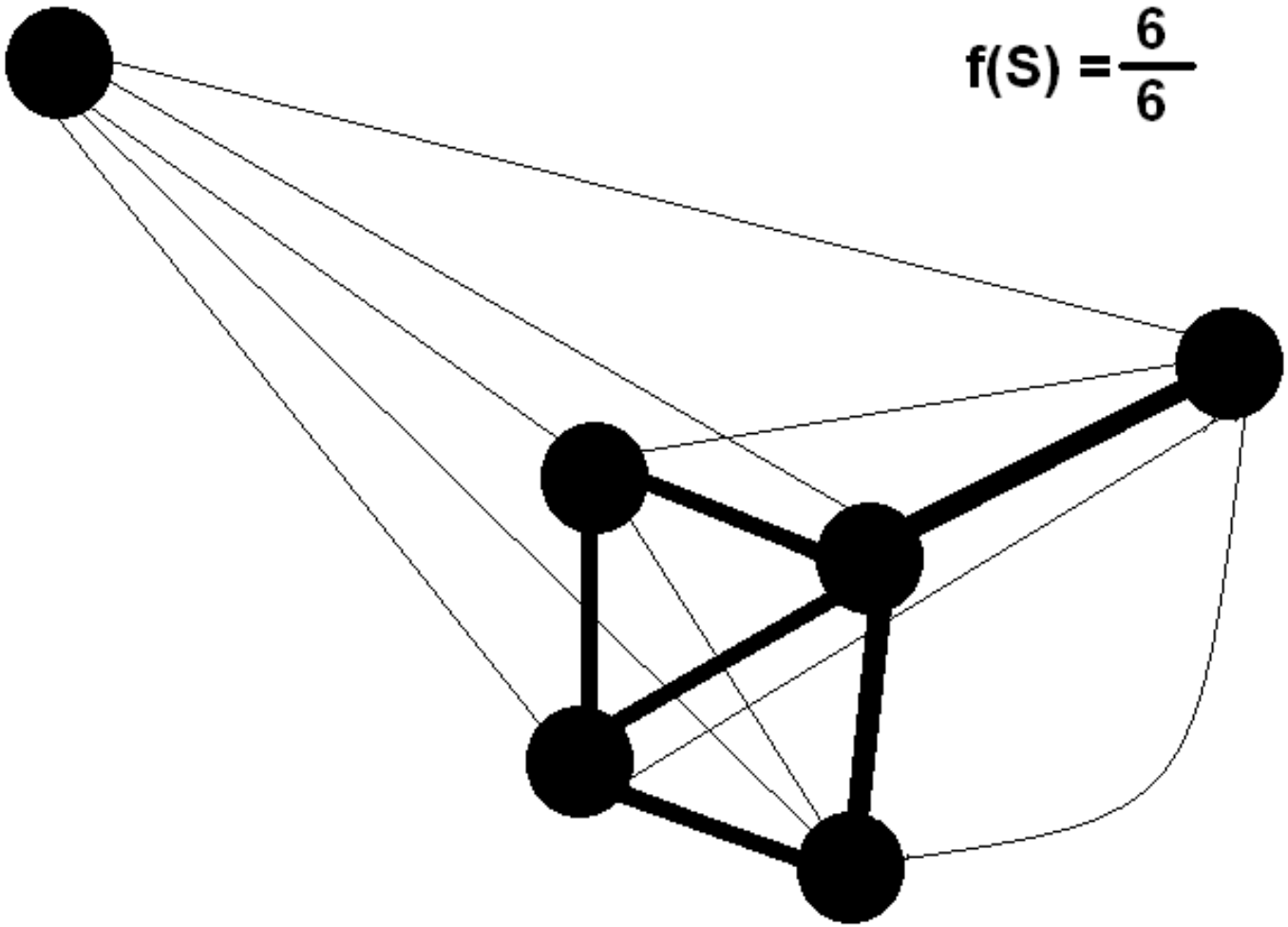
Step 3 : Finding Densest Component in the Graph



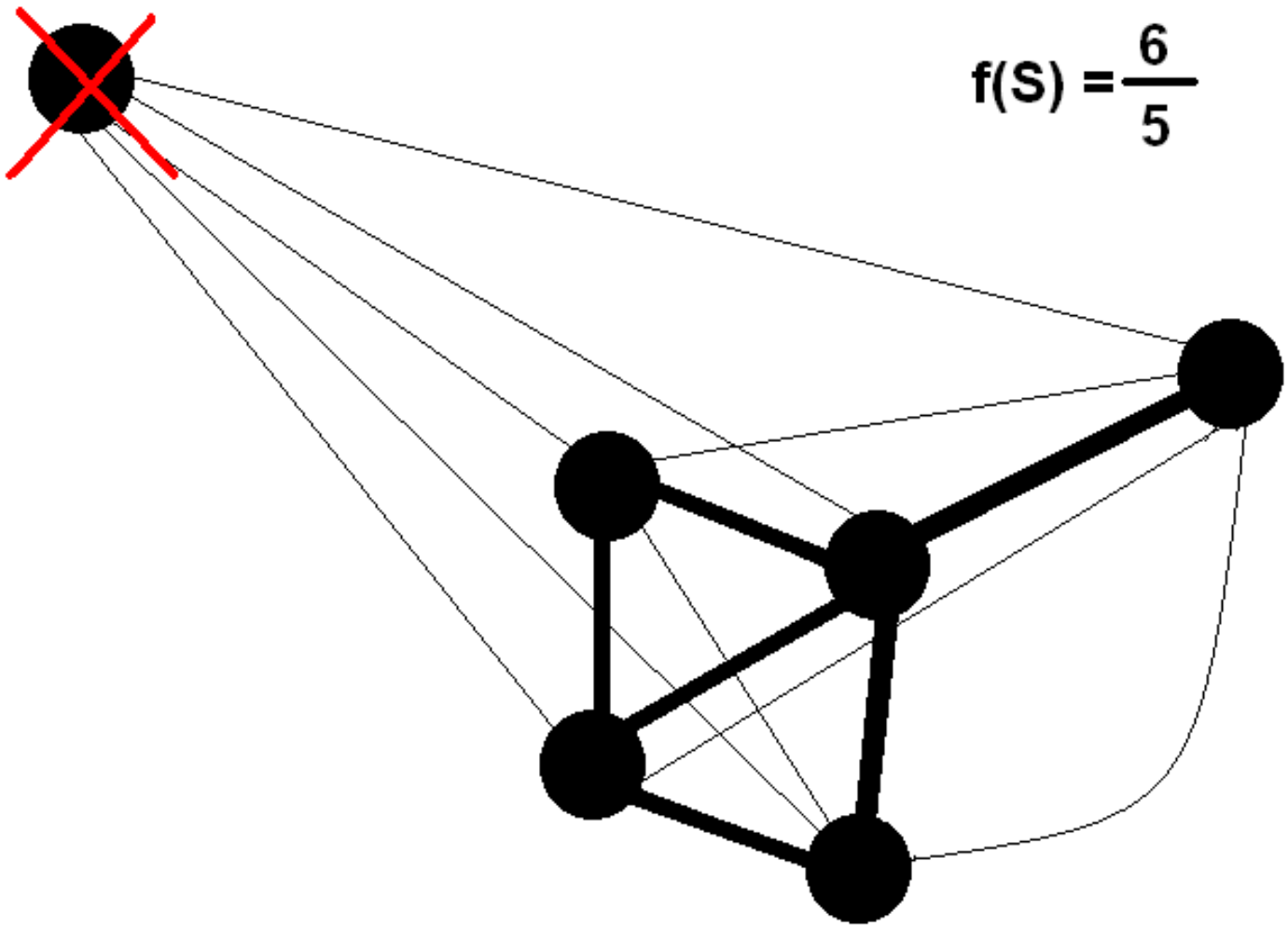
Step 3 : Finding Densest Component in the Graph



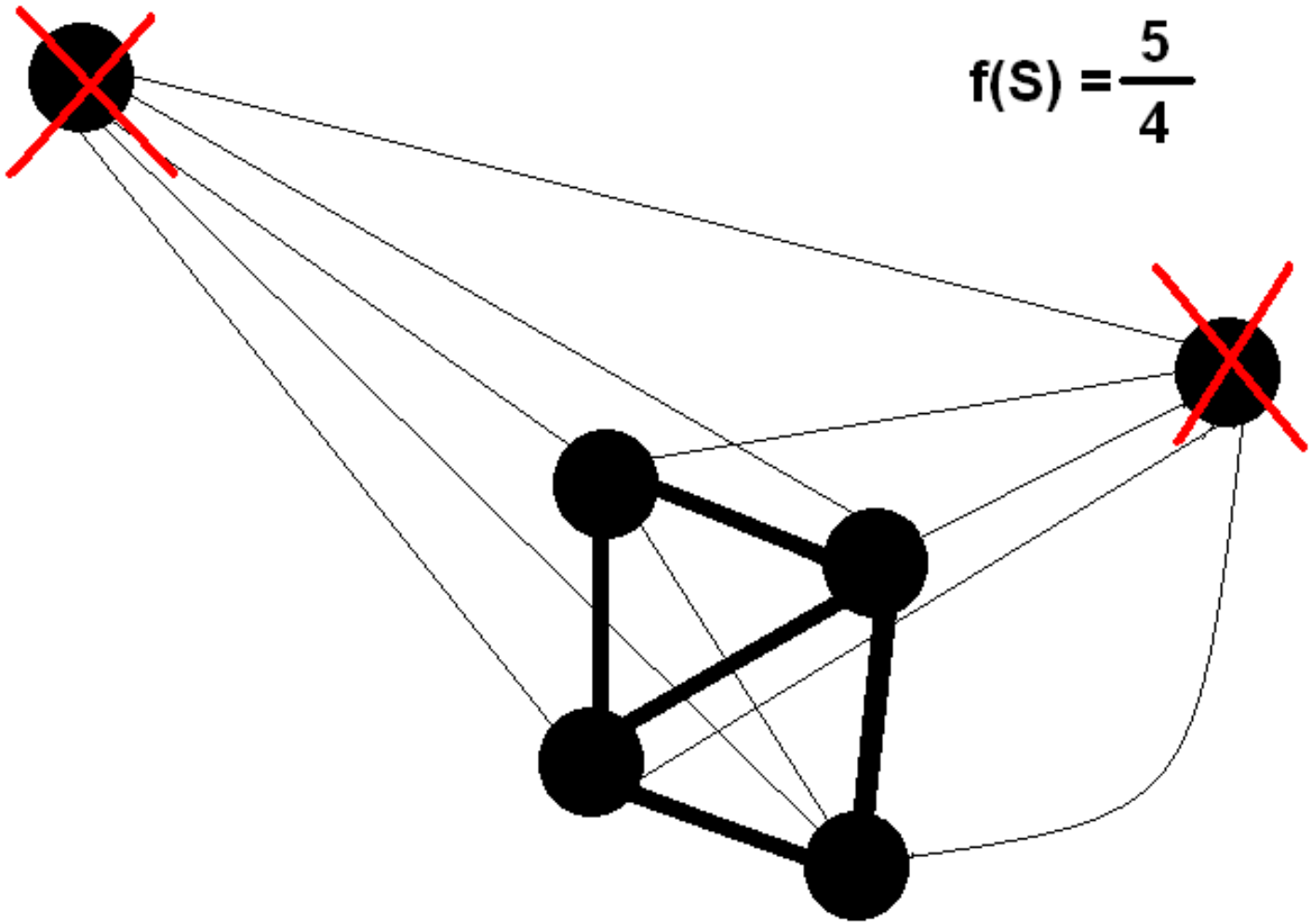
Step 3 : Finding Densest Component in the Graph



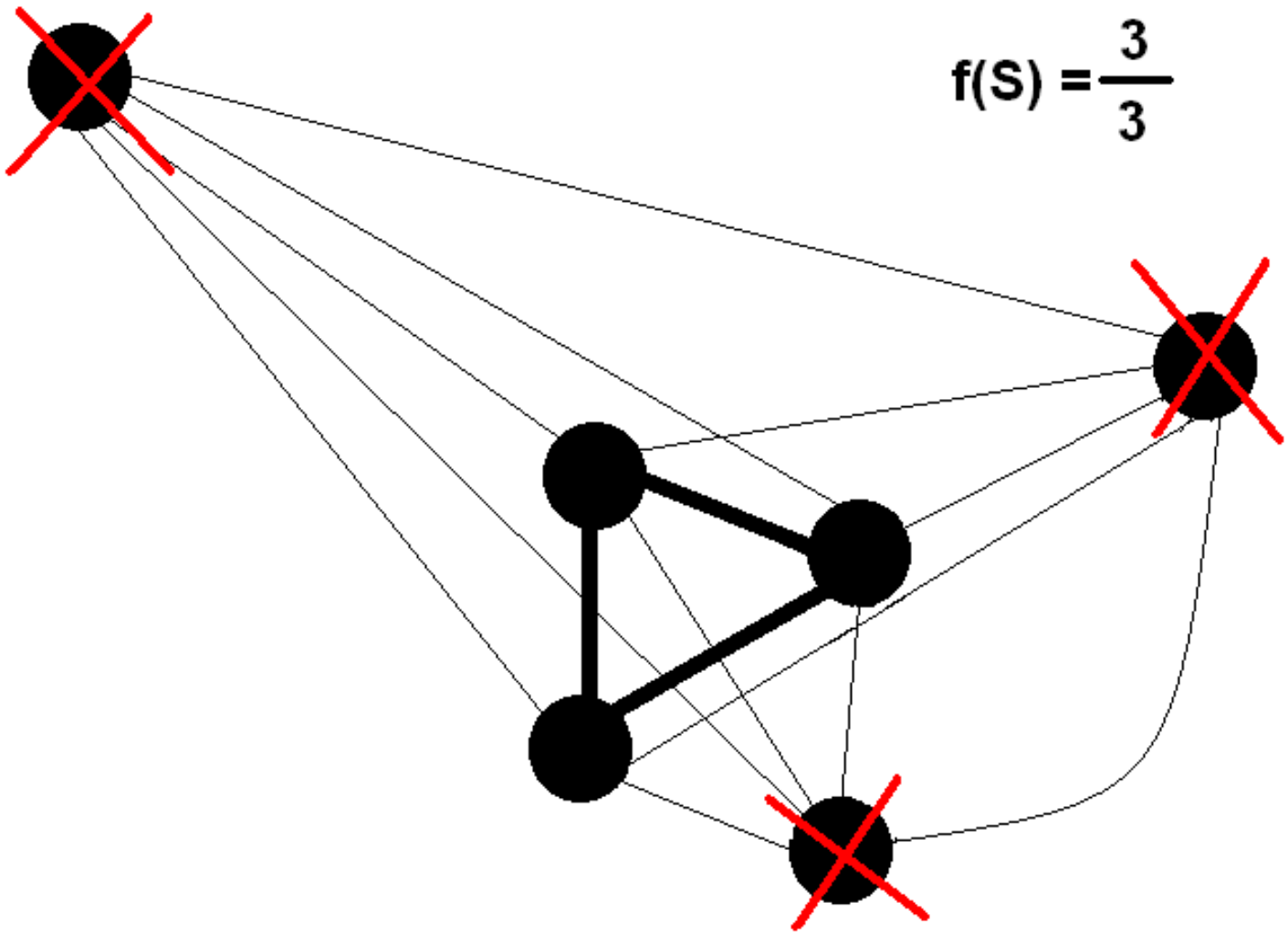
Step 3 : Finding Densest Component in the Graph



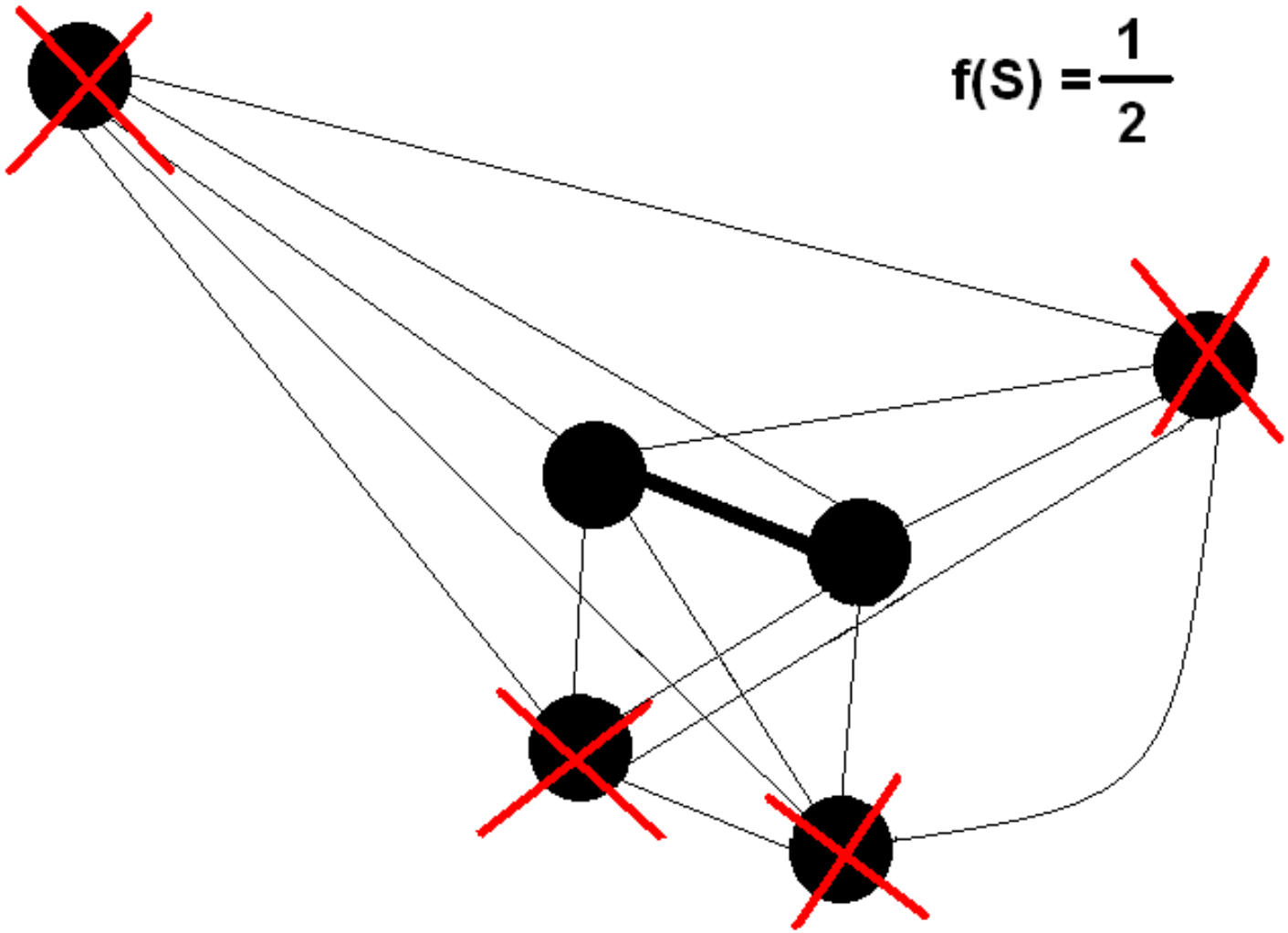
Step 3 : Finding Densest Component in the Graph



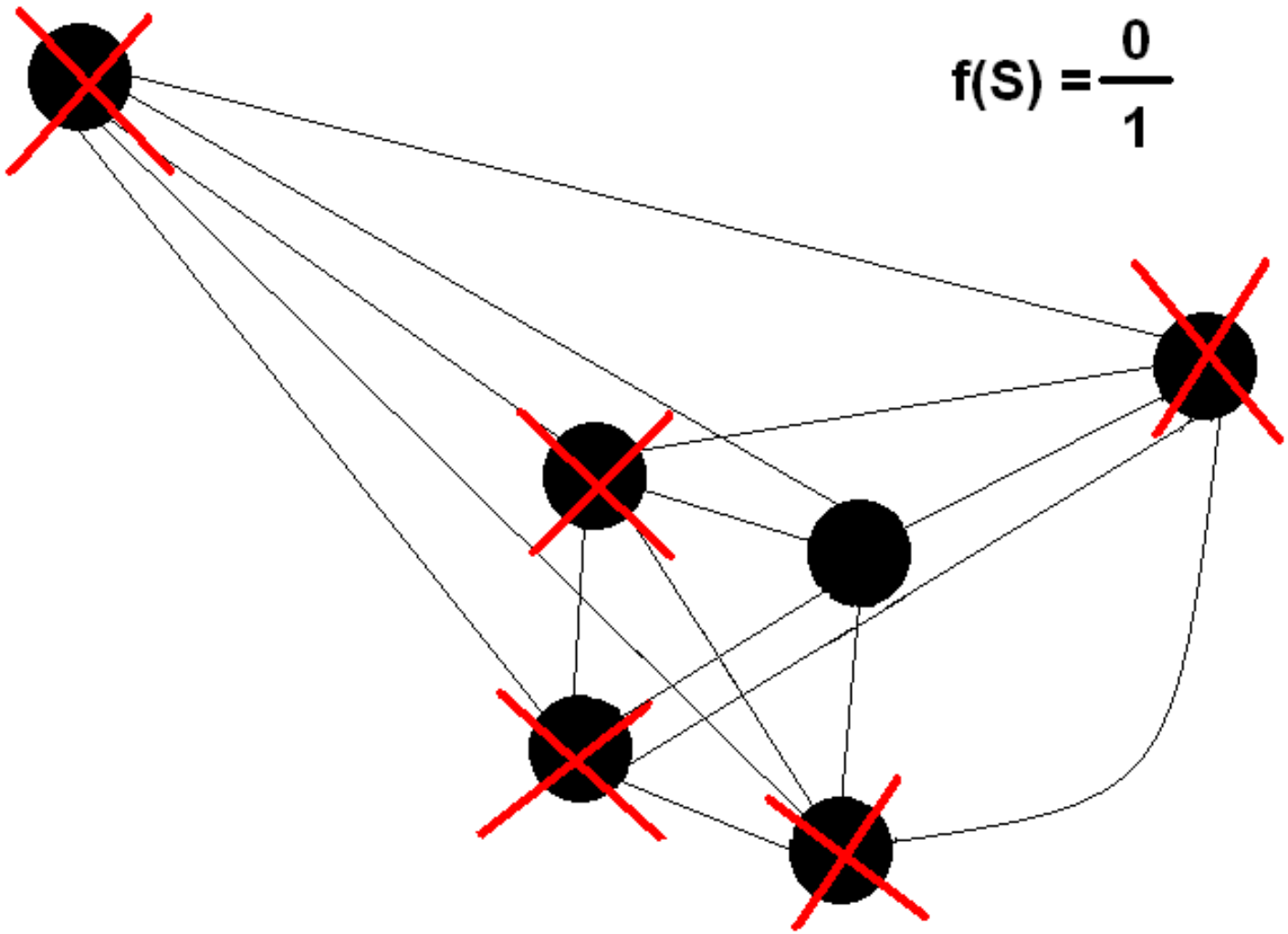
Step 3 : Finding Densest Component in the Graph



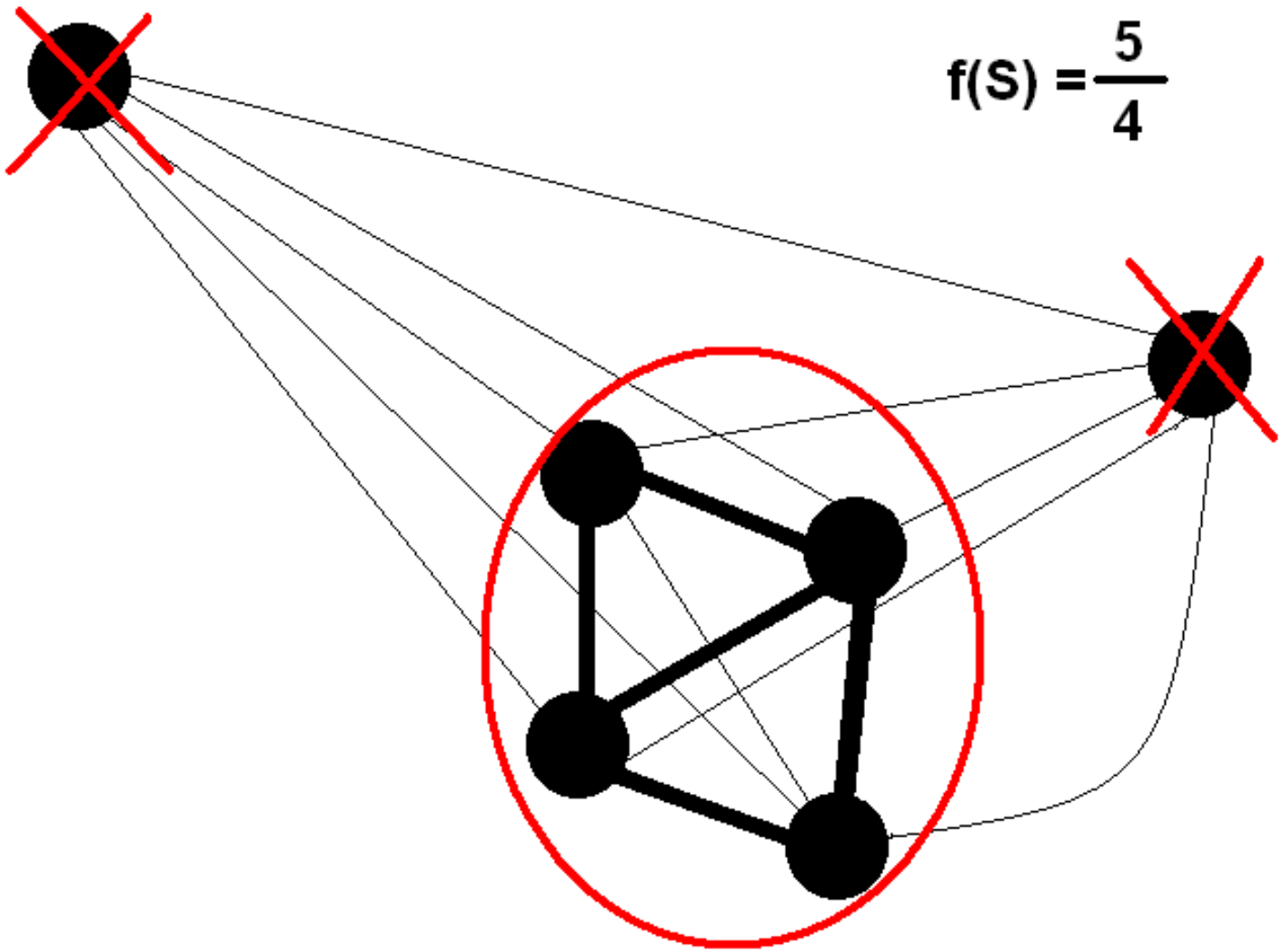
Step 3 : Finding Densest Component in the Graph



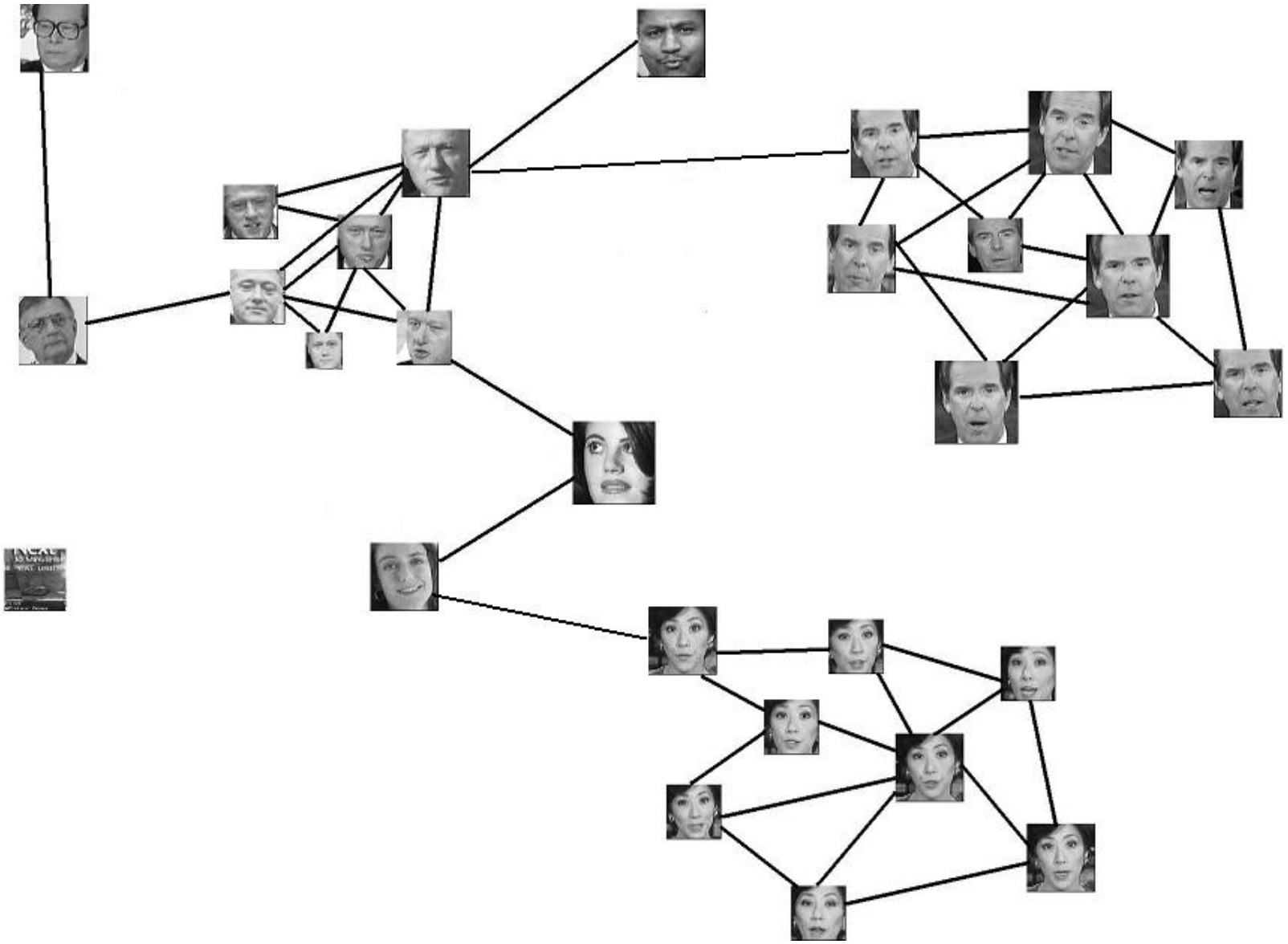
Step 3 : Finding Densest Component in the Graph



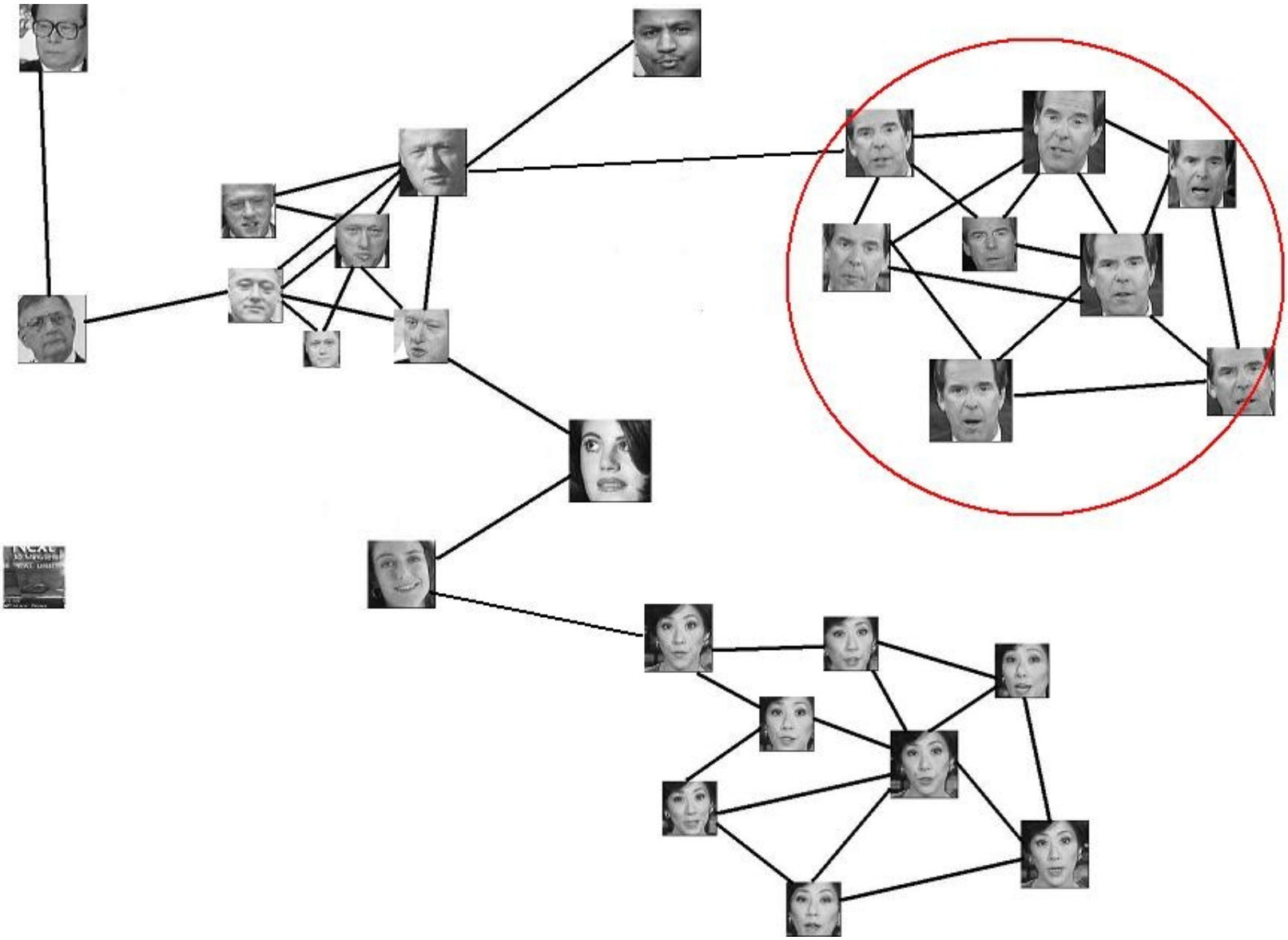
Step 3 : Finding Densest Component in the Graph



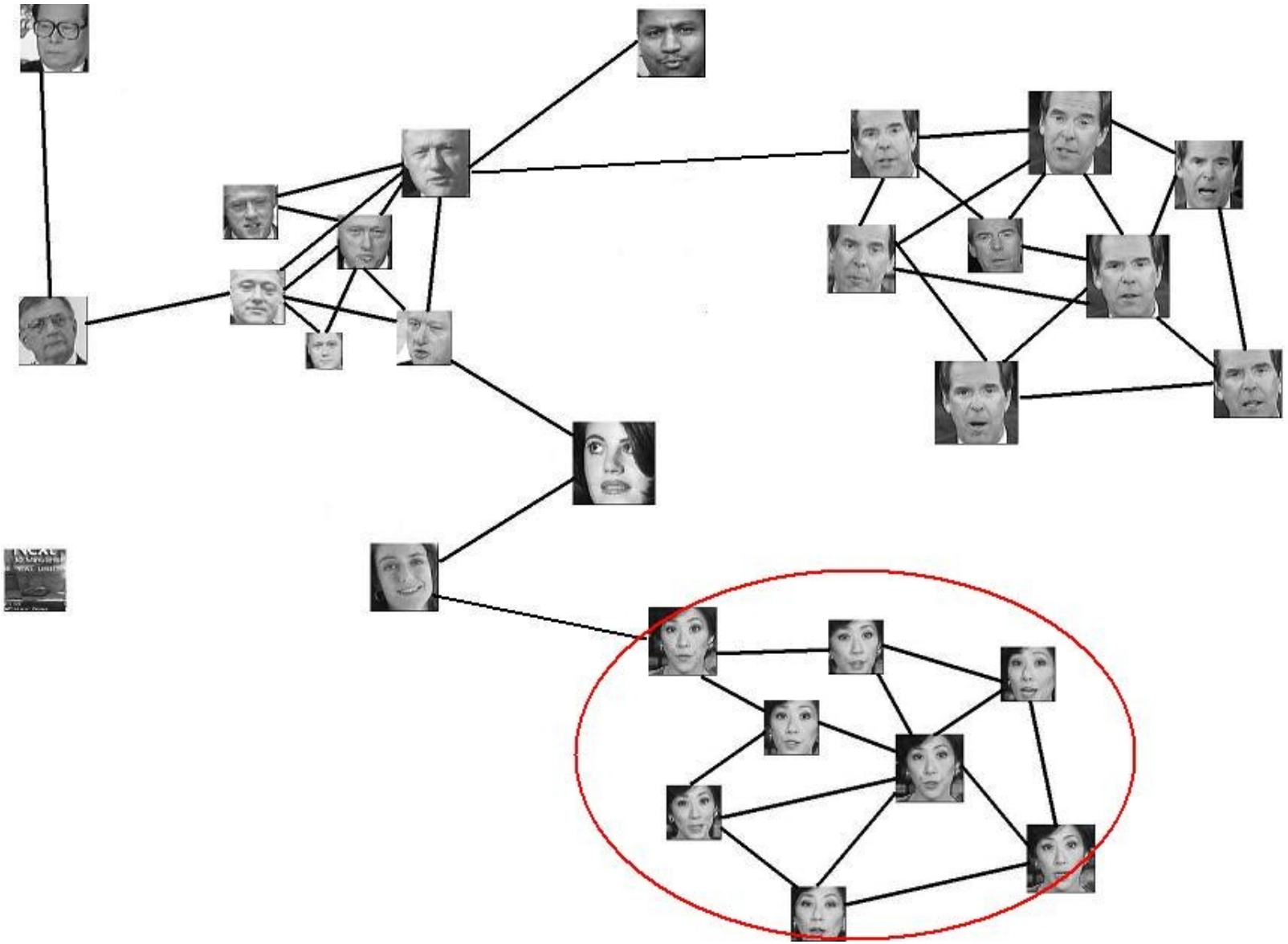
Problem: Anchor person



Problem: Anchor person

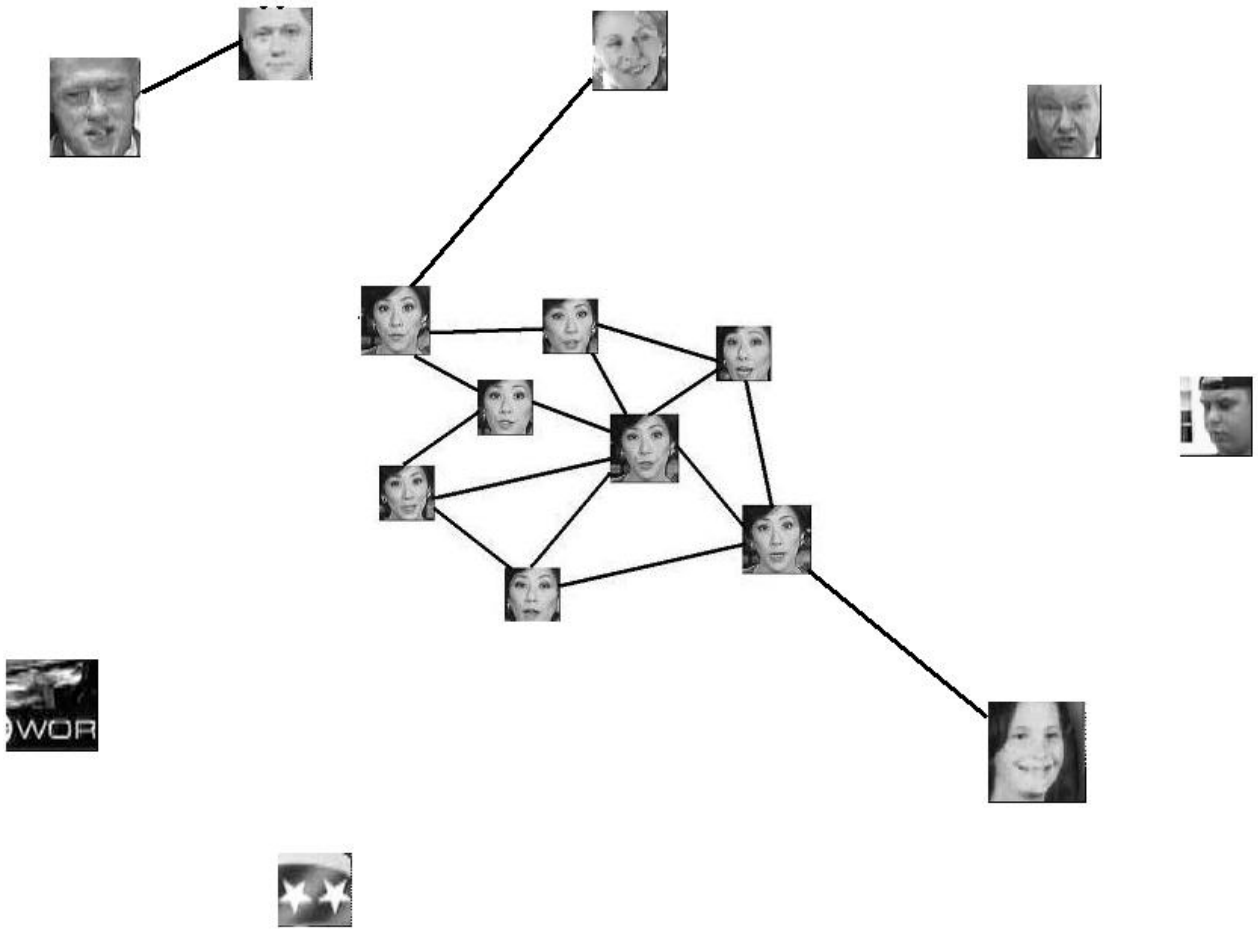


Problem: Anchor person



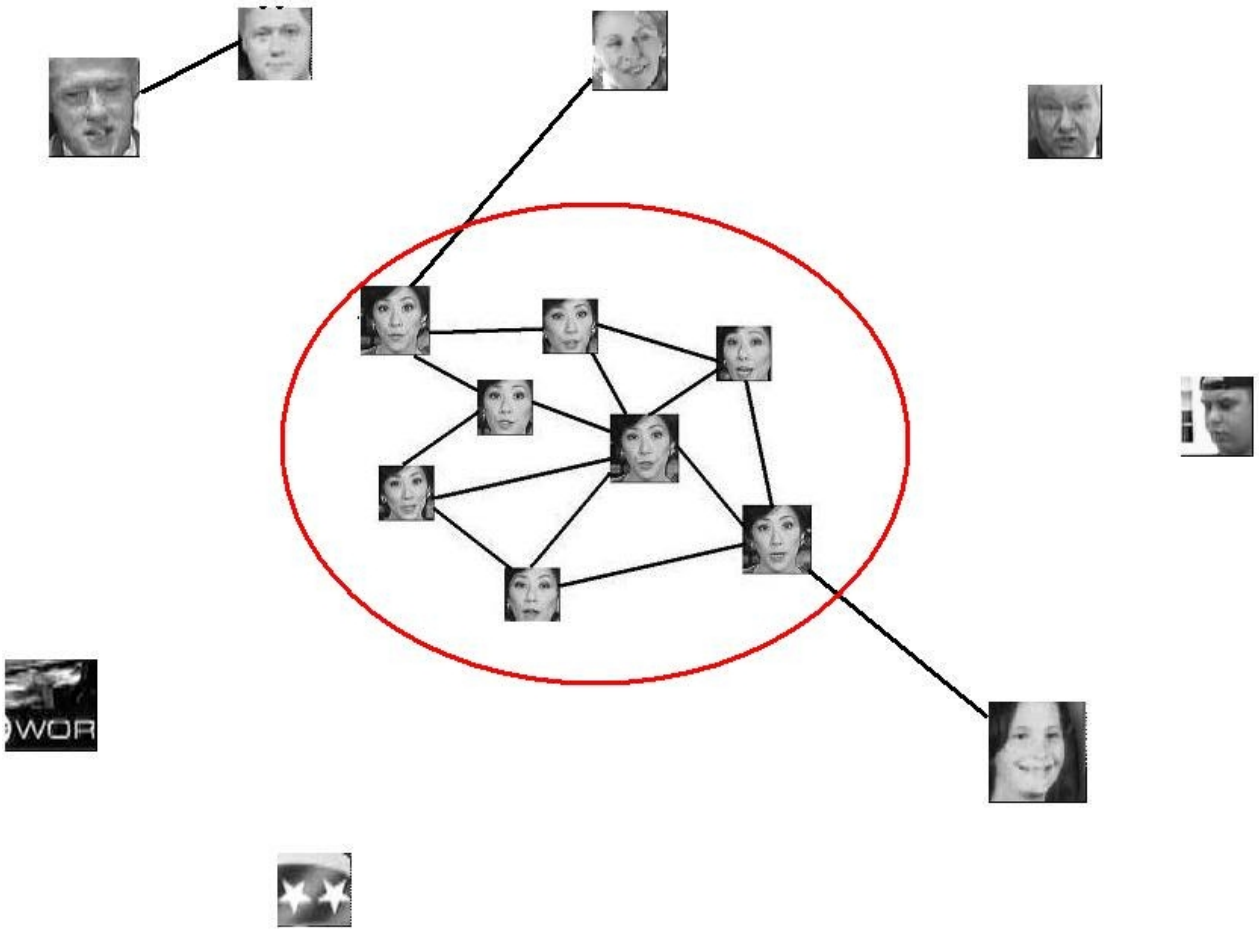
Solution: Anchor person detection

- Construct similarity matrix for each news video.



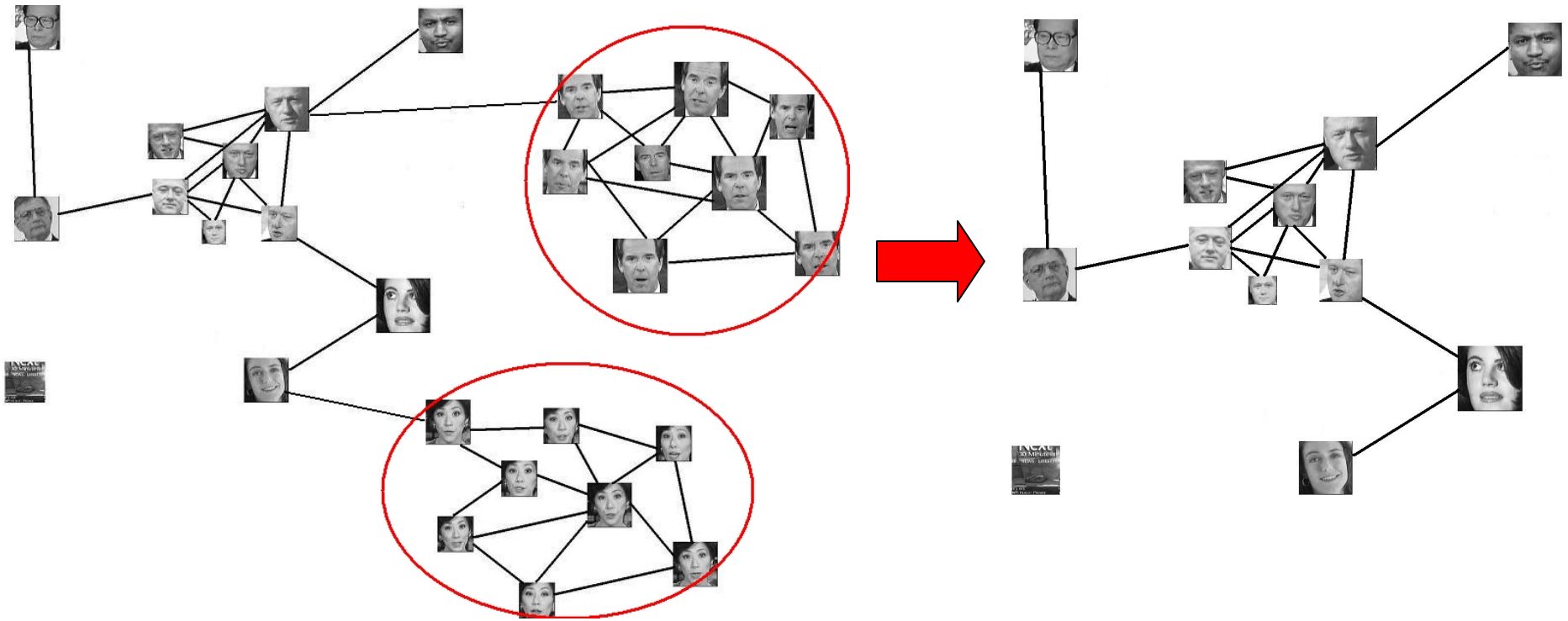
Solution: Anchor person detection

- Find the densest component which corresponds to the anchor person.



Solution: Anchor person detection

- Remove detected anchor persons of all videos from the search space of queried person.

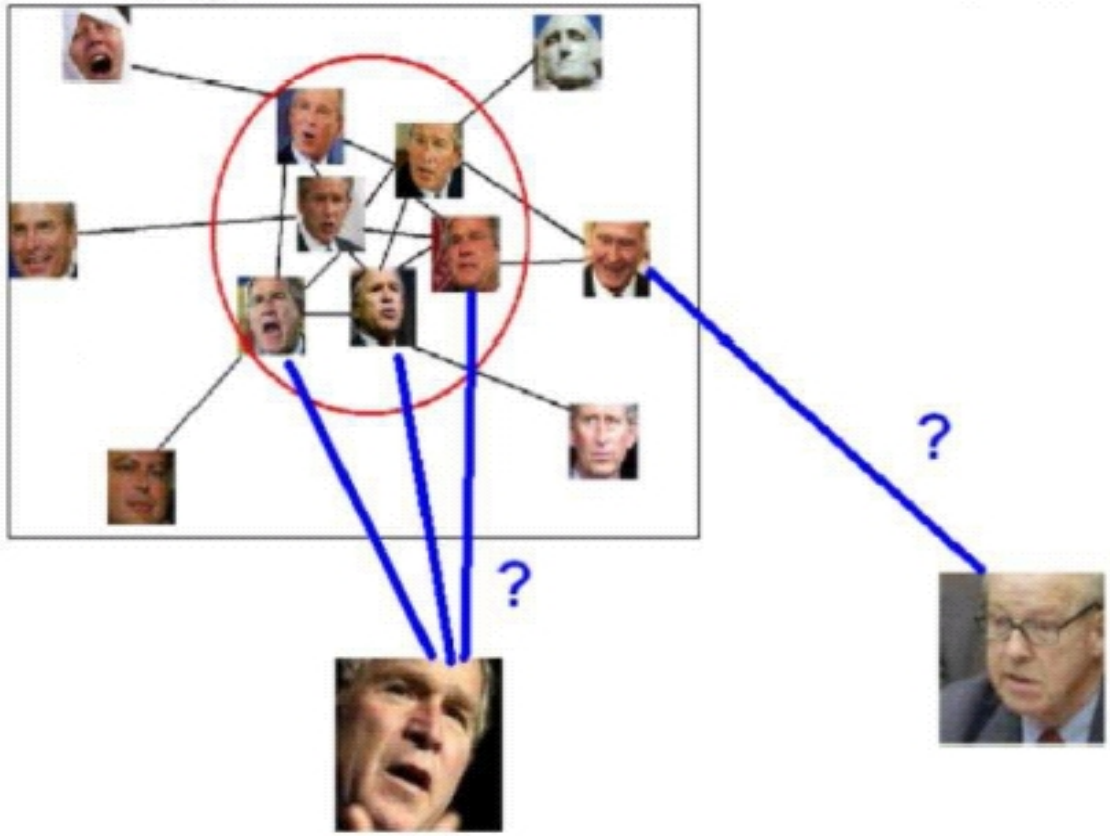


Proposed Approach

Step 1: Limit the search space of a query person by using name

Step 2: Construct a similarity graph among faces in this search space.

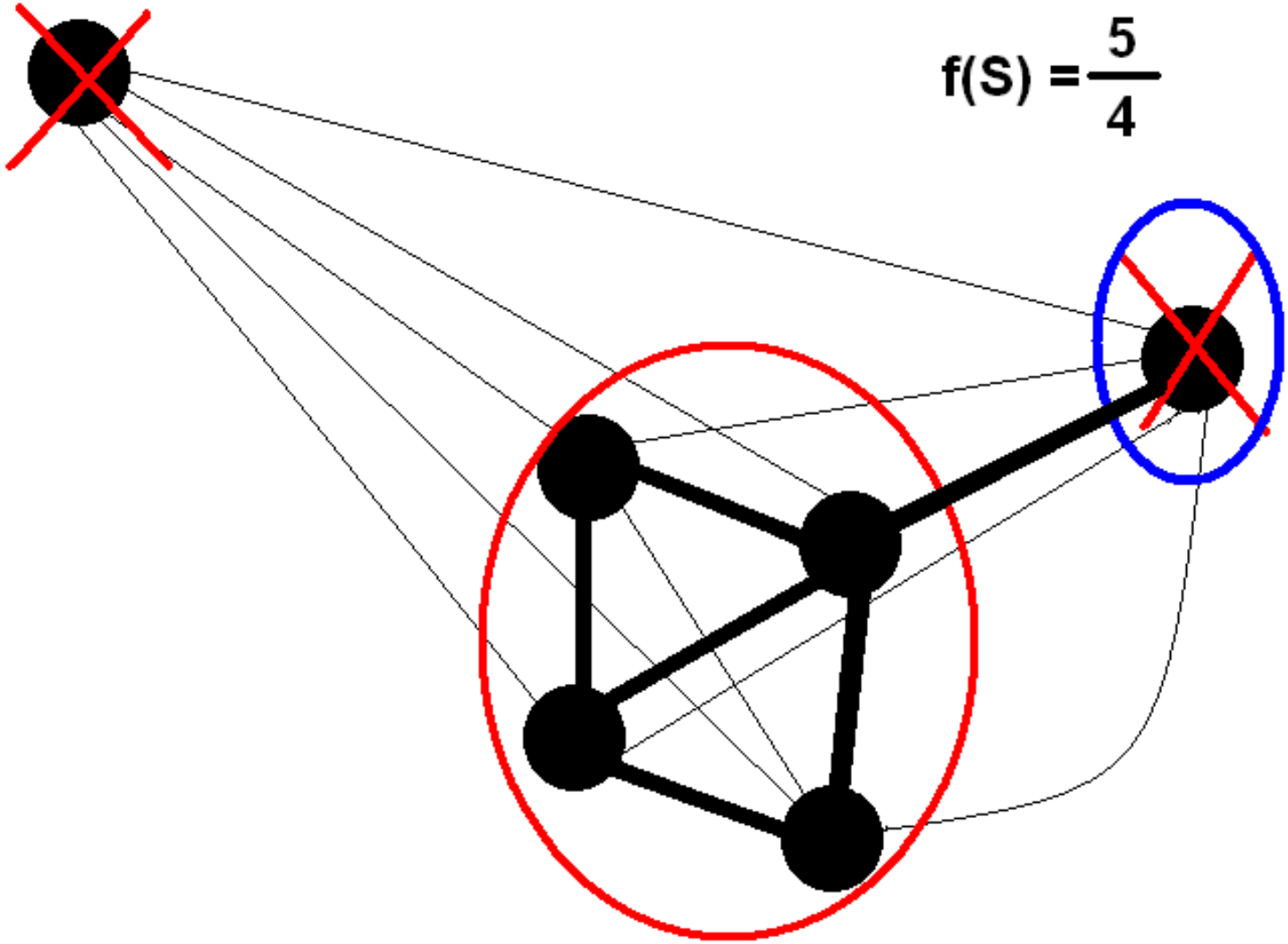
Step 3: Find the densest component in this graph.



Step 4: Use the result in recognizing new faces.

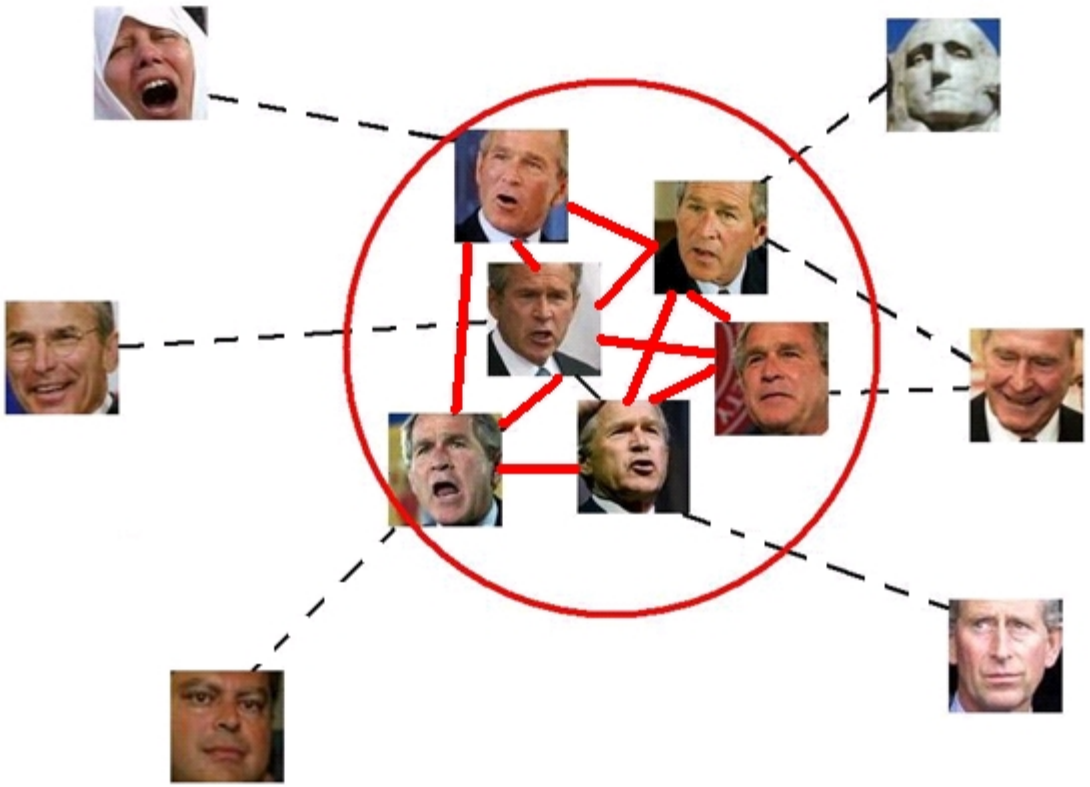
Step 4: Online Recognition

1. Degree Modeling



Step 4: Online Recognition

2. Distance Modeling



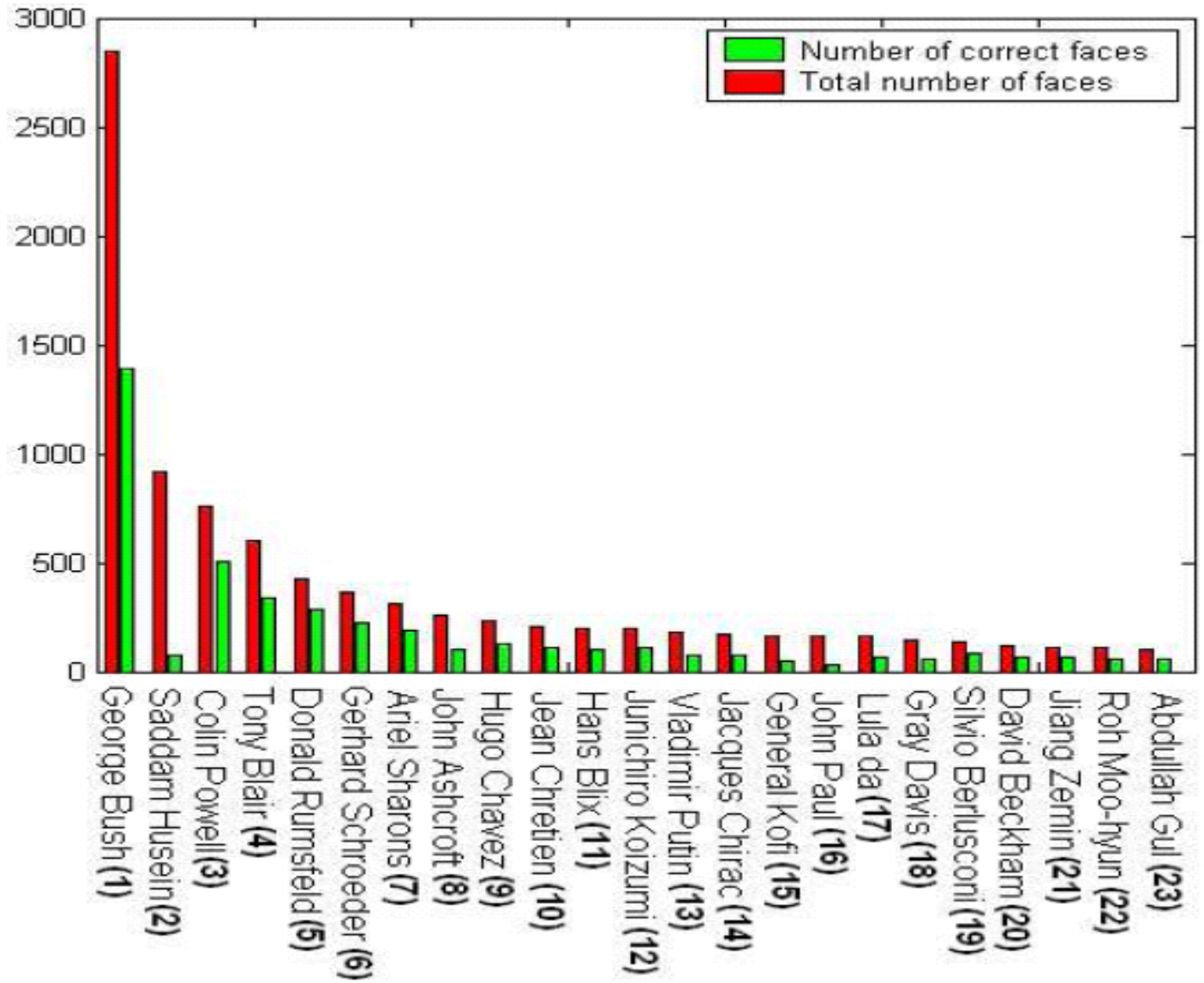
Data Set

- **News Photographs**
  - about half a million captioned news images from Yahoo! News on the web collected by Berg et al. [2].
  - 30,281 detected faces
  - 13,292 different names are used for association.

- In the experiments, the top 23 people appearing with the highest frequencies (more than 200 times)

- [2] T. Berg, A. C. Berg, J. Edwards, and D.A. Forsyth. Who's in the picture. In *Neural Information Processing Systems (NIPS), 2004.*

Data Set


Number of correct faces
Total number of faces
3000
2500
2000
1500
1000
500
0
George Bush (1)
Saddam Husein (2)
Colin Powell (3)
Tony Blair (4)
Donald Rumsfeld (5)
Gerhard Schroeder (6)
Ariel Sharons (7)
John Ashcroft (8)
Hugo Chavez (9)
Jean Chretien (10)
Hans Blix (11)
Junichiro Koizumi (12)
Vladimir Putin (13)
Jacques Chirac (14)
General Kofi (15)
John Paul (16)
Lula da (17)
Gray Davis (18)
Silvio Berlusconi (19)
David Beckham (20)
Jiang Zemin (21)
Roh Moo-hyun (22)
Abdullah Gul (23)

Data Set

- **News Videos**
  - Broadcast news videos provided by NIST for TRECVID video retrieval evaluation competition 2004.
  - 229 videos (30 minutes each) from ABC and CNN news.
  - The shot boundaries & key-frames are provided by NIST.
  - Speech transcripts extracted by LIMSI are used to obtain the associated text for each shot.
  - The face detection algorithm provided by Mikolajcyzk is used to extract faces from key-frames. 57% accuracy is recorded for randomly selected 10 videos.

Data Set

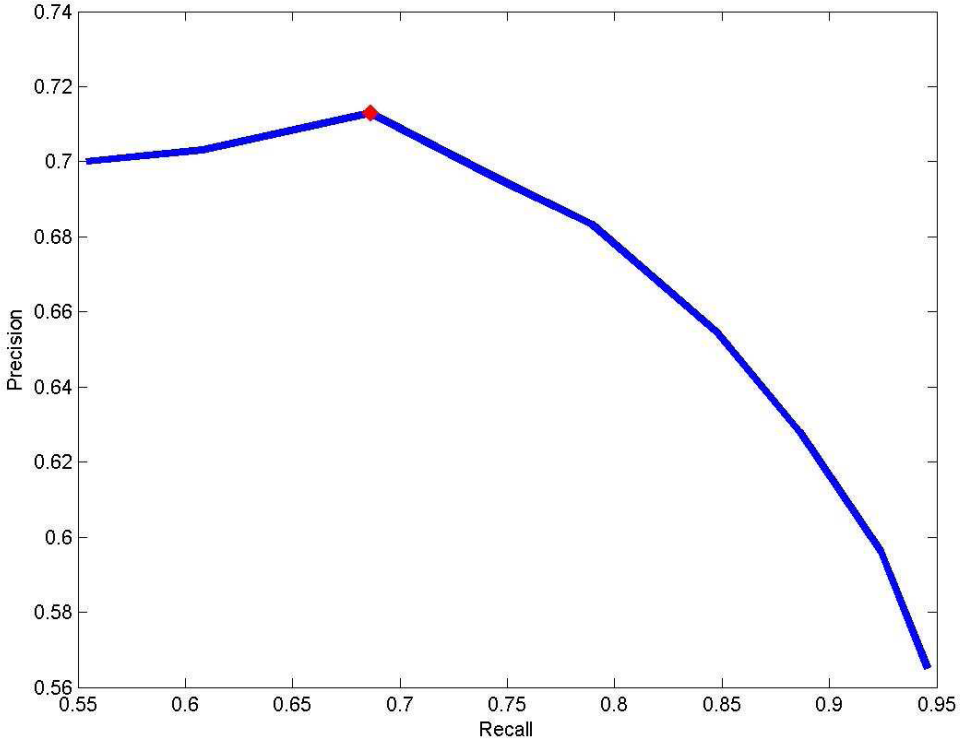
- **News Videos**
- 5 people are used:
  - Bill Clinton, (991)
  - Benjamin Netanyahu, (51)
  - Sam Donaldson, (100)
  - Saddam Hussein, (149)
  - Boris Yeltsin, (78)

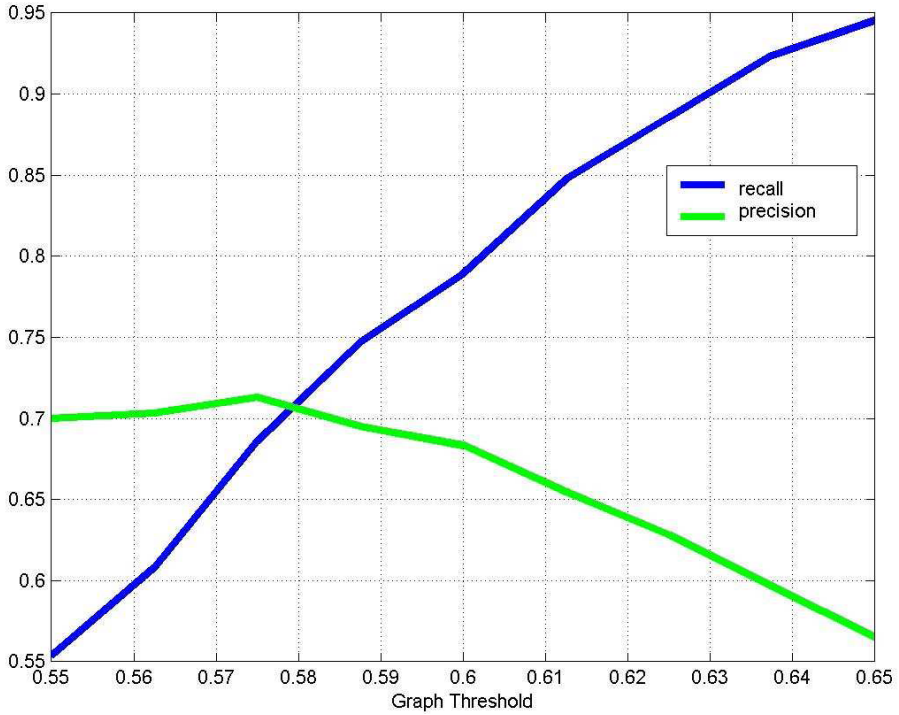
Experimental Results on News Photographs

- **Matching Points**
  - After the constraints, 73% of all possible true matches are kept and we lose only 27% of true matches.
  - Among these assignments, we achieved a correct matching rate of 72%.

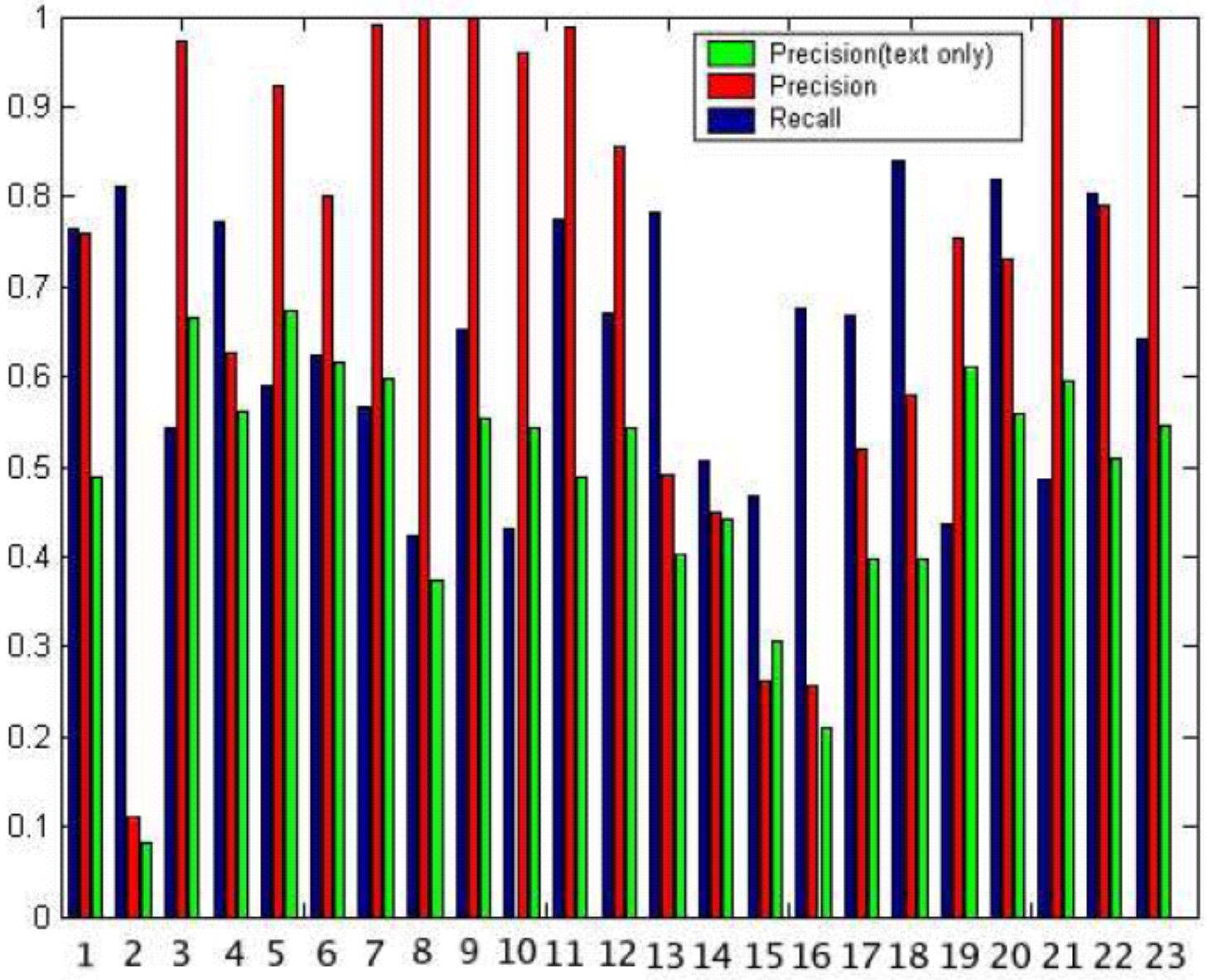
Experimental Results on News Photographs

- **Graph Approach**





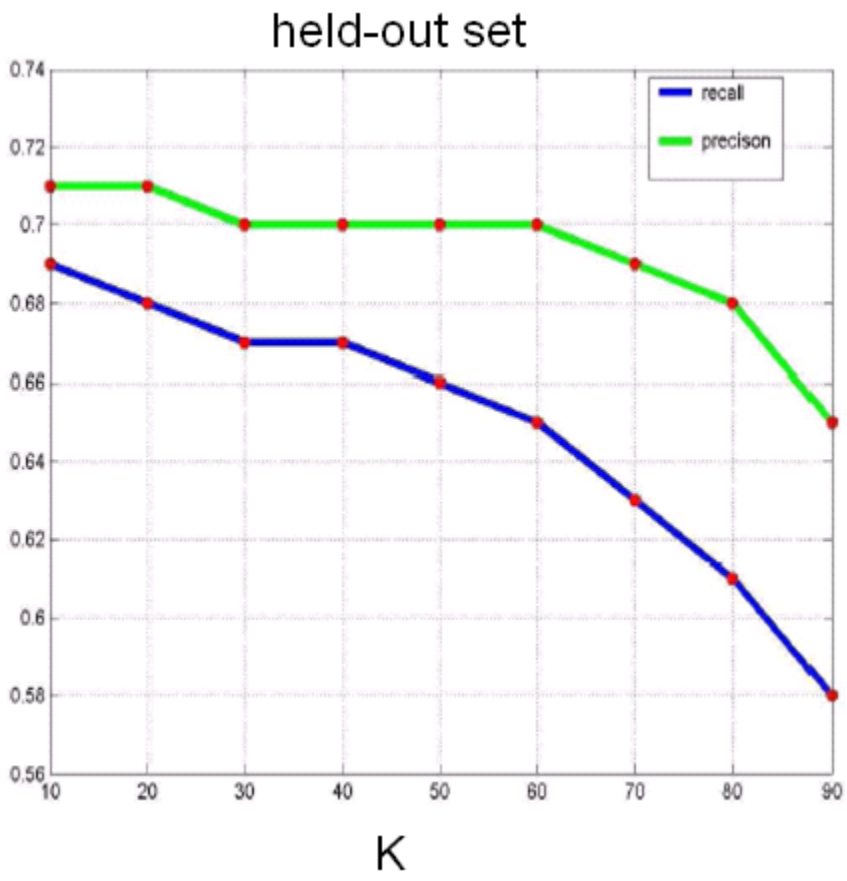
Experimental Results on News Photographs

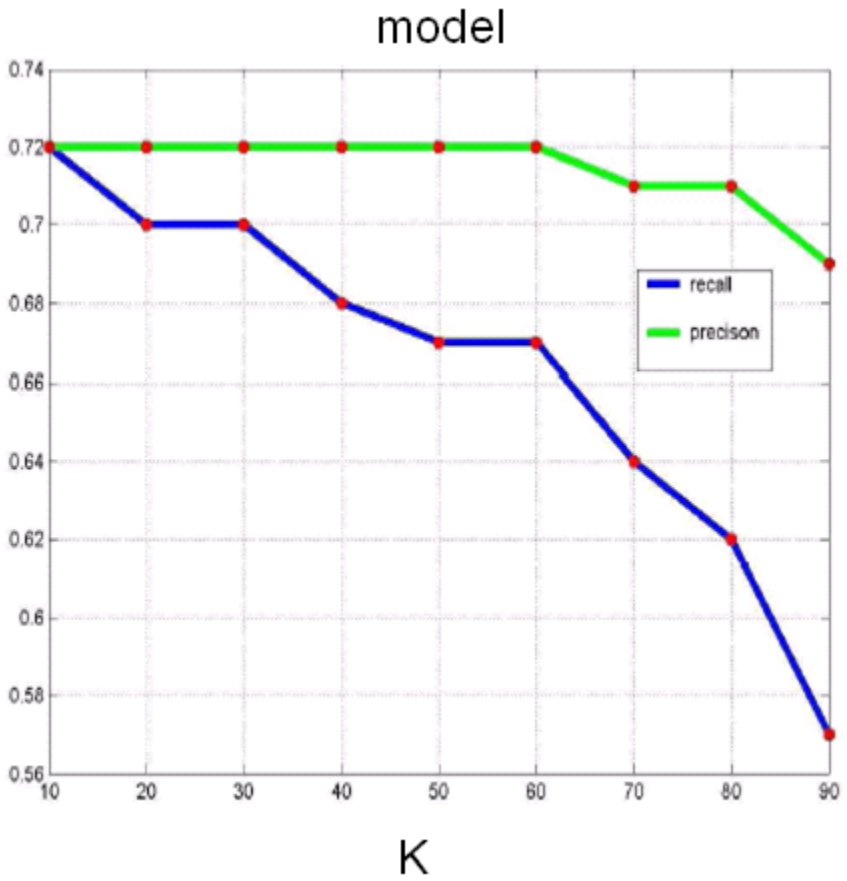


Experimental Results on News Photographs

Experimental Results on News Photographs

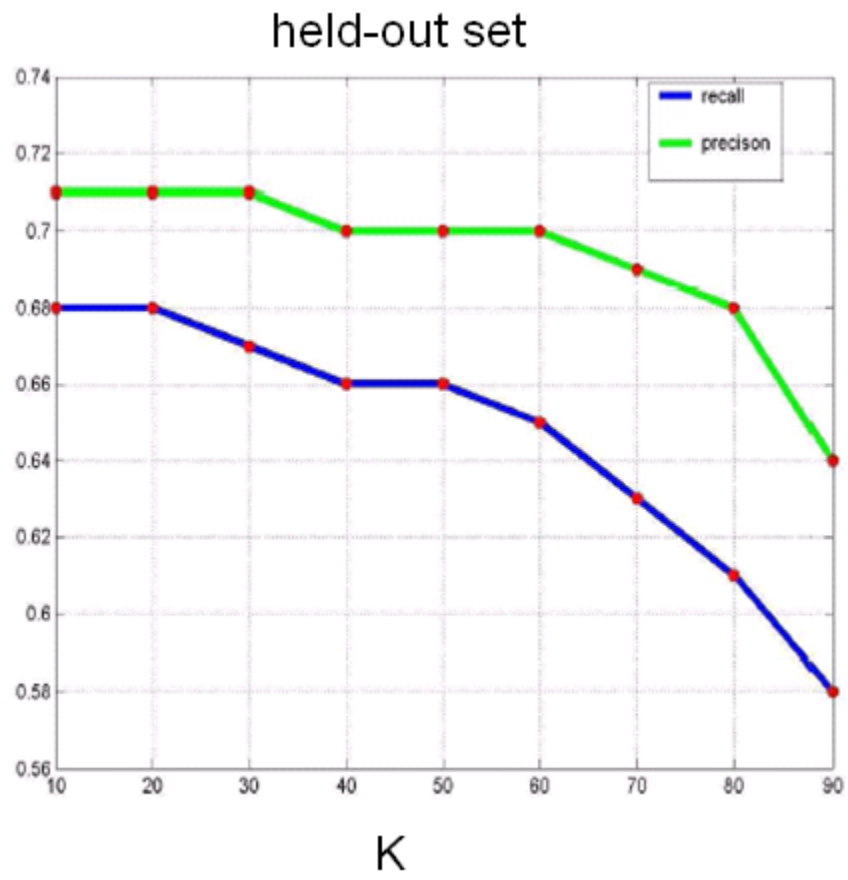
- **Online Recognition: Distance Modeling**

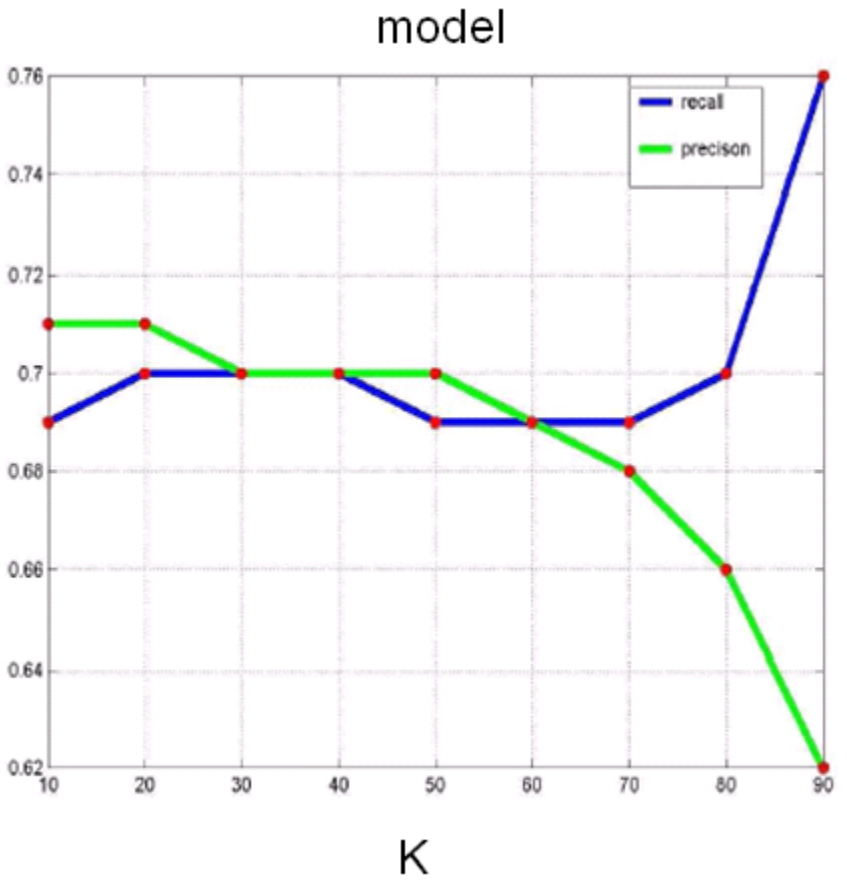




Experimental Results on News Photographs

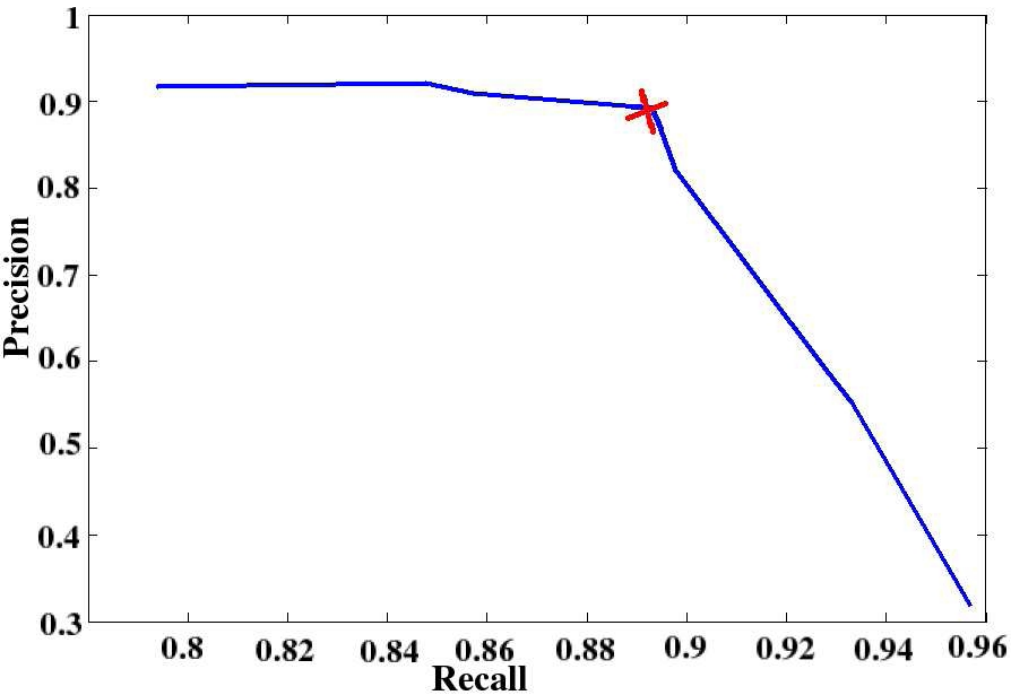
- **Online Recognition: Degree Modeling**

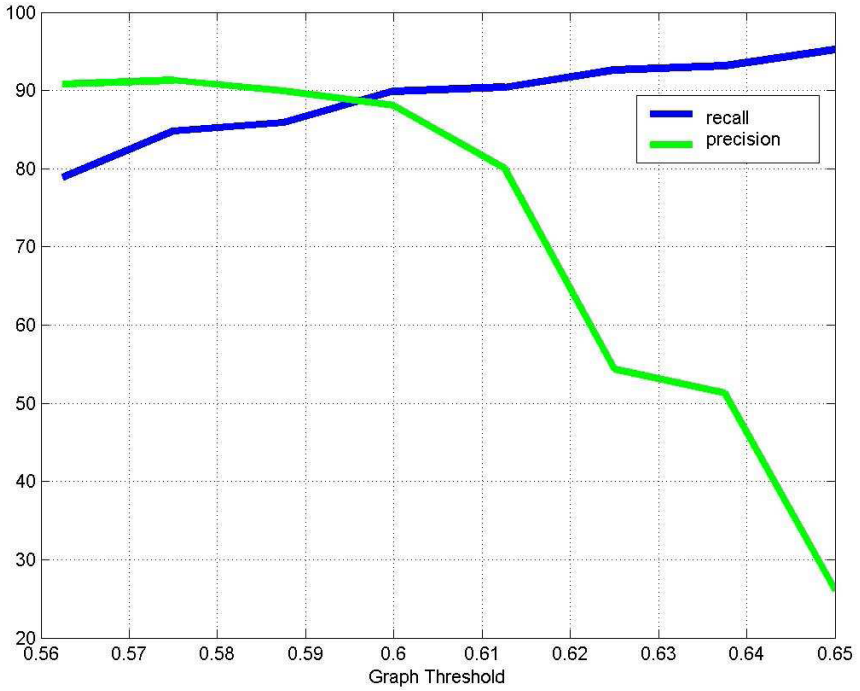




Experimental Results on News Videos

- **Graph Approach**

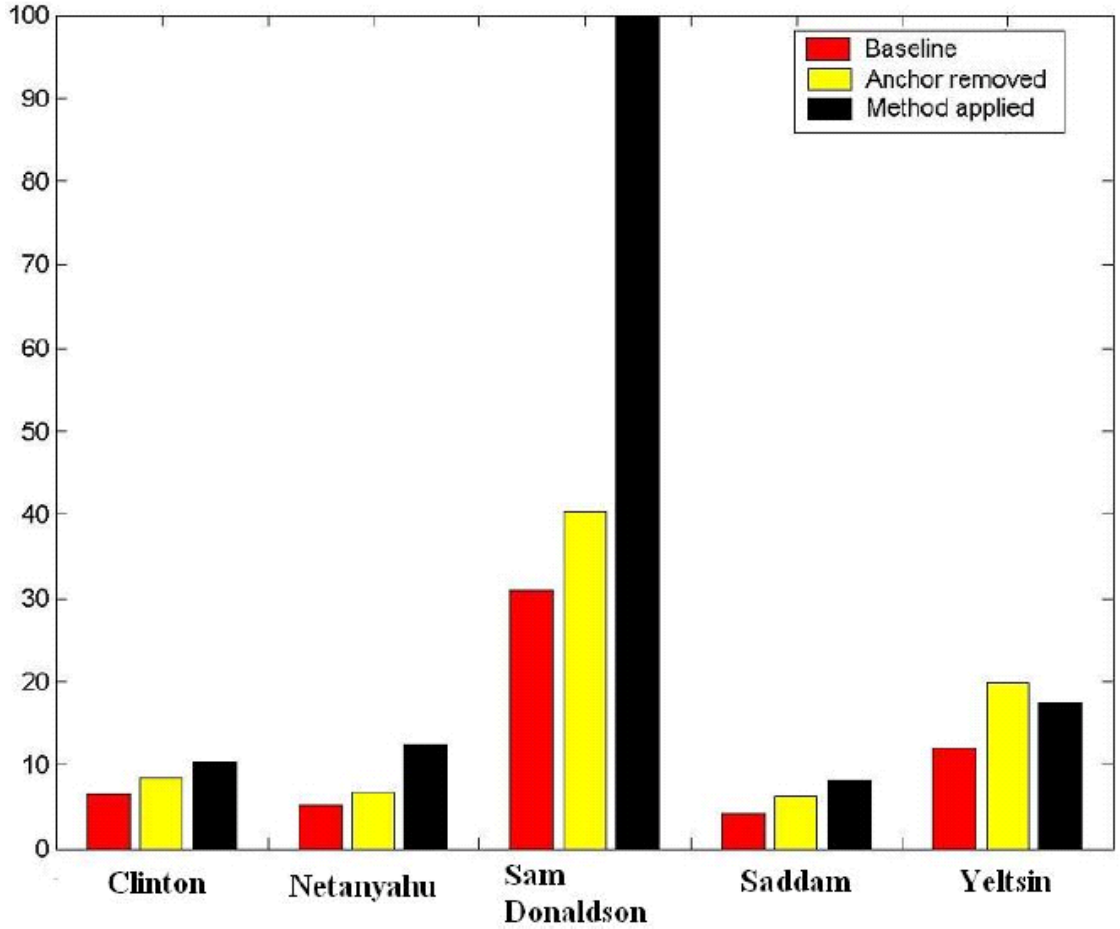




Experimental Results on News Videos

- **Anchorperson detection**
- Average recall : 90%, Average precision : 85%

Person queries



Average precisions

- Baseline: 11.8 %
- After anchor persons removed: 15 %
- Overall: 29 %

29%

152%

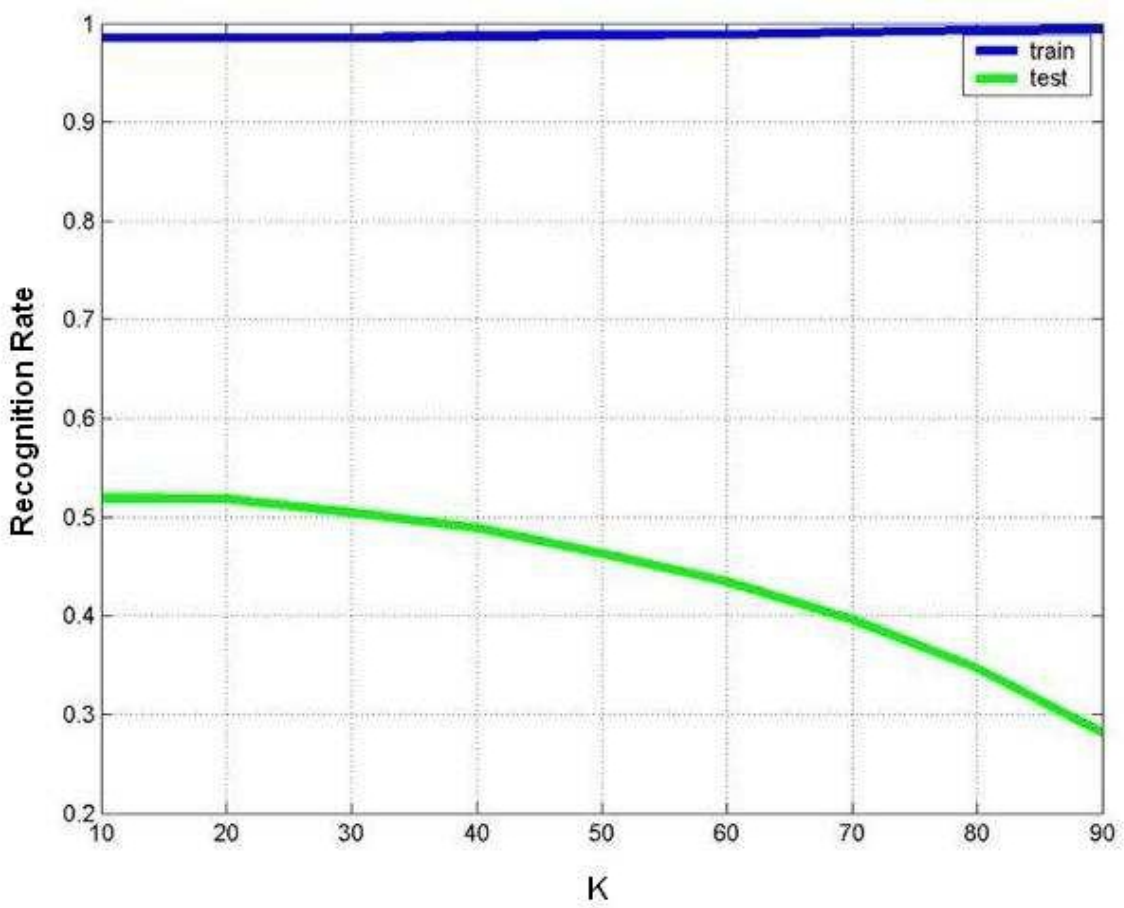
Sample Images Retrieved

Comparisons

- **Overview:**
- Baseline Method
- Feature Selection and Similarity Matrix Construction
  - Finding True Matching Points
  - Facial Features
- Extracting Similar Group of Faces
  - k-nn Approach
  - One-class Classification
- Comparison with Related Studies

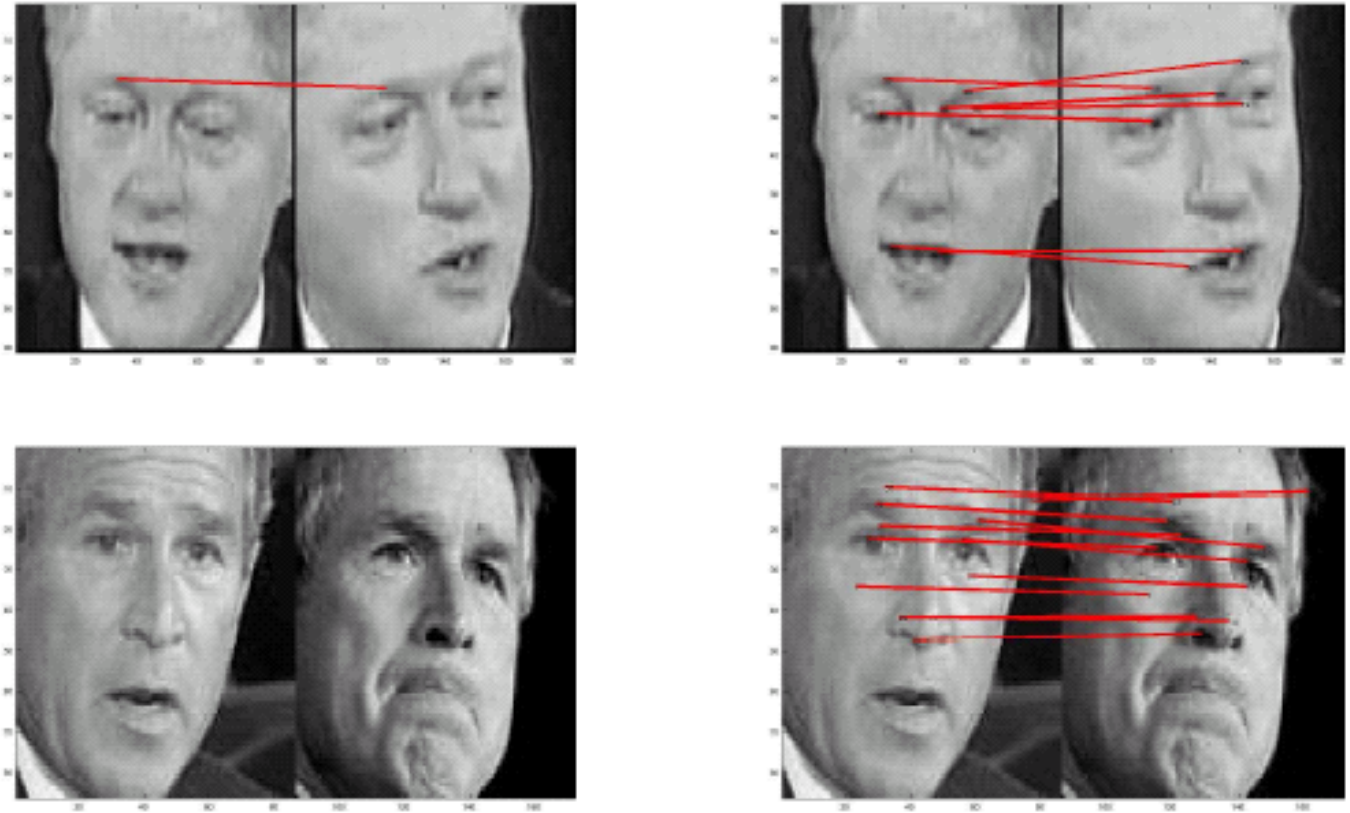
Comparisons

- **Baseline Method : Eigenface method**
- The experiments are conducted on the ground truth faces of the top 23 people used in news photos.



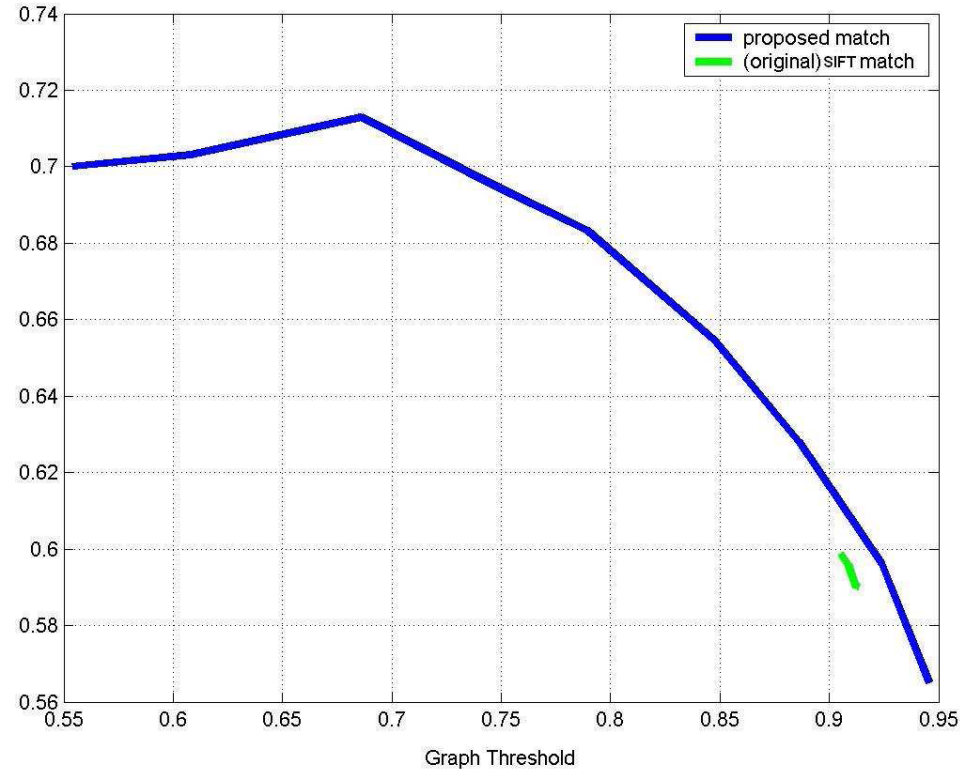
Comparisons

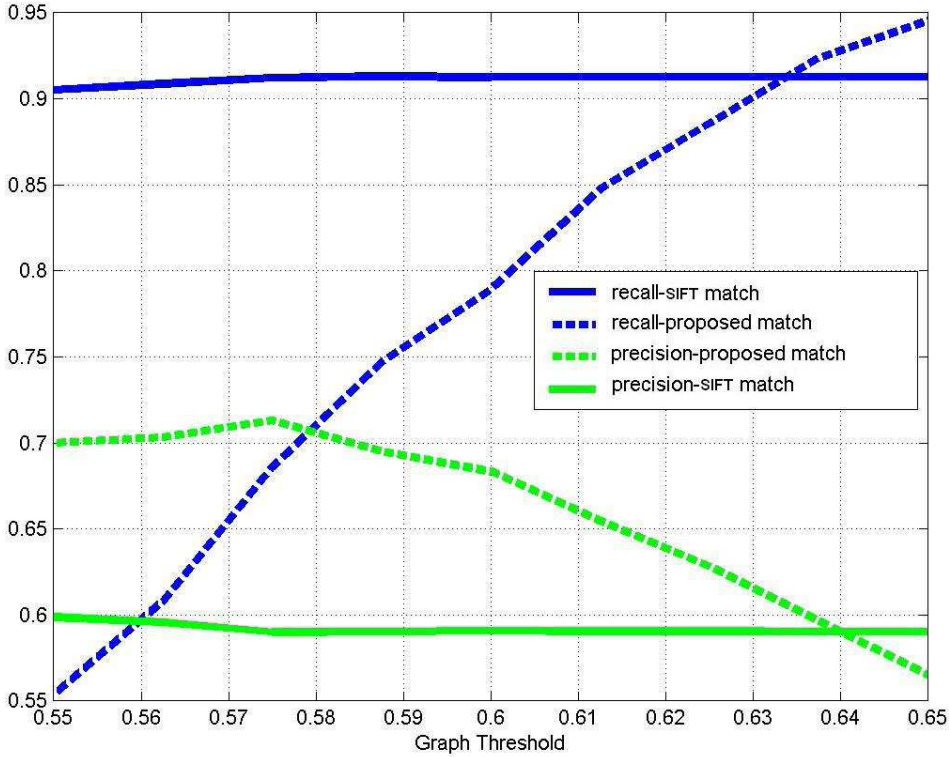
- **Finding True Matching Points**
- We constructed the similarity graph with using the matches of the original SIFT metric and then applied the densest component algorithm.



Comparisons

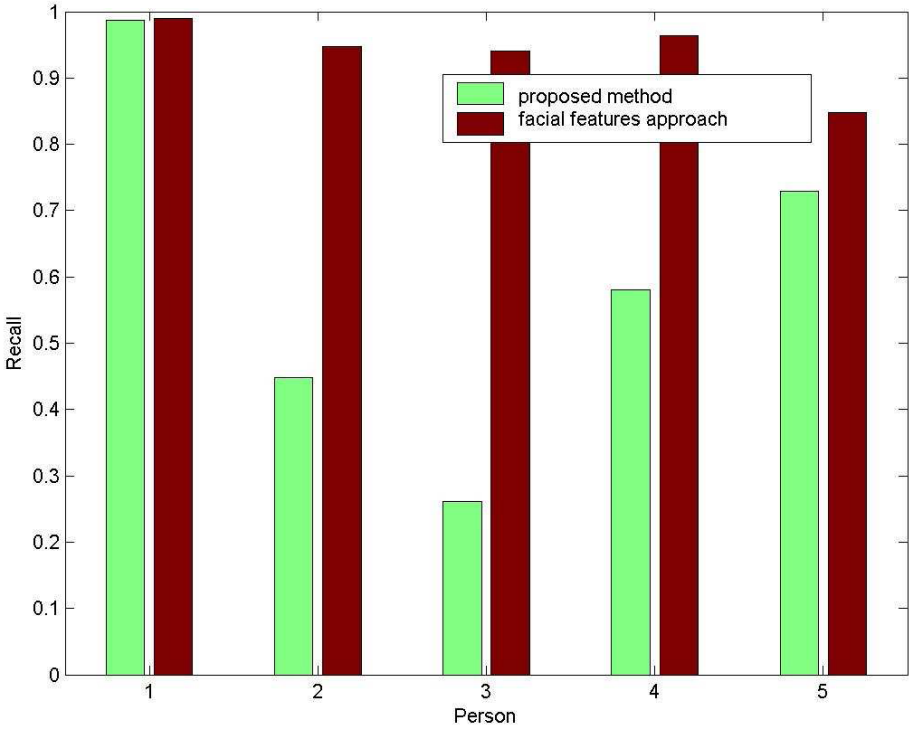
- **Finding True Matching Points**

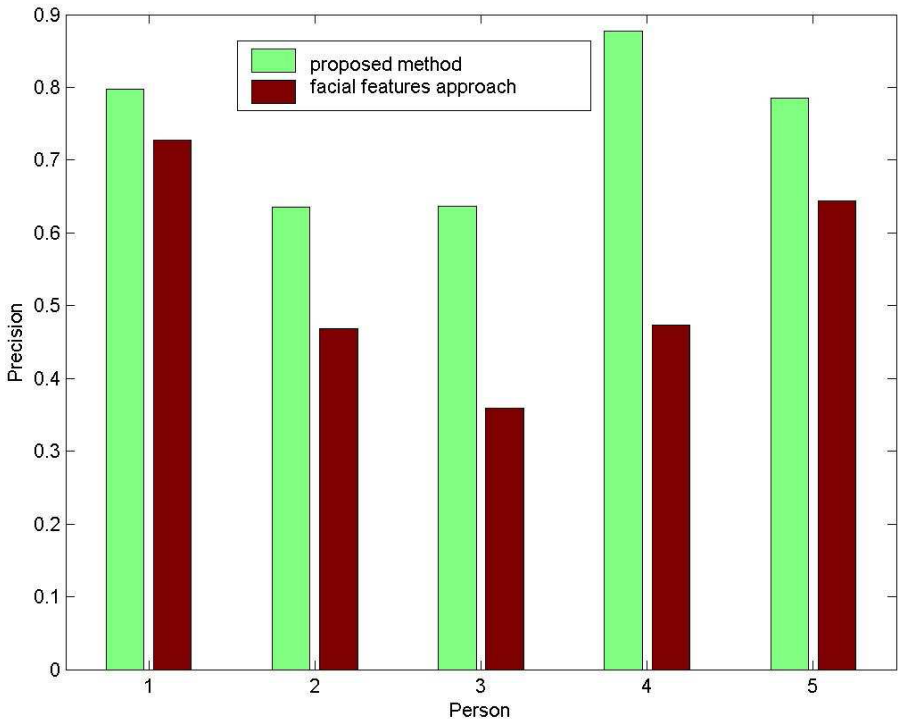




Comparisons

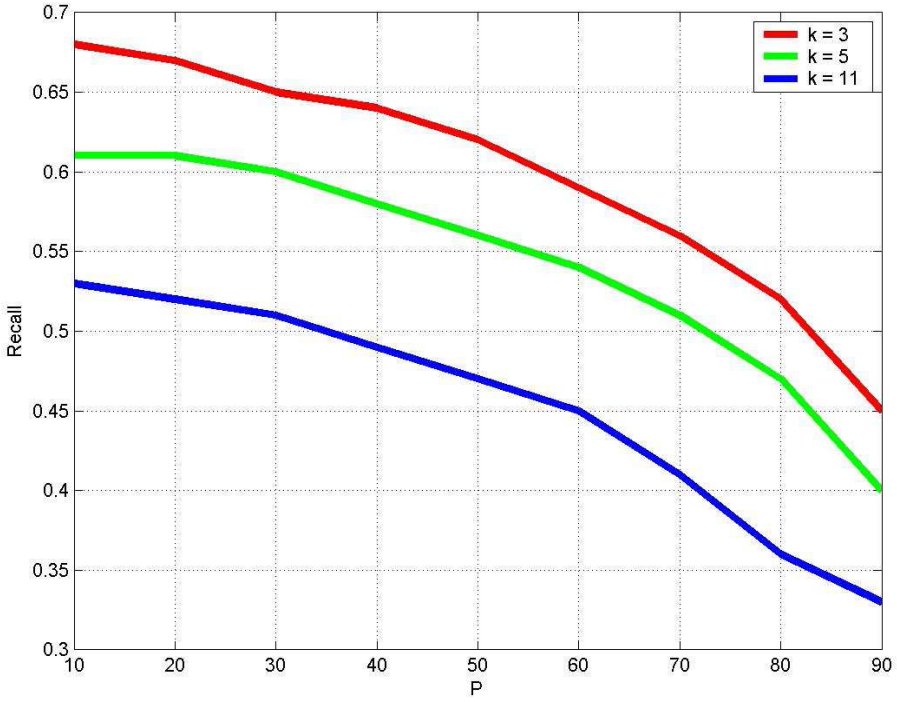
- **Facial Features**

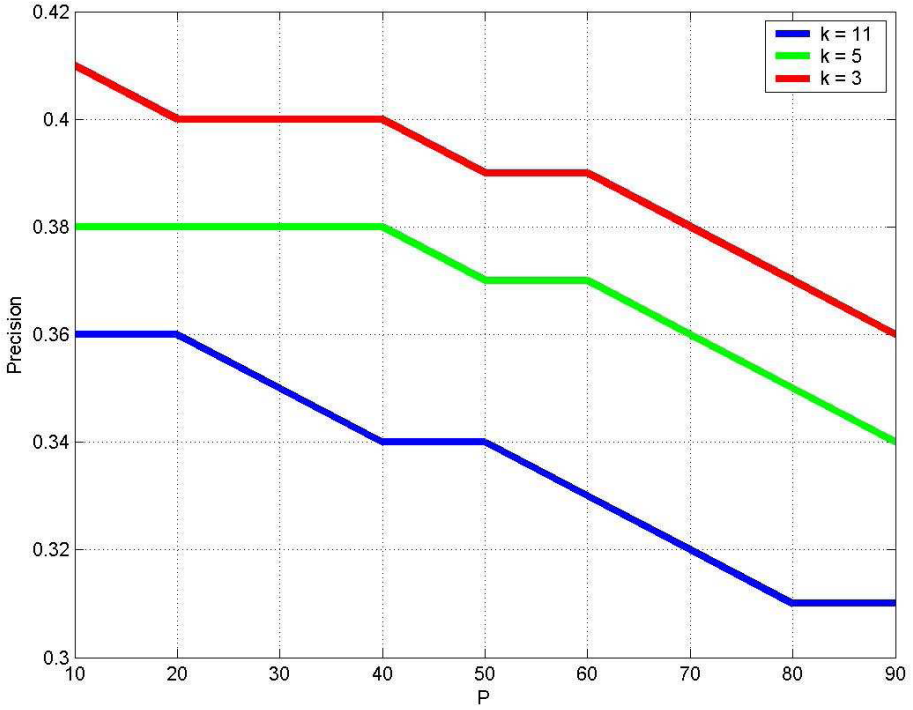




Comparisons

- **Extracting Similar Group of Faces: k-NN**





P is the percentage of the images used for testing, and k is the number of neighbors in k-nn.

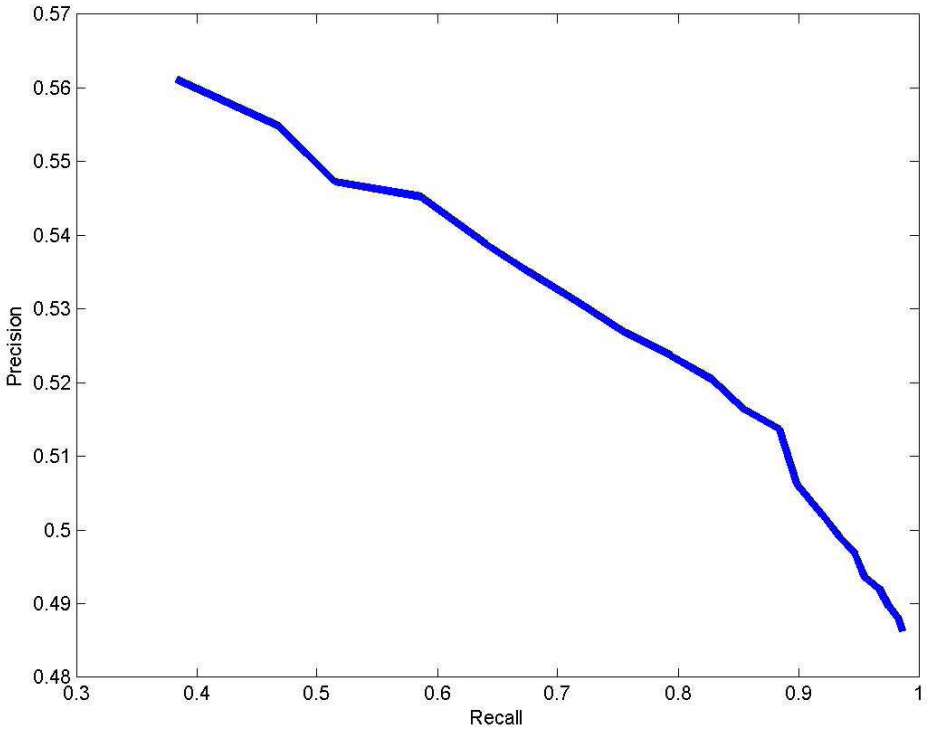
Comparisons

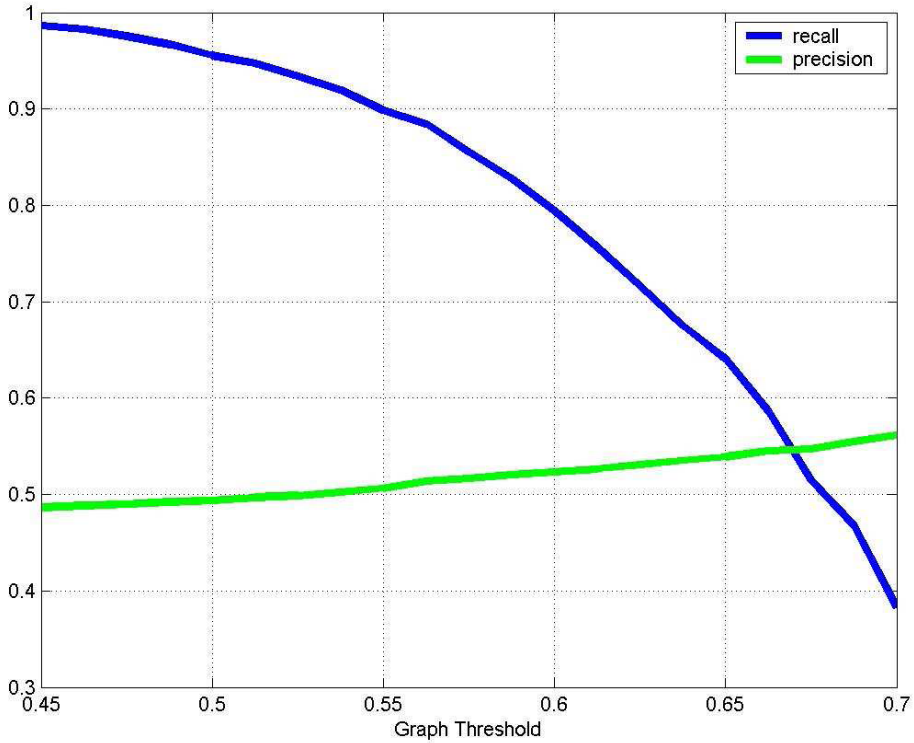
- **Extracting Similar Group of Faces: one-Class**

- Bag-Of-Features approach is used for graph construction:
  - extract sift features from each face image
  - cluster these features using k-means clustering into 50 clusters. => a histogram of the size of number of clusters (50) is formed for each image
  - the frequencies of the clusters are weighted by *'term frequency inverse document frequency (tf-idf)'*

Comparisons

- **Extracting Similar Group of Faces: one-Class**





Comparisons

- **Extracting Similar Group of Faces: one-Class**

- w1 => nearest neighbor data description method
- w2 => k-nearest neighbor data description method

training set										
	K = 10		K = 20		K = 30		K = 40		K = 50	
	rec	pre	rec	pre	rec	pre	rec	pre	rec	pre
w1	1.00	0.54	1.00	0.53	1.00	0.52	1.00	0.52	1.00	0.51
w2	0.90	0.57	0.90	0.55	0.90	0.54	0.90	0.54	0.90	0.53
test set										
	K = 10		K = 20		K = 30		K = 40		K = 50	
	rec	pre	rec	pre	rec	pre	rec	pre	rec	pre
w1	.90	0.50	0.91	0.50	0.90	0.50	0.90	0.50	0.90	0.49
w2	0.84	0.53	0.88	0.54	0.87	0.53	0.86	0.53	0.86	0.52

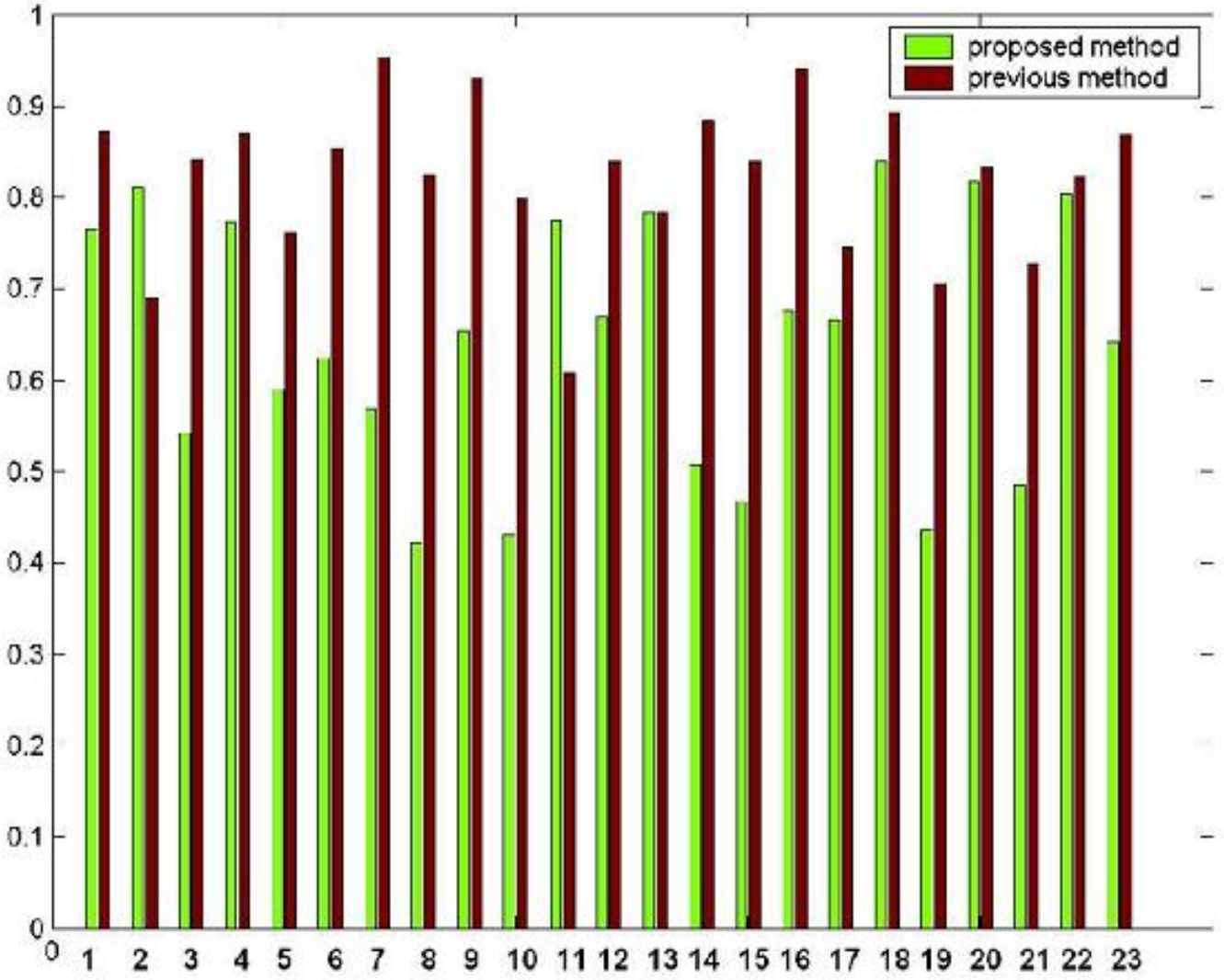
Comparisons

- **Related Study I:** Berg et al. in [2]
- Reduce dimensionality with PCA and LDA
- Represent each image with both
  - a vector gained after the kPCA and LDA processes, and
  - a set of associated names extracted from the caption
- assign a label for each image with k-means clustering.

[2] T. Berg, A. C. Berg, J. Edwards, and D.A. Forsyth. Who's in the picture. In *Neural Information Processing Systems (NIPS), 2004.*

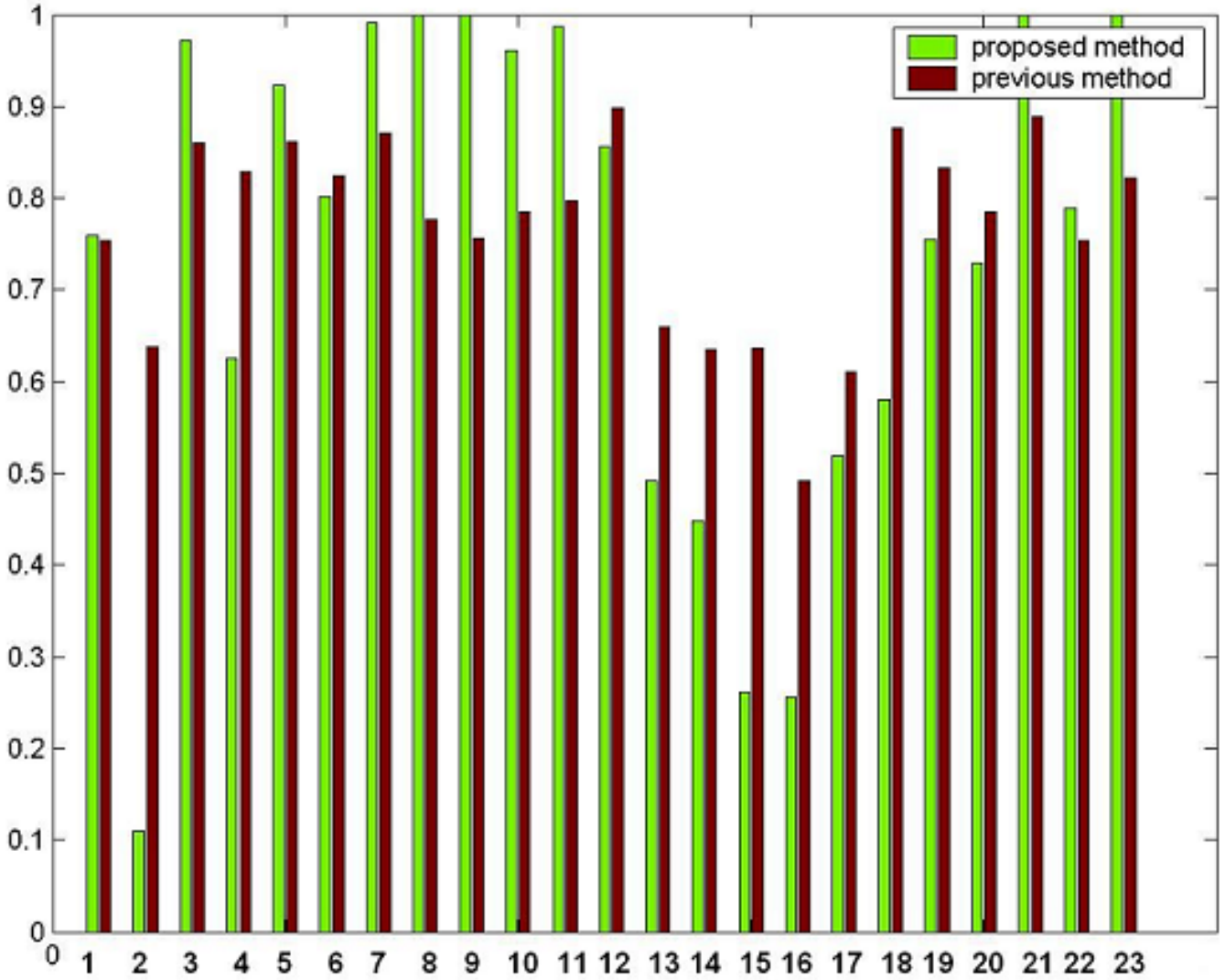
Comparisons

- **Related Study I:**



Comparisons

- **Related Study I:**



Comparisons

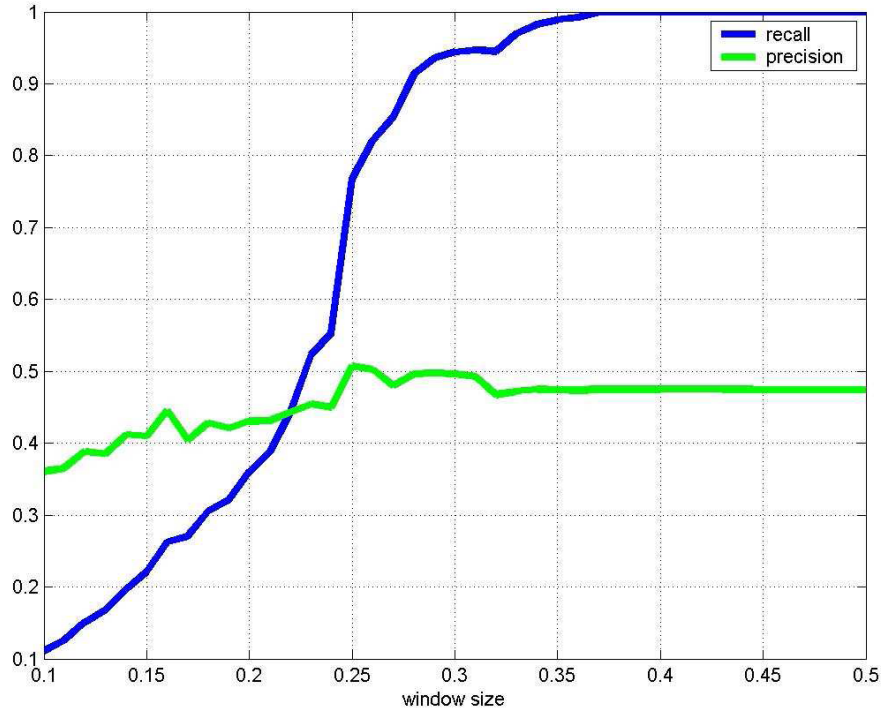
- **Related Study II:** Belongie et al. in [3]

- 500 images from Yahoo!
- Segments each image into blobs and builds a color histogram for each blob
- Clusters the blobs via mean-shift clustering
- Finds the cluster of blobs corresponding to the largest number of parent images
- Re-ranking all the search results based on distance of each blob in each image to the significant cluster

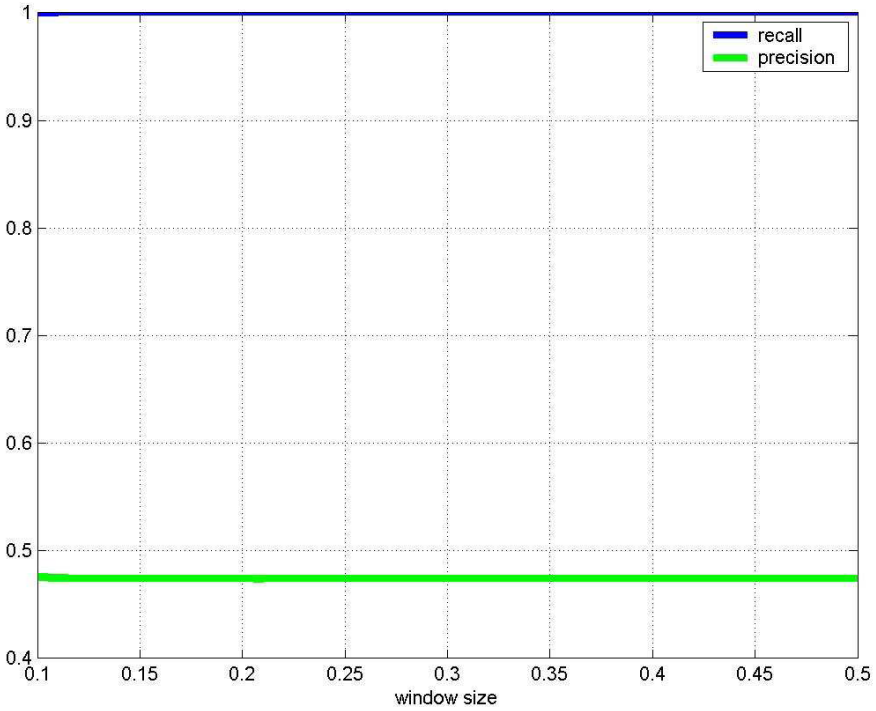
[3] B. Babenko N. B. Haim and S. Belongie. Improving web-based image search via content based clustering. In *Semantic Learning Applications in Multimedia 2006, page 106, 2006.*

Comparisons

- **Related Study II:** Belongie et al. in [3]



multi-dimensional scaling coordinate features



Bag-Of-Features

Conclusions

- In this thesis, we propose a graph based method for querying people in large news photograph and video collections with associated captions or speech transcript texts.

1. constructing a limited search space for a query person by using text
2. defining similarities between faces in this search space to form a similarity graph of faces,
3. finding the densest component of this graph corresponding to the faces of the query person,
4. using the result as a model further in recognizing new faces.

Conclusions

- The proposed method is an overall scheme. Hence, each step of the method can be commited with another tecnique.

- We have proposed a similarity definition based on minimum distance metric of SIFT features and additional constraints;

- And, the greedy graph algorithm is comparable to one other the most possible approach, namely one-class classification.

Future Work

- The approach can also be applied to other problems such as *object recognition or image region annotation*.

- In the context of region annotation, annotations of images can be used for limiting the search space for a region. Then, the region of interest is expected to form the largest similar group of regions.

Future Work

- Instead of taking a single face from each shot by only considering the key-frames, face detection can be applied to all frames in a shot to obtain more instances of the same person.

  This approach can help to find better matching interest points and more examples that can be used in the graph algorithm.

- **Book Chapters**

  Pinar Duygulu, Muhammet Bastan and Derya Ozkan, "Integrating image and text for semantic labeling of images and videos", to appear as a chapter in Machine Learning for Multimedia, Editor: Matthieu Cord, Publisher: Springer-Verlag

- **Conference Papers**

  Derya Ozkan, Pinar Duygulu, "Finding People Frequently Appearing in News", In Proceedings of International Conference on Image and Video Retrieval (CIVR 2006), Arizona State University, Tempe, AZ, USA, July 13-15 2006. Also published in Lecture Notes in Computer Science, Volume 4071/2006.

  Derya Ozkan, Pinar Duygulu, "A Graph Based Approach for Naming Faces in News Photos",In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, NY, June 17-22, 2006.

  Derya Ozkan, Pinar Duygulu, "Yuz ve Isim Iliskisi Kullanarak Haberlerdeki Kisilerin Bulunmasi" (in Turkish), In Proceedings of IEEE 14. Sinyal Isleme ve Iletisim Uygulamalari Kurultayi (SIU 2006), Antalya, Turkey, April 17-19, 2006.

THANK YOU FOR LISTENING !!!