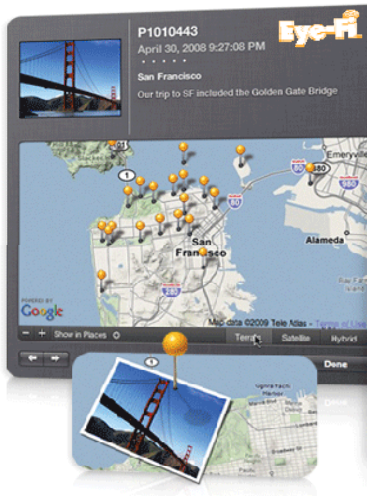
Geotag Propagation in Social Networks Based on User Trust Model

Ivan Ivanov, Peter Vajda, Jong-Seok Lee,
Lutz Goldmann, Touradj Ebrahimi

Multimedia Signal Processing Group
Ecole Polytechnique Federale de Lausanne, Switzerland

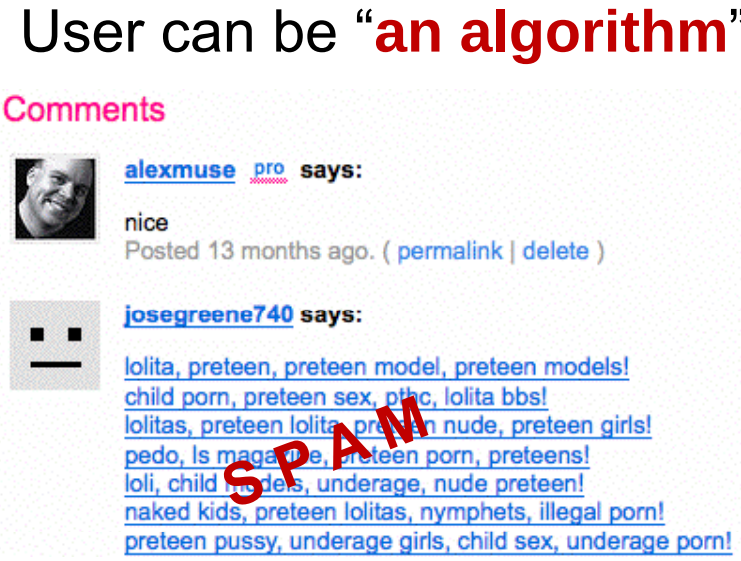
Motivation

- We introduce users in our system for **geotagging** in order to simulate a real social network
- **GPS coordinates** to derive geographical annotation, which are not available for the majority of web images and photos
- A GPS sensor in a camera provides only **the location of the photographer** instead of that of the captured landmark
- Sometimes **GPS and Wi-Fi geotagging** determine wrong location due to noise



Motivation

- **Tag** – short textual annotation (**free-form keyword**) used to describe photo in order to provide meaningful information about it
- User-provided tags may sometimes be **spam annotations** given on purpose or **wrong tags** given by mistake
- User can be "**an algorithm**"



http://www.flickr.com/photos/scriptingnews/2229171225

http://code.google.com/p/spamcloud

Goal

- Consider **user trust information** derived from users' tagging behavior for the tag propagation
- Build up an **automatic tag propagation system** in order to:
  - ***Decrease the annotation time****, and*
  - ***Increase the accuracy of the system***

http://www.costadevault.com/blog/2010/03/listening-to-strangers

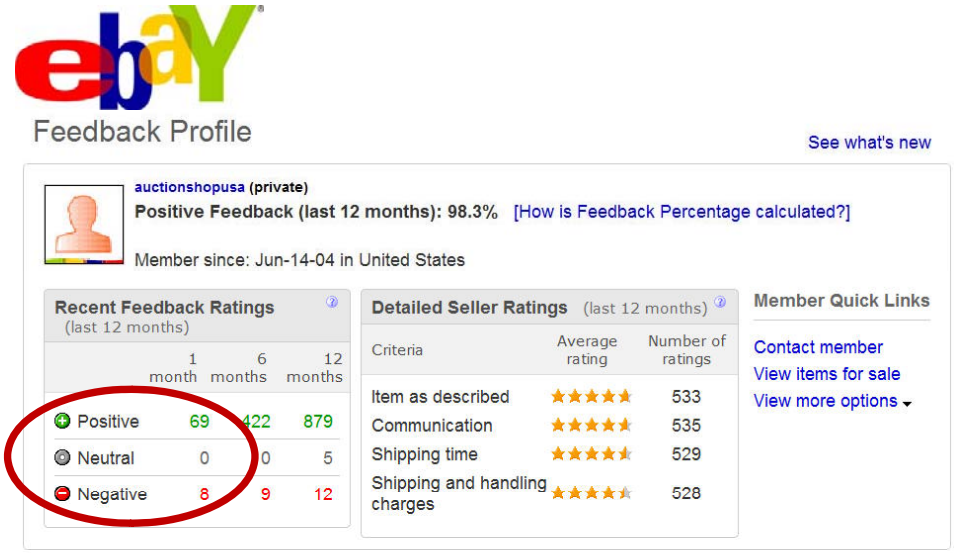
Geotags

Tags
- summer school
- June
- 2009
- sea

Geotags
- Hagia Sophia
- Blue Mosque
- Istanbul
- Turkey
- geo:lat = 41.01667
- geo:lon = 28.96667

- **IPTC schema** is considered:
  - *City* – name of the city
  - *Sublocation* – area or name of the landmark
  - *Examples:* Istanbul (Hagia Sophia), San Francisco (Golden Gate Bridge)

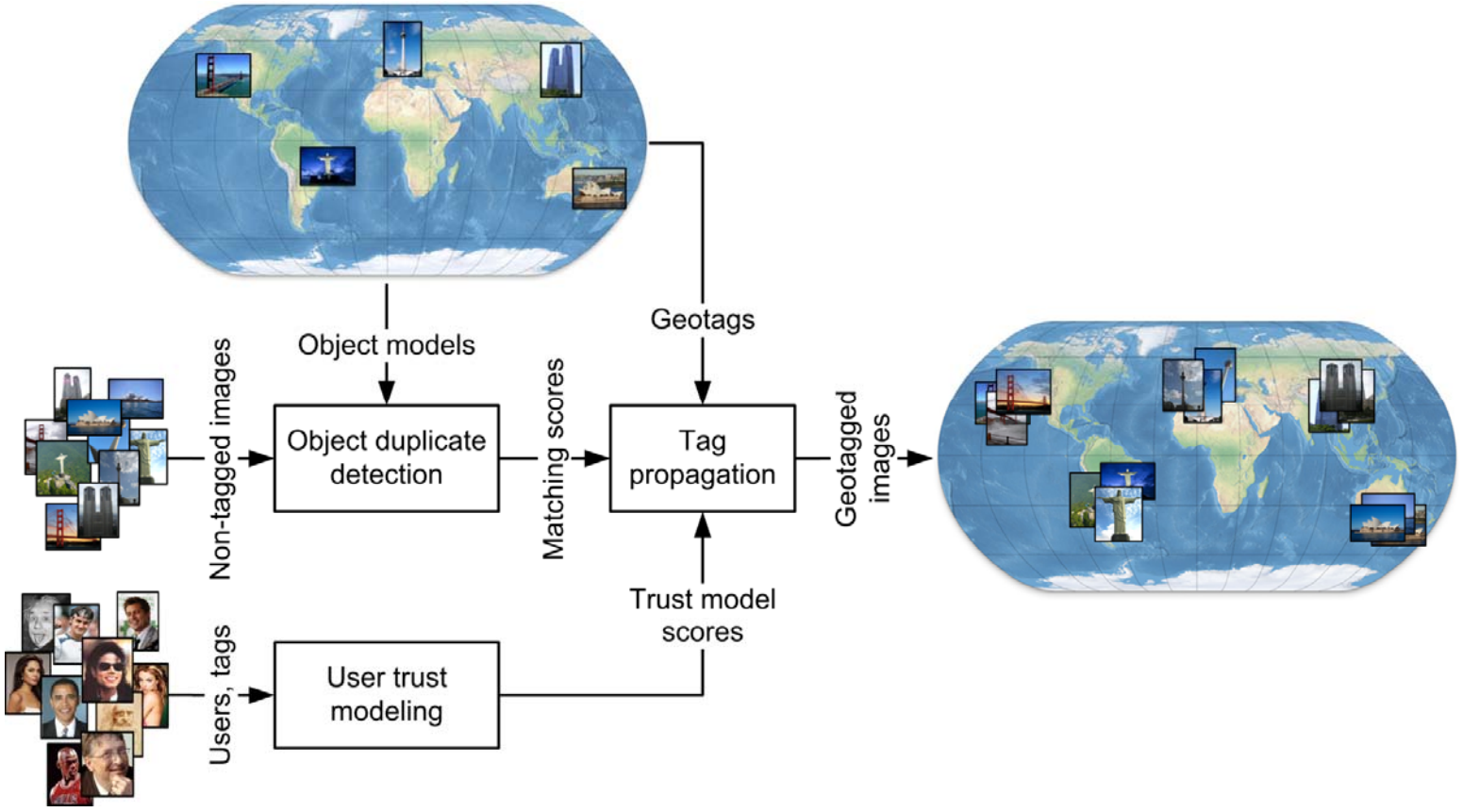
Trust and reputation systems



http://www.wpconfig.com/2010/07/04/tips-to-increase-your-pagerank

Email spam filtering

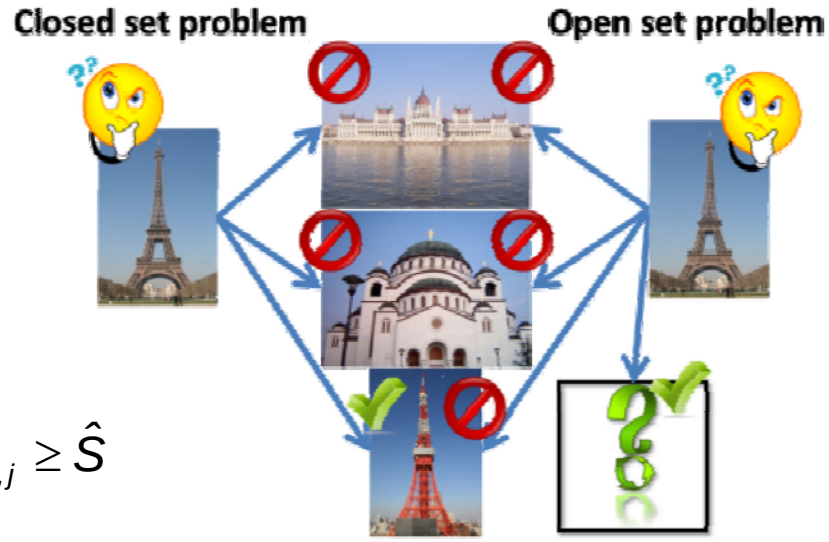
System overview



Tag propagation

- Object duplicate detection module provides a matching score matrix $S_{i,j}$, $i \in [1,M]$, $j \in [1,N]$
- Two application scenarios:
  - ***Closed set problem***:

$$C_{i,j} = \begin{cases} 1, & \text{if } S_{i,j} = \max\limits_{i \in [1,M]} \{S_{i,j}\} \\ 0, & \text{otherwise} \end{cases}$$



  - ***Open set problem***:

$$O_{i,j} = \begin{cases} 1, & \text{if } S_{i,j} = \max\limits_{i \in [1,M]} \{S_{i,j}\} \wedge S_{i,j} \geq \hat{S} \\ 0, & \text{otherwise} \end{cases}$$

- Three steps:
  - ***User evaluation*** – comparing the predicted tags to manually defined ground truth (for a real photo sharing system, such as Panoramio, it is not necessary to collect ground truth data since user feedback can replace them):

$$U_{i,j} = \begin{cases} 1, & \text{if user } j \text{ tags image } i \text{ correctly} \\ 0, & \text{otherwise} \end{cases}$$

  - ***Trust model creation*** – the percentage of the correctly tagged images by particular user out of the overall number of tagged images M:

$$T_j = \frac{\sum_{i=1}^{M} U_{i,j}}{M}$$

  - ***Tag propagation*** – only tags from users who are trusted ($T_j > \hat{T}$) are propagated to other photos in the dataset

User trust modeling in Panoramio

http://www.panoramio.com/photo/23286122

User trust modeling in Panoramio

- Idea:

true flag (correct tag)

false flag (wrong tag + suggest a correct tag)

- The more misplacements a user has, the more untrusted he/she is
- User trust ratio:

$$T_j = \left(\frac{\#\ \text{true flags}}{\#\ \text{all associated flags}} \right)_{\text{all images tagged by user } j}$$

Eliminate spammers in Panoramio

- Users can collaboratively eliminate a spammer:

Bob

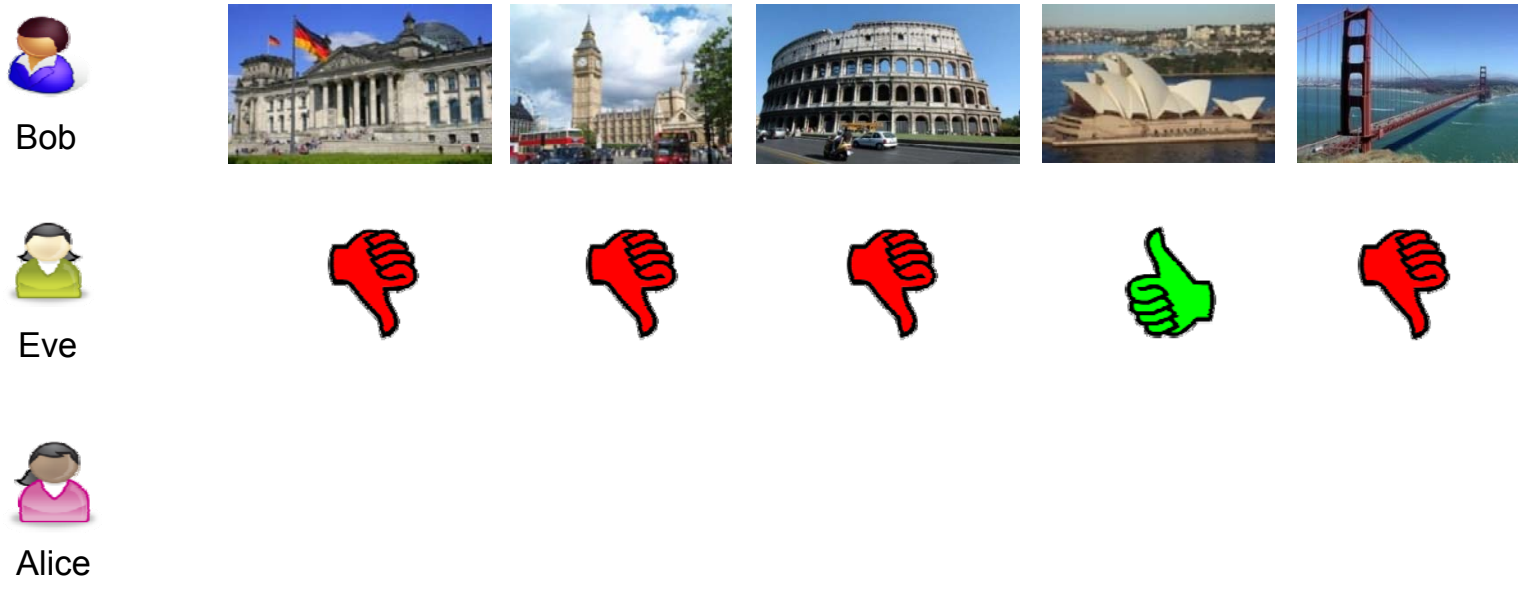
Eve

Alice

$T\left(\text{Bob}\right) = 0.9$ $T\left(\text{Alice}\right) = 0.7$ $T\left(\text{Eve}\right) = 0.5$

Eliminate spammers in Panoramio

- Users can collaboratively eliminate a spammer:



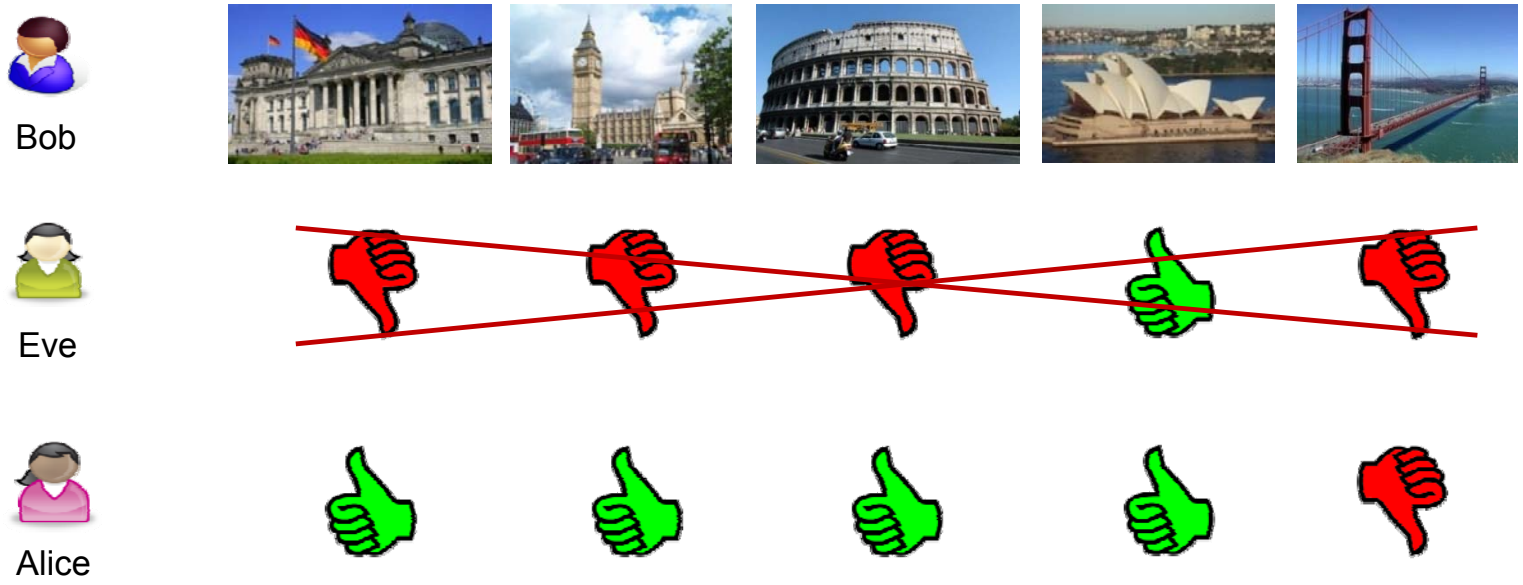
$T($ $) = 0.3$ ↓ Bob

$T($ $) = 0.7$ Alice

$T($ $) = 0.9$ ↑ Eve

Eliminate spammers in Panoramio

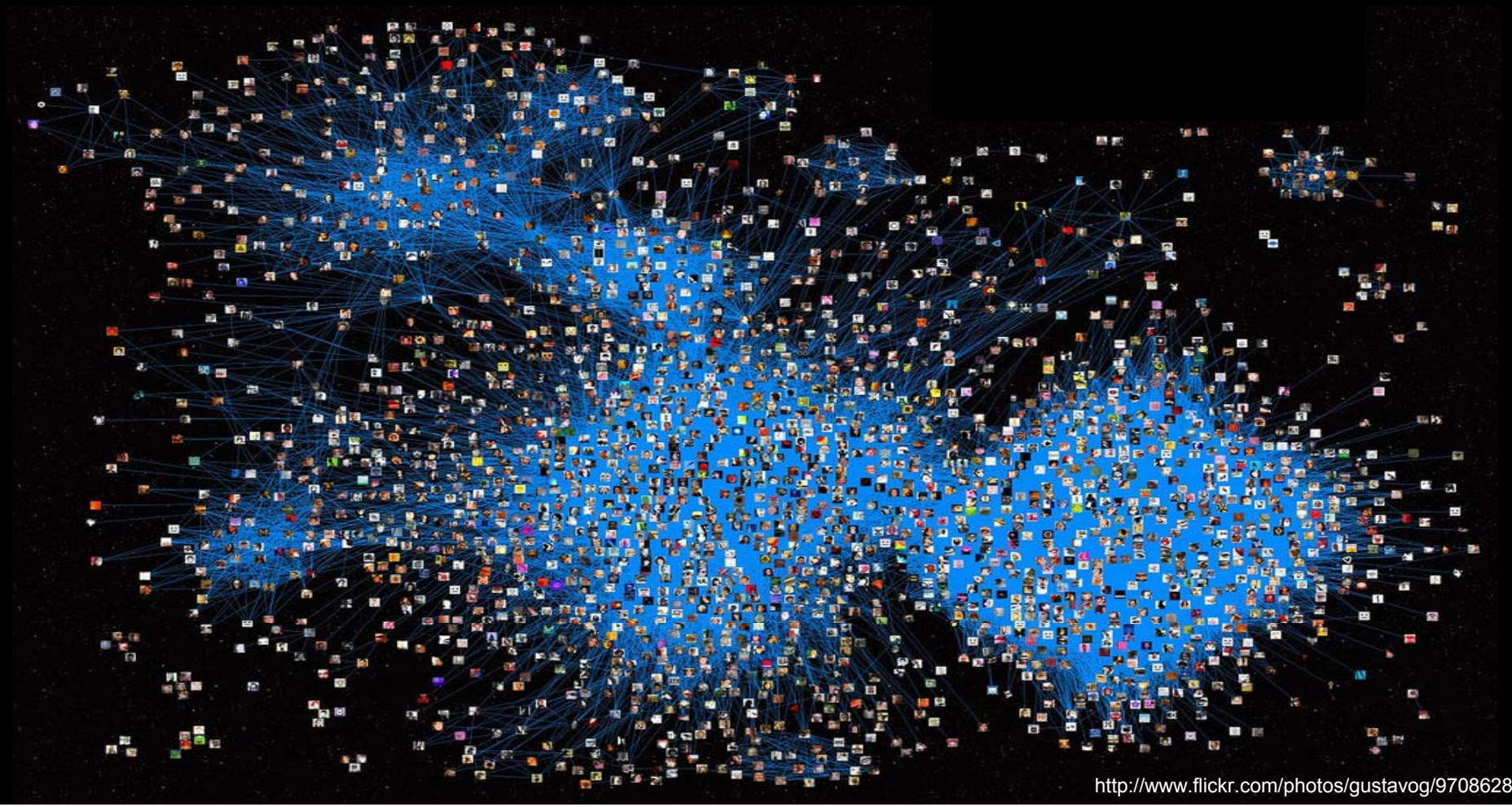
- Users can collaboratively eliminate a spammer:



$T(\text{Bob}) = 0.7 \uparrow$ $T(\text{Alice}) = 0.8 \uparrow$ $T(\text{Eve}) = 0.4 \downarrow$

Tag propagation based on user trust model

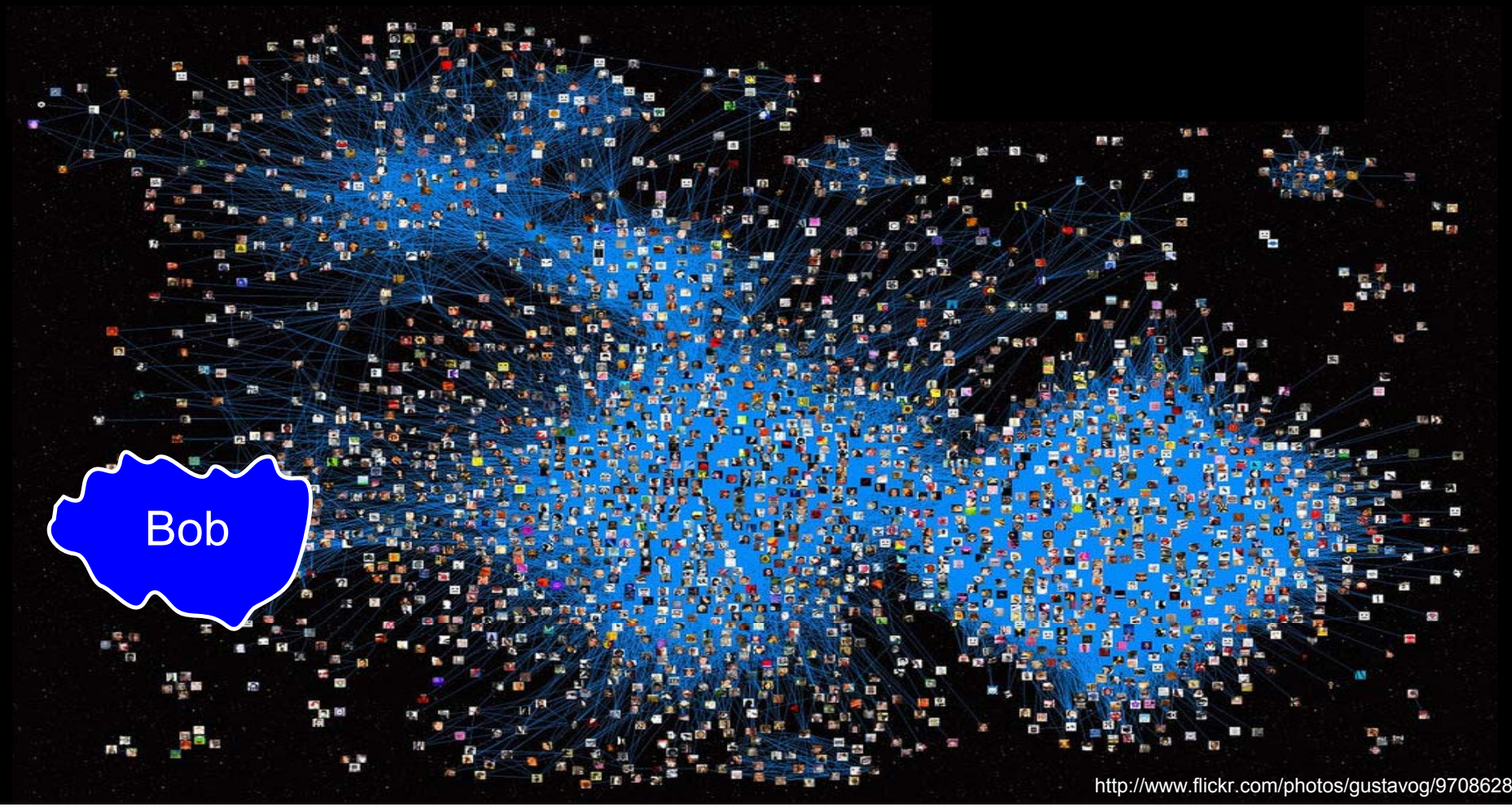
$T(\text{Bob}) = 0.7$ $T(\text{Alice}) = 0.8$ $T(\text{Eve}) = 0.4$ $\hat{T} = 0.6$



http://www.flickr.com/photos/gustavog/9708628

Tag propagation based on user trust model

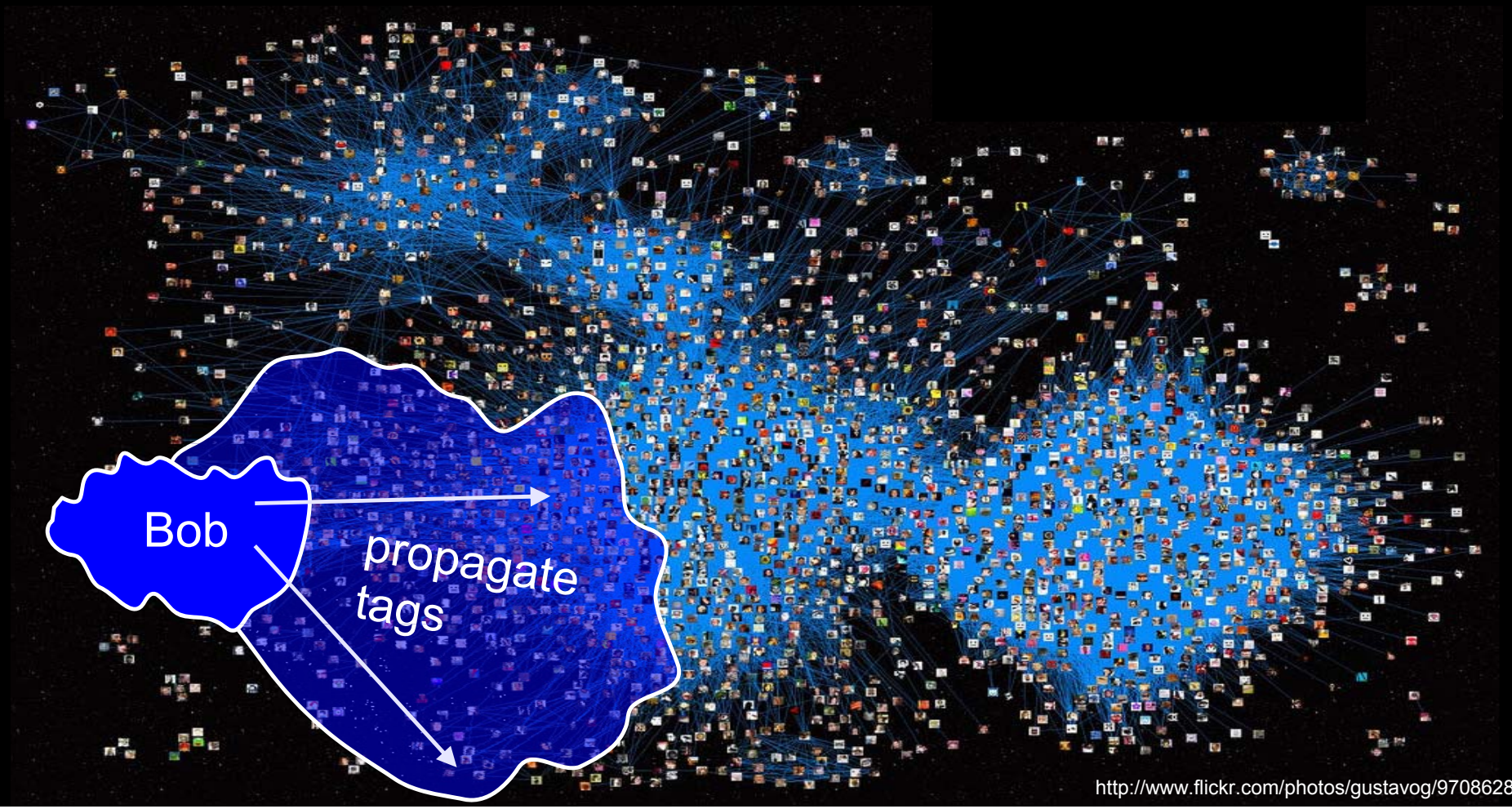
$T(\text{Bob}) = 0.7$ $T(\text{Alice}) = 0.8$ $T(\text{Eve}) = 0.4$ $\hat{T} = 0.6$



http://www.flickr.com/photos/gustavog/9708628

Tag propagation based on user trust model

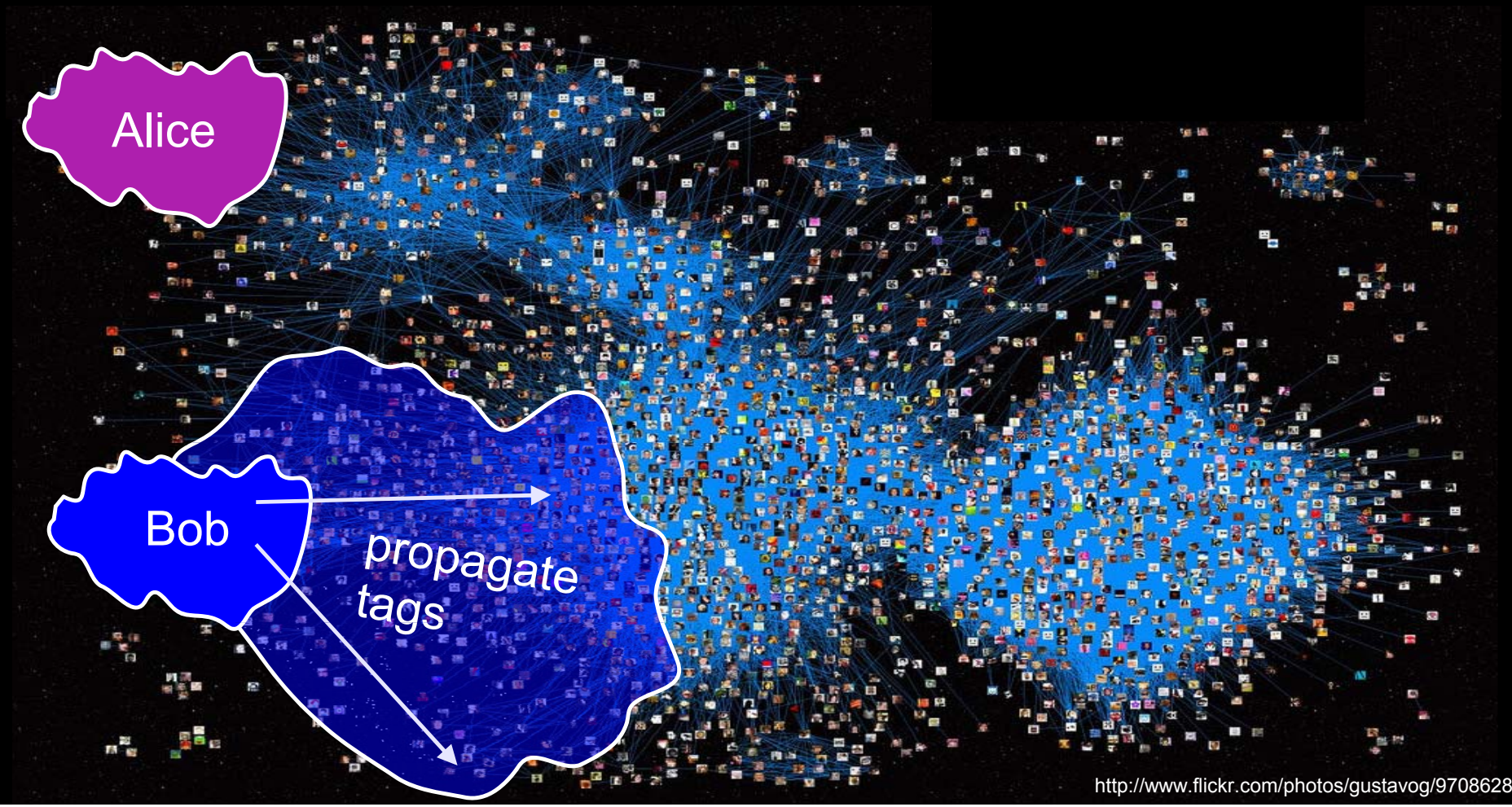
$T(\text{Bob}) = 0.7$ $T(\text{Alice}) = 0.8$ $T(\text{Eve}) = 0.4$ $\hat{T} = 0.6$



http://www.flickr.com/photos/gustavog/9708628

Tag propagation based on user trust model

$T(\text{Bob}) = 0.7$ $T(\text{Alice}) = 0.8$ $T(\text{Eve}) = 0.4$ $\hat{T} = 0.6$



http://www.flickr.com/photos/gustavog/9708628

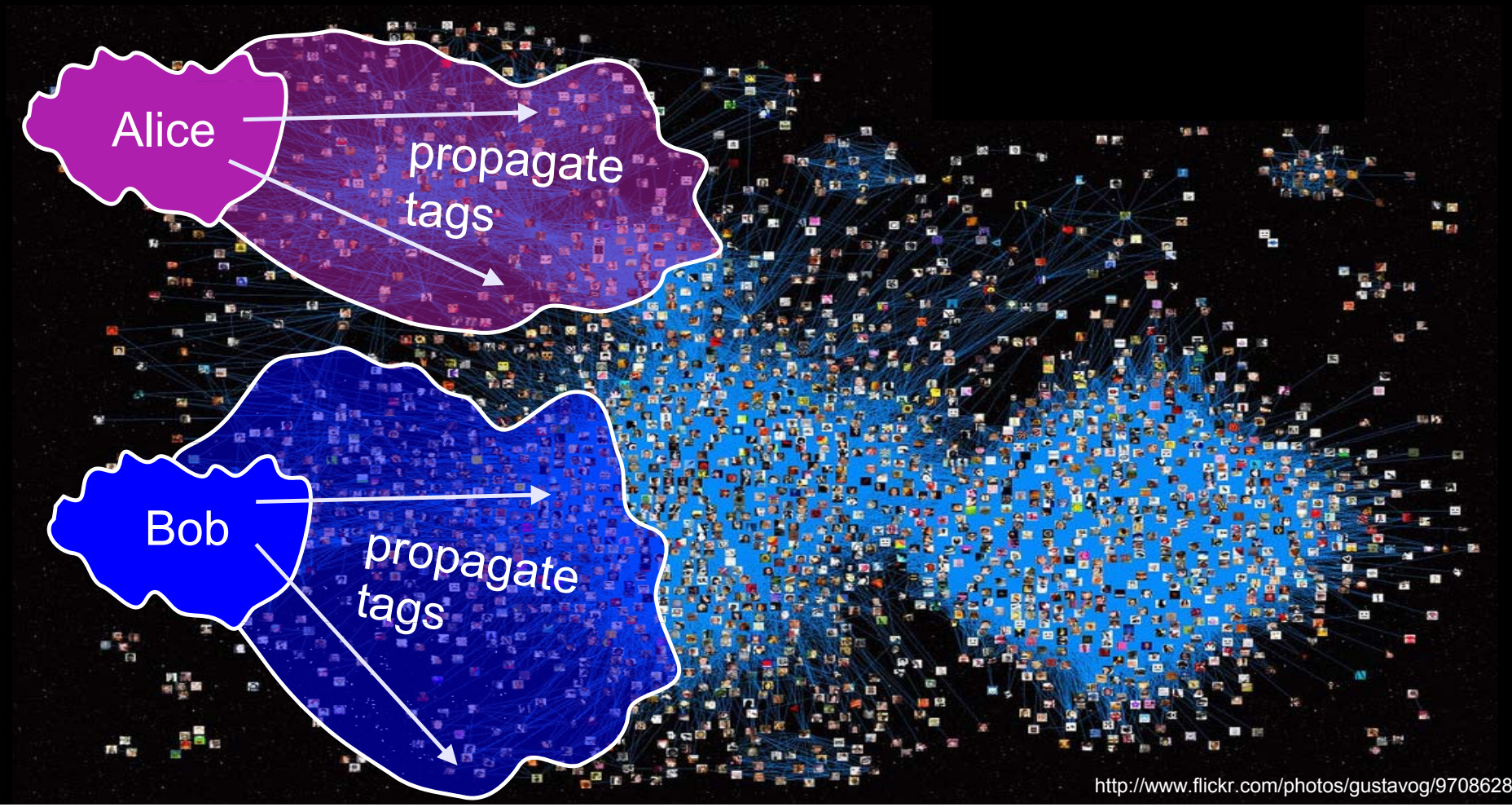
Tag propagation based on user trust model

$T($ $) = 0.7$ Bob

$T($ $) = 0.8$ Alice

$T($ $) = 0.4$ Eve

$\hat{T} = 0.6$



http://www.flickr.com/photos/gustavog/9708628

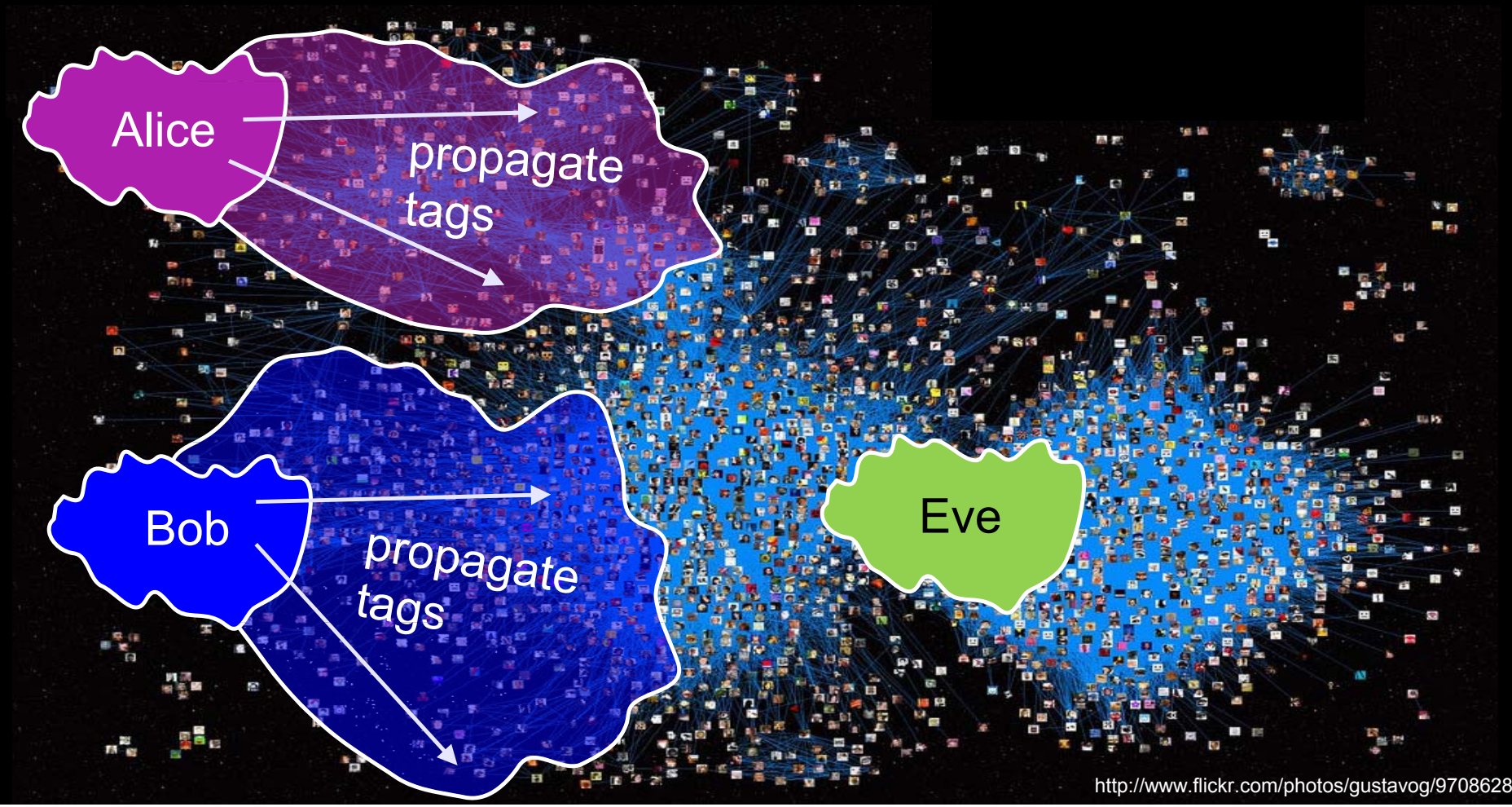
Tag propagation based on user trust model

$T($ $) = 0.7$ Bob

$T($ $) = 0.8$ Alice

$T($ $) = 0.4$ Eve

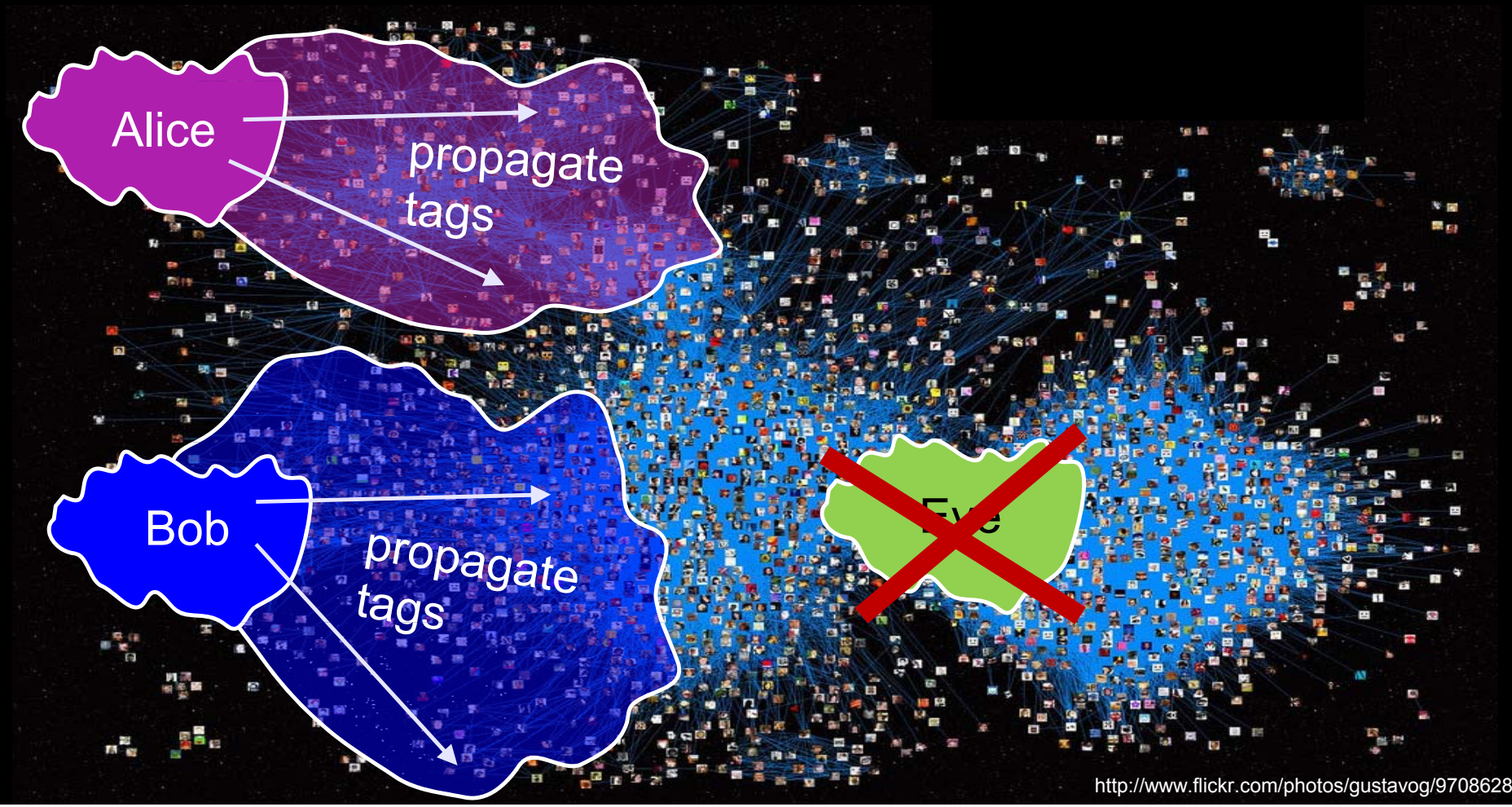
$\hat{T} = 0.6$



http://www.flickr.com/photos/gustavog/9708628

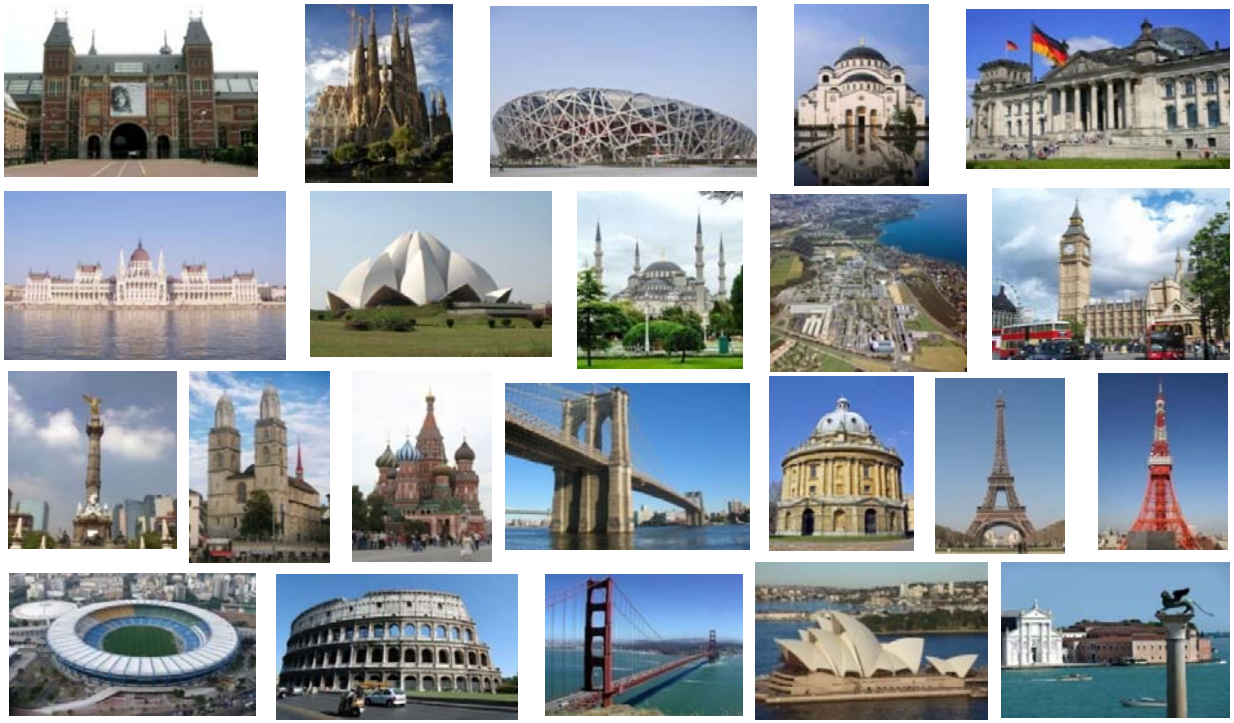
Tag propagation based on user trust model

$T(\text{Bob}) = 0.7$ $T(\text{Alice}) = 0.8$ $T(\text{Eve}) = 0.4$ $\hat{T} = 0.6$

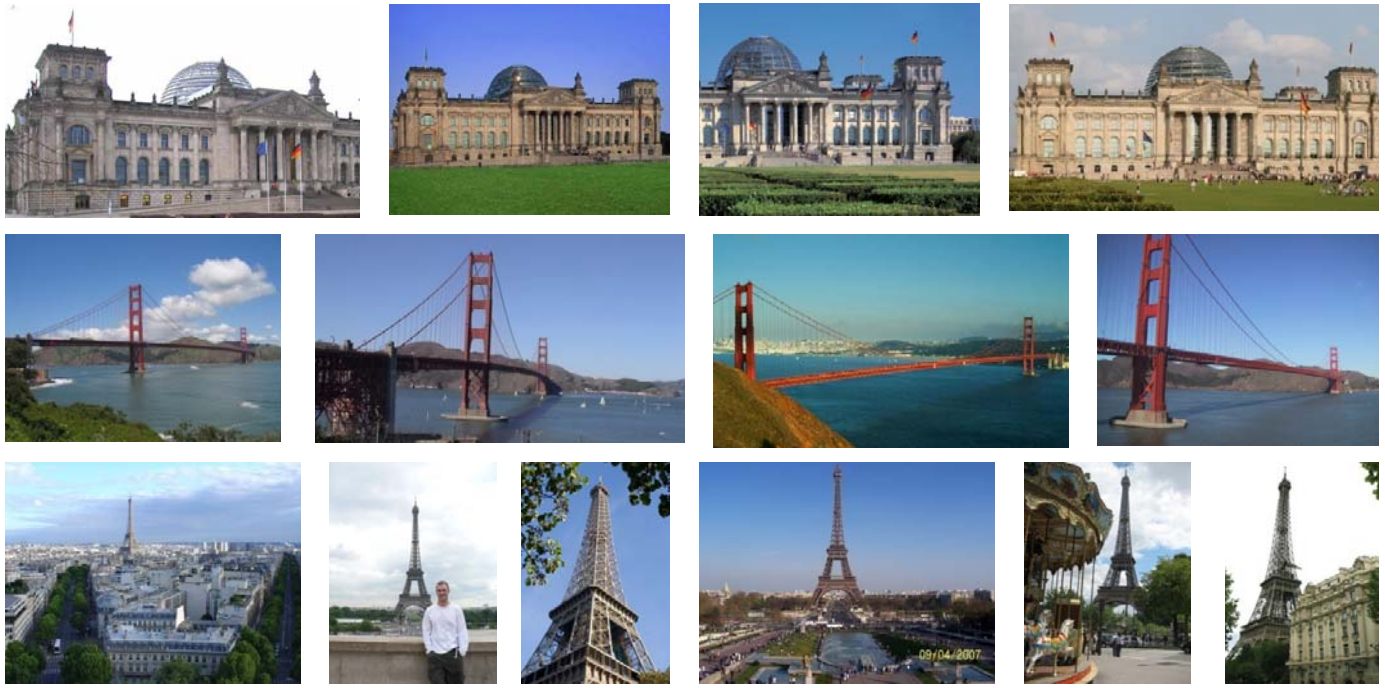


http://www.flickr.com/photos/gustavog/9708628

- **Dataset**
  - *1320 images* – 22 cities, 3 sublocations for each city, 20 sample images for each sublocation



- **Dataset**
  - *Large variety of views, distances and partial occlusions*



- **Methodology**
  - *Extract a sub network from a large social network, in a way that every user in this subsystem annotates every monument in the subset of the dataset*
  - *Upon this sub network, we build up an automatic propagation system in order to decrease the annotation time and increase the accuracy of the system*
  - *44 subjects were asked to tag 66 photos, chosen in advance*
  - *If either one of geotags (name of the area or name of the monument depicted in the image) is correlated with the object in the image, we assume that the image is correctly tagged*
- **Evaluation**
  - *Recognition rate (accuracy):*

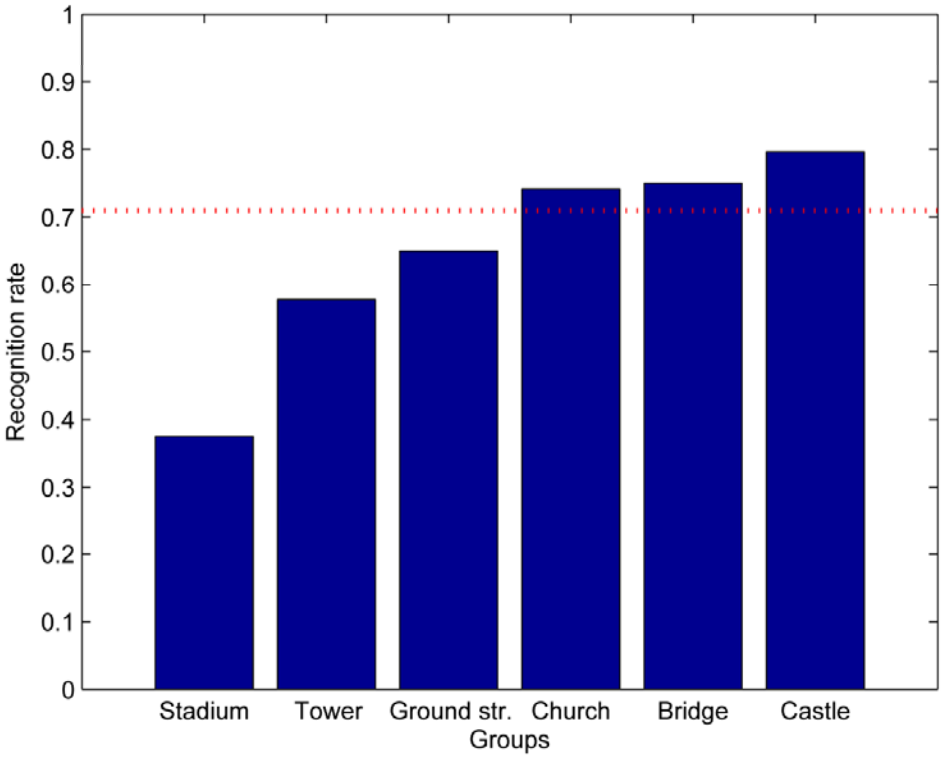
$$RR = \frac{T}{A}$$

Results

- The recognition rate for all landmarks

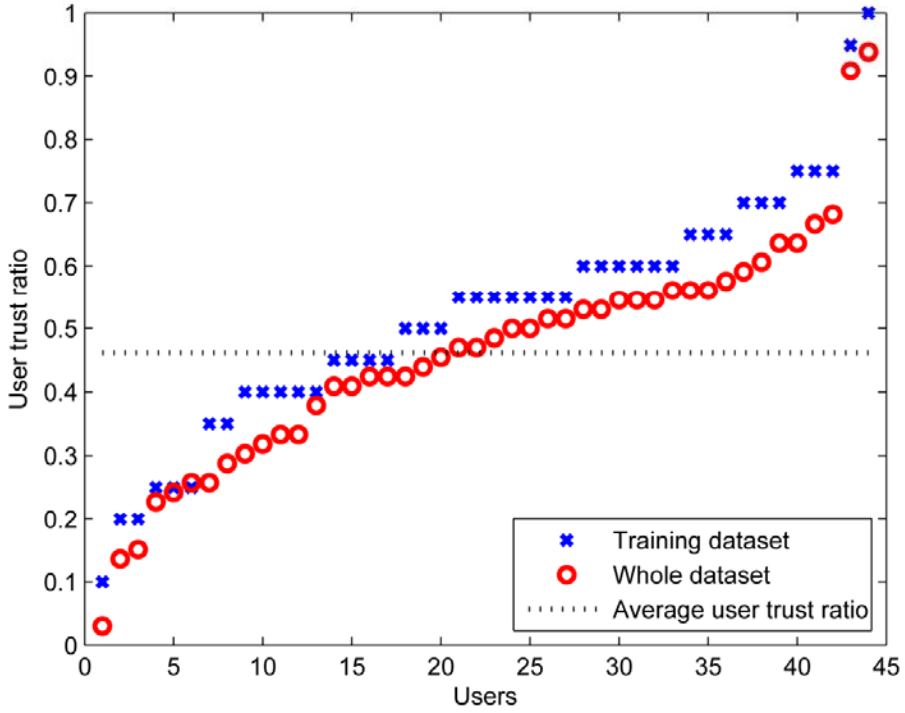
Sydney	Harbour Bridge	Luna Park	Opera House
Oxford	Radcliffe	All Souls College	Ashmolean Museum
Budapest	Parliament	Buda Castle	Hero Square
Paris	Eiffel Tower	Louvre	Arc De Triomphe
Moscow	Christ Savior Cathedral	St. Basil	Kremlin
Delhi	Lotus Temple	Akshardham Temple	Humayun Tomb
Venice	Lion Statue	Campanile Di San Marco	St. Mark Bell Tower
Rome	Pantheon	St. Peter Basilica	Colosseum
London	Big Ben	Buckingham Palace	Tower Bridge
Berlin	TV Tower	Brandenburg Gate	Reichstag
Beijing	Temple Of Heaven	Birds Nest Stadium	Tiananmen
Barcelona	Sagrada Familia	Casa Mila	Olympic Communication Tower
Mexico City	Angel De La Independencia	Torre Latinoamericana	Palace Of Fine Arts
San Francisco	Coit Tower	Golden Gate Bridge	Twin Peaks
Amsterdam	Church Of St. Nicholas	Rijksmuseum	Royal Palace
Rio De Janeiro	Cristo Redentor	Paco Imperial	Maracana
Belgrade	Parliament	Winner Statue	St. Sava Church
Zurich	St. Peter	Fraumunster	Grossmunster
Tokyo	Tower	Metropolitan Government Center	National Museum
Istanbul	Blue Mosque	Hagia Sofia	Galata Tower
Lausanne	EPFL	Riponne	Cathedral
New York	Brooklyn Bridge	Statue Of Liberty	Twin Towers

- The recognition rate across the different locations groups (bars) and the recognition rate of all locations (dashed line)



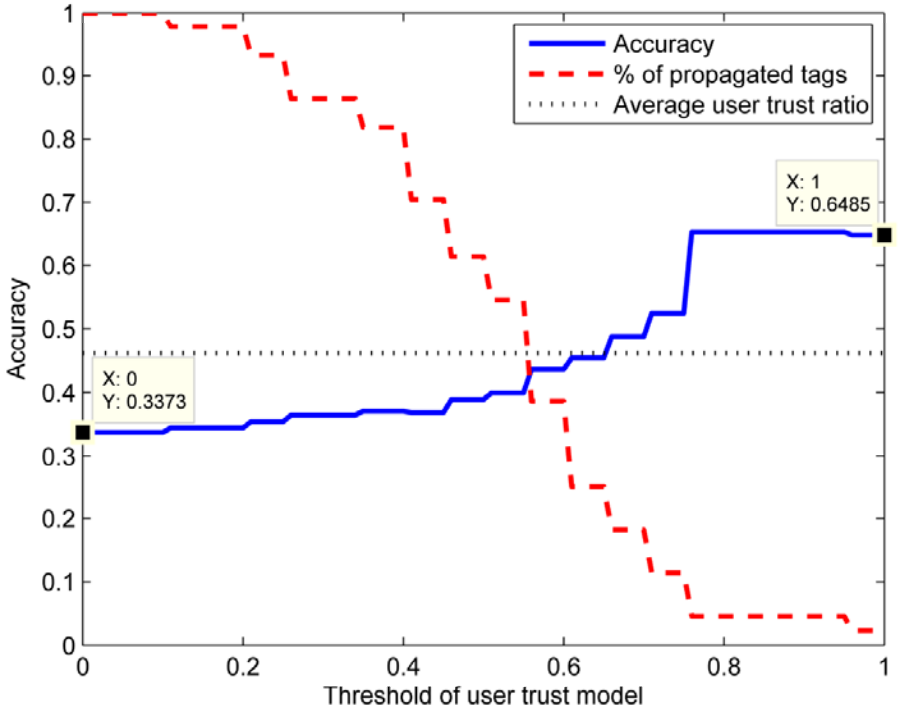
Results

- Trust ratios for the different users for a reduced (first 20 images) and the complete set of training images



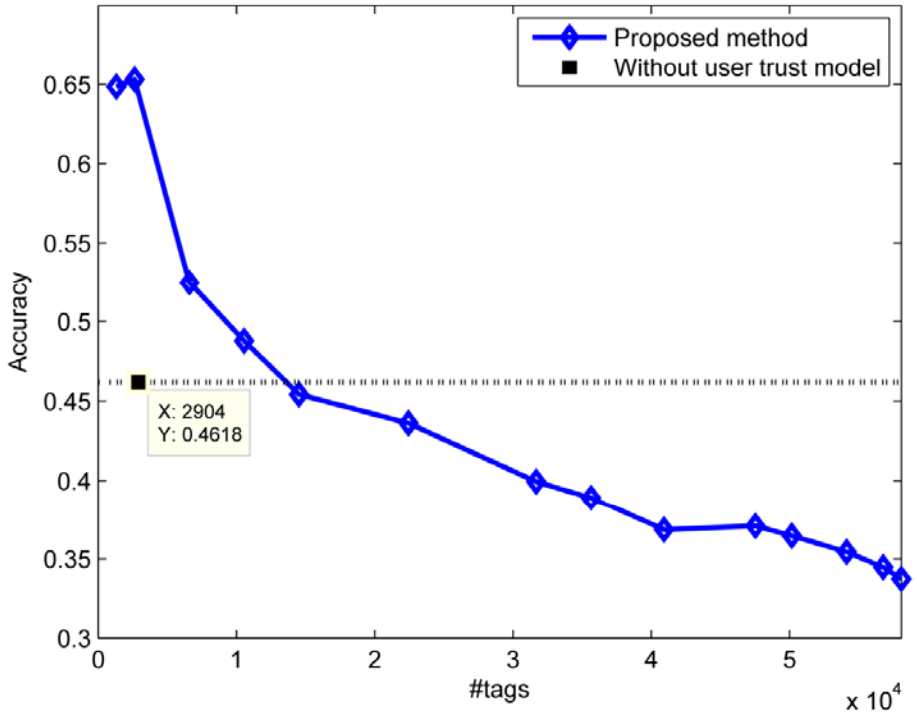
Results

- The recognition rate of the geotag propagation system and the percentage of the propagated tags versus the threshold involved in user trust model



Results

- The recognition rate of the geotag propagation system versus the number of tags



Conclusion

- An efficient system for **automatic geotag propagation** by associating locations with distinctive landmarks and using **object duplicate detection** for tag propagation
- **User trust information** derived from users' tagging behavior for the tag propagation
- Only reliable geotags are propagated
- **Increased accuracy** of the tag propagation and a **decrease of tagging efforts**
- The proposed user trust model can be generalized to photo sharing platforms such as Panoramio

Thank you for your attention!

Ivan Ivanov, PhD student, EPFL
ivan.ivanov@epfl.ch