

# Surprise, You've Got Friends!

## Problem Definition:

Want to present the serendipitous connections?  
How do we define and mine surprise?

## Aspects of Surprise:

Uncanny vs Unexpected  
Surprise can be measured as the change in prior to posterior probability, is time dependent.

## Approaches/Main Idea

Explicit communities, e.g. flickr group, are also connected by implicit groups, e.g. personomy measures, co-location via geo tags. Can find two people at this moment, this place, with the greatest Explicit Distance, and smallest Implicit Distance, this may surprise or abhor these people.

## Weighting

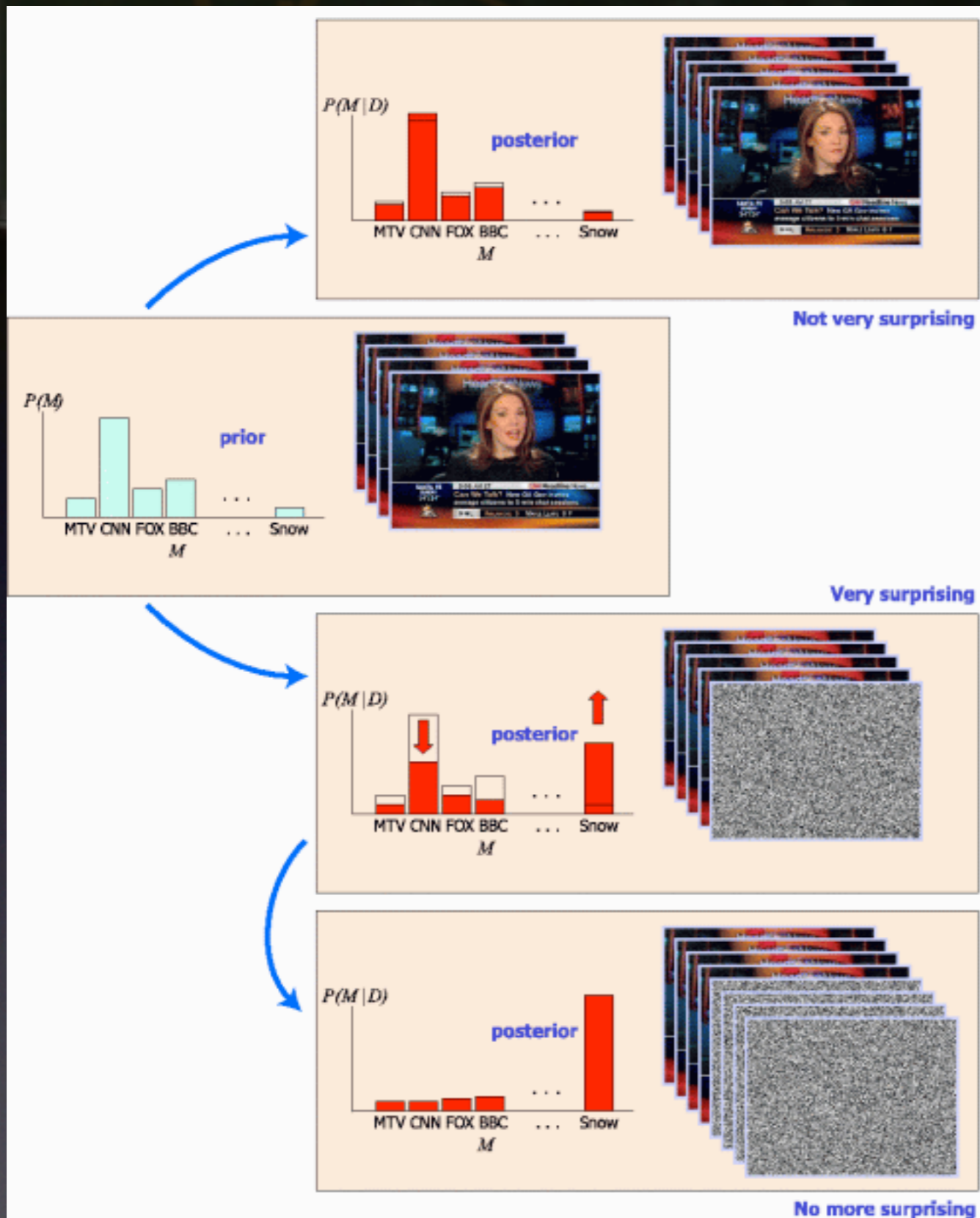
Consider geo similarity. Inverse weight geo location by radius of bounding box, or "fame" of location.

## Test Case

Use geo tagged flickr pictures to find two strangers at a gathering that have a visited the same unusual location in the past.

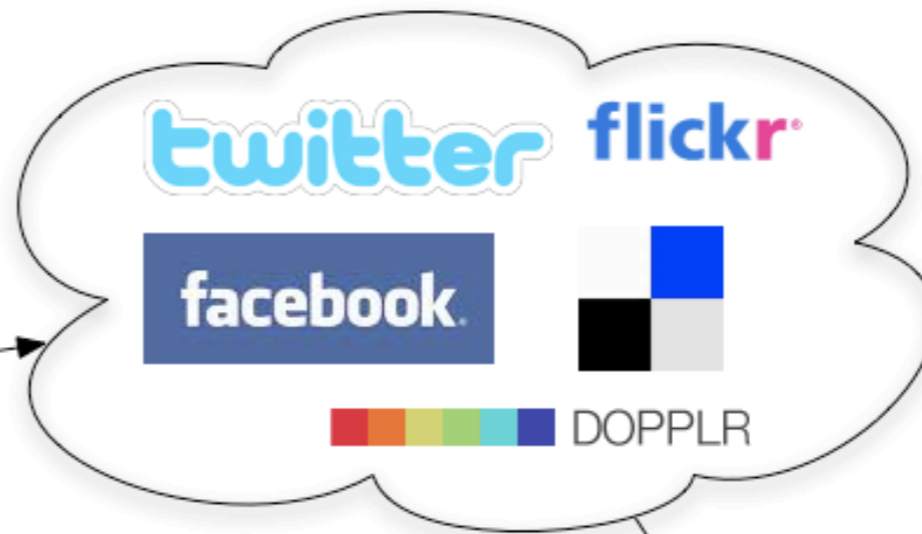
Then

Now



A unit of surprise --- a wow may be defined for a single model  $M$  as the amount of surprise corresponding to a two-fold variation between  $P(M|D)$  and  $P(M)$ , i.e., as  $\log P(M|D)/P(M)$  (with log taken in base 2)

- [L. Itti & P. Baldi, Proc. IEEE-CVPR, 2005](#)
- [L. Itti & P. Baldi, Proc. NIPS, 2006](#)

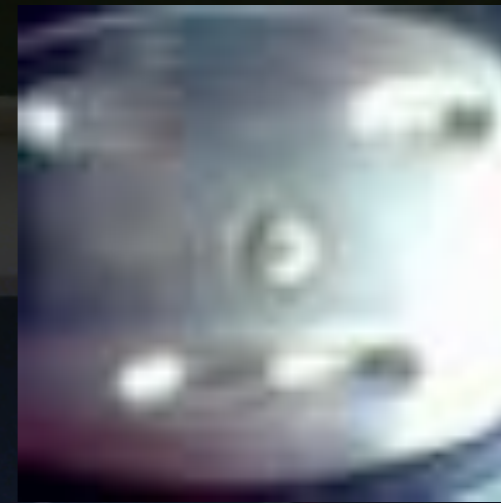


## First approximation

Took the flickr group ISWC

1. Extract the tag distribution of the event based on the photos in the group(E)
2. For each user:  
    Took the user vector and computer the TF-IDF using the group factor as the IDF
3. For each pair of users  
    Compute the cosine similarity

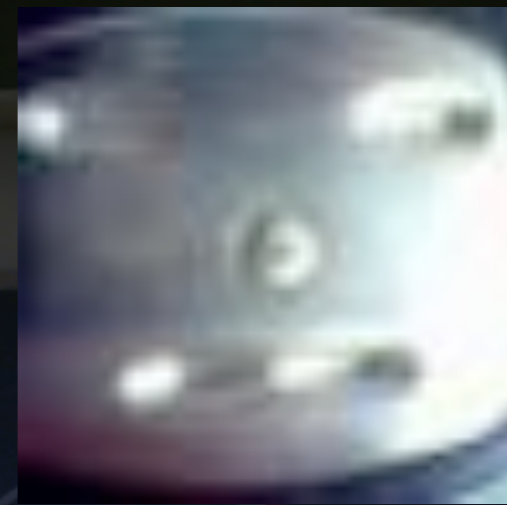
## Results



The two “Closest - furthest” people in this group are from the same research institute in Ireland, DERI



## Base Case



The base-case, using all the information, gives the same result, however the first pass analysis was naive



Plenty of signals that we wanted to take into account:

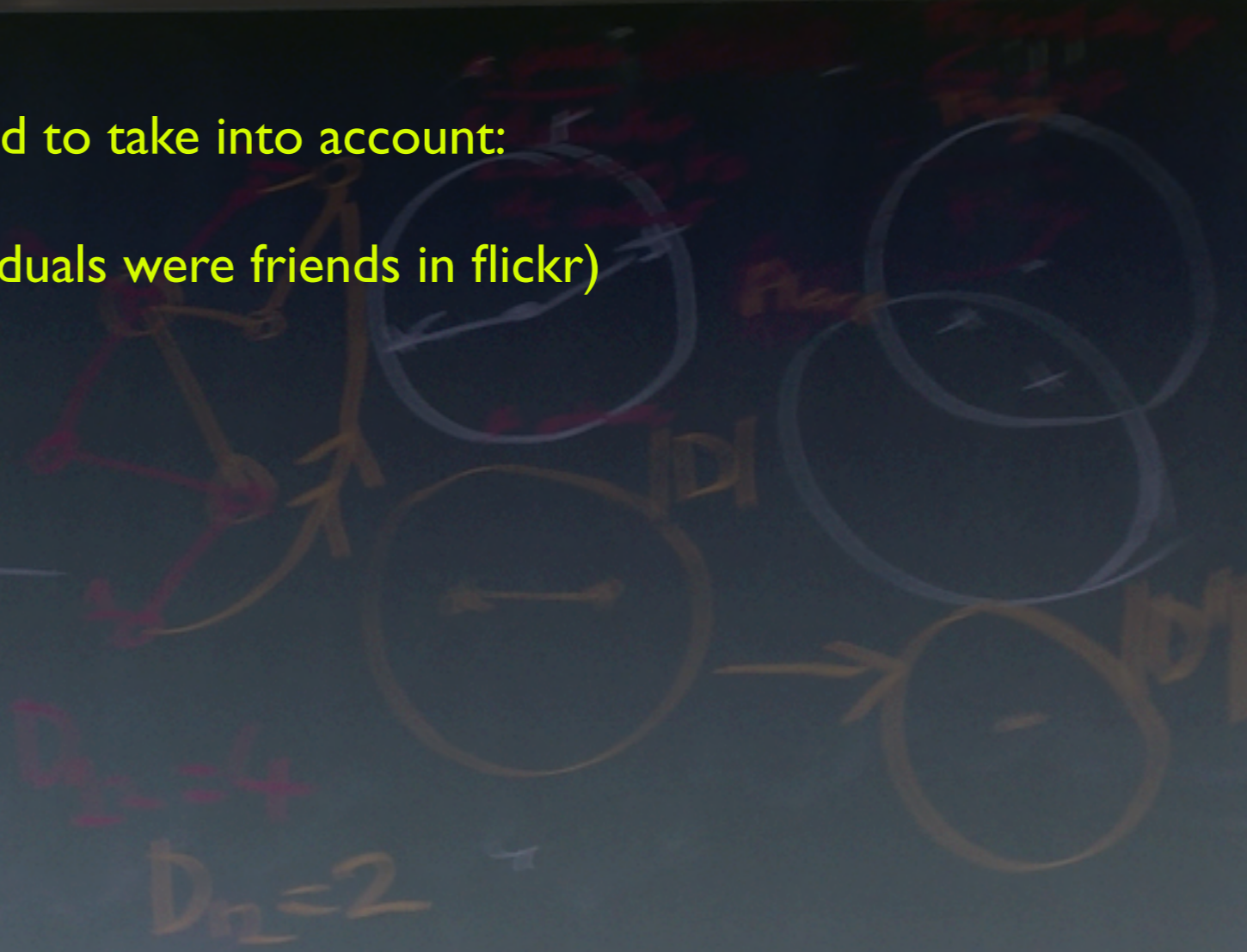
Network of friends (two individuals were friends in flickr)

All tags from all users

Geo Coordinates

etc etc etc

theory → Unoperated



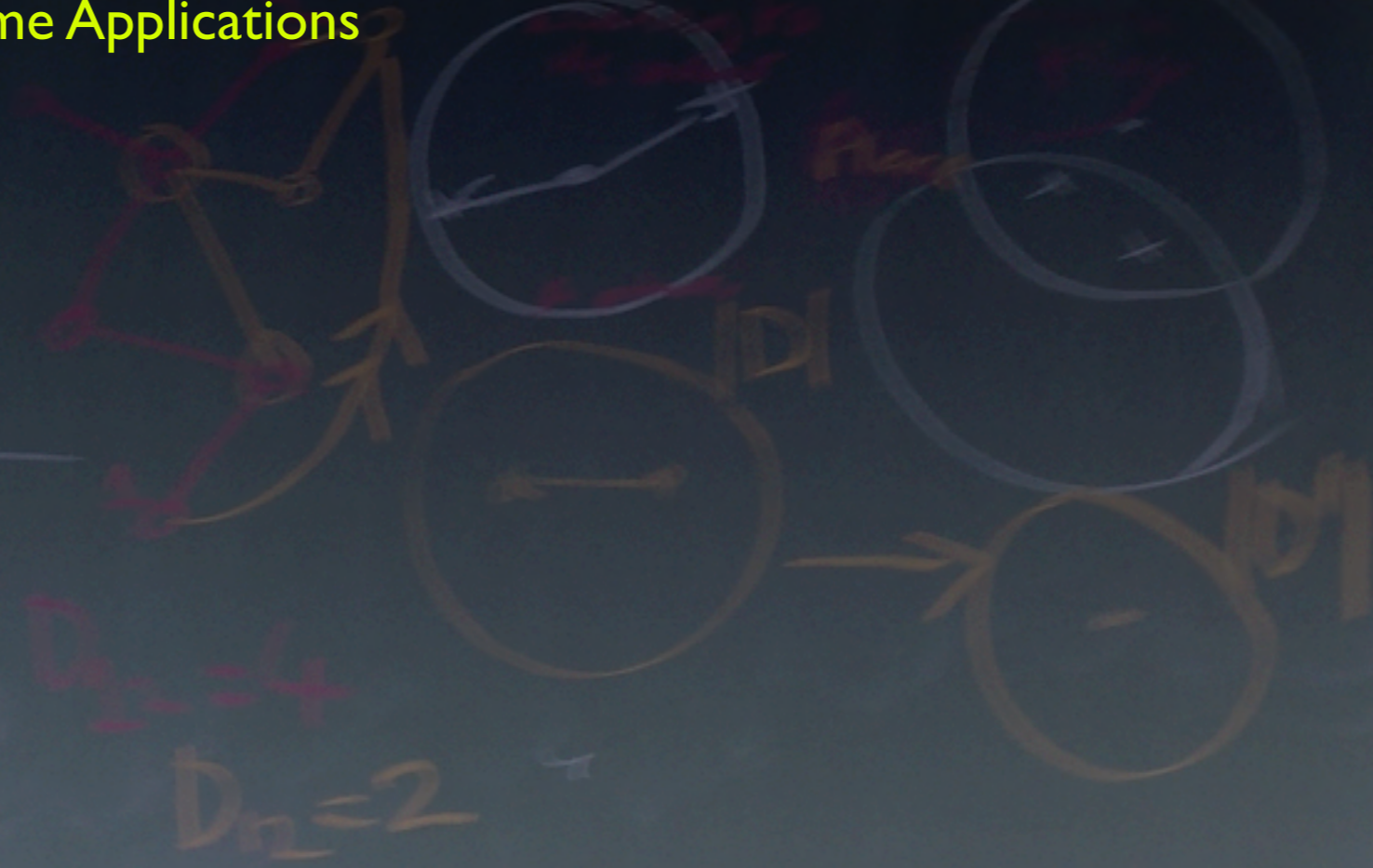
theory → Unoperated

Unoperated Conventions

### Some Applications

Dating Service

For Academics



## Some Broader Issues

Surprise is time dependent, current signals from some SNS are highly bursty, e.g. batch upload in flickr

Connecting this analysis to real world sensors would seem to offer a method for extracting better information on calculating priors.

Connecting across data sources certainly will

A ~~Pleasure Device~~ "Surprise Engine" may become a tunable mess, much like current search engines.