A. NOULAS, M. MUSOLESI, M. PONTIL, C. MASCOLO

Inferring Interests from Mobility and Social Interactions

University of Cambridge, Computer Lab, NetOS talklet

December 1, 2009 in Cambridge, UK

Anastasios (aka Tassos) Noulas, PhD Student

Technological Network

People Interests

Goal : Exploit social interactions to infer interests

A multi-label binary classification problem

Domain : Conference environments

Outline

- Create a similarity graph based on user mobility
- Extract interest labels from data sets
- Full knowledge scenario : predict interests of generic user
- Partial knowledge scenario : predict interests of unlabelled users

Mobility based Similarity Graphs

- user representation:

location 1	location 2	location 3	...			...	location n
23	10	0					4

Representations : duration, frequency and colocation.

- distance in multi-dimensional space : euclidean, manhattan, hamming and many others

- similarity inverse of distance

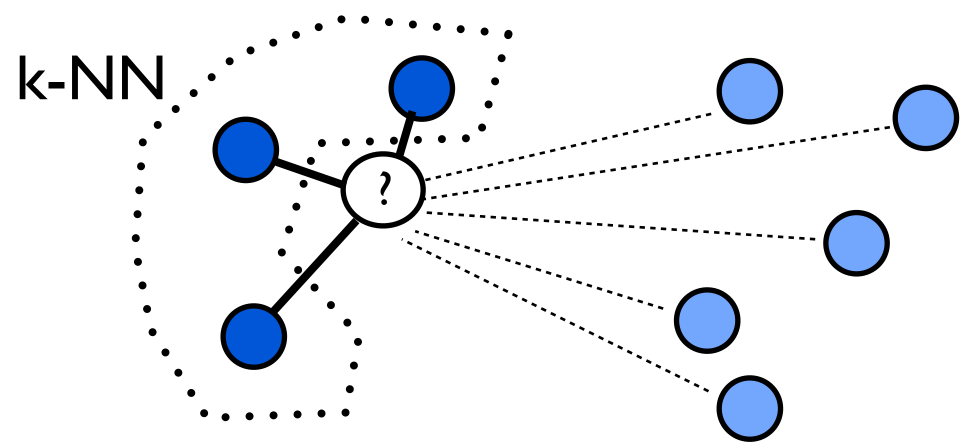
Questionnaires : from interests to binary labels

Interest	Answer	Label
"Ad Hoc Nets"	☑	+1
"Multimedia"	☐	-1
"Sensor Nets"	☑	+1
"Security"	☐	-1
"Traffic Analysis"	☑	+1
...	...	...

Datasets

Dataset	Infocom'06	AMD HOPE
Total Users	78	1200
Active Users	62	410
# Interests	35	21
# locations	17	21
Tracking	Bluetooth	RFID
Duration	3 days	4 days

Scenario 1 : Full knowledge on user interests (Supervised Learning)

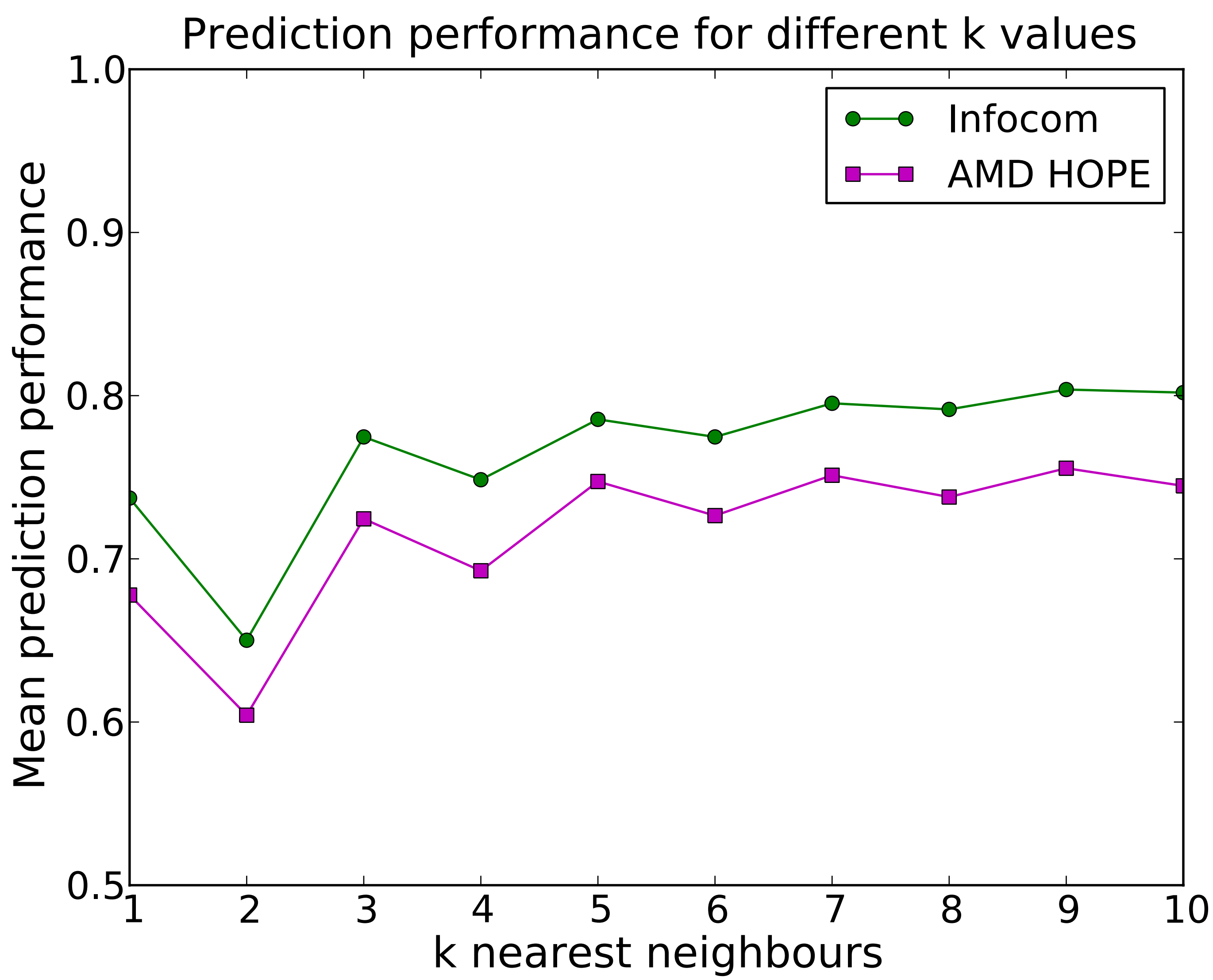


Goal : Predict interest of generic user

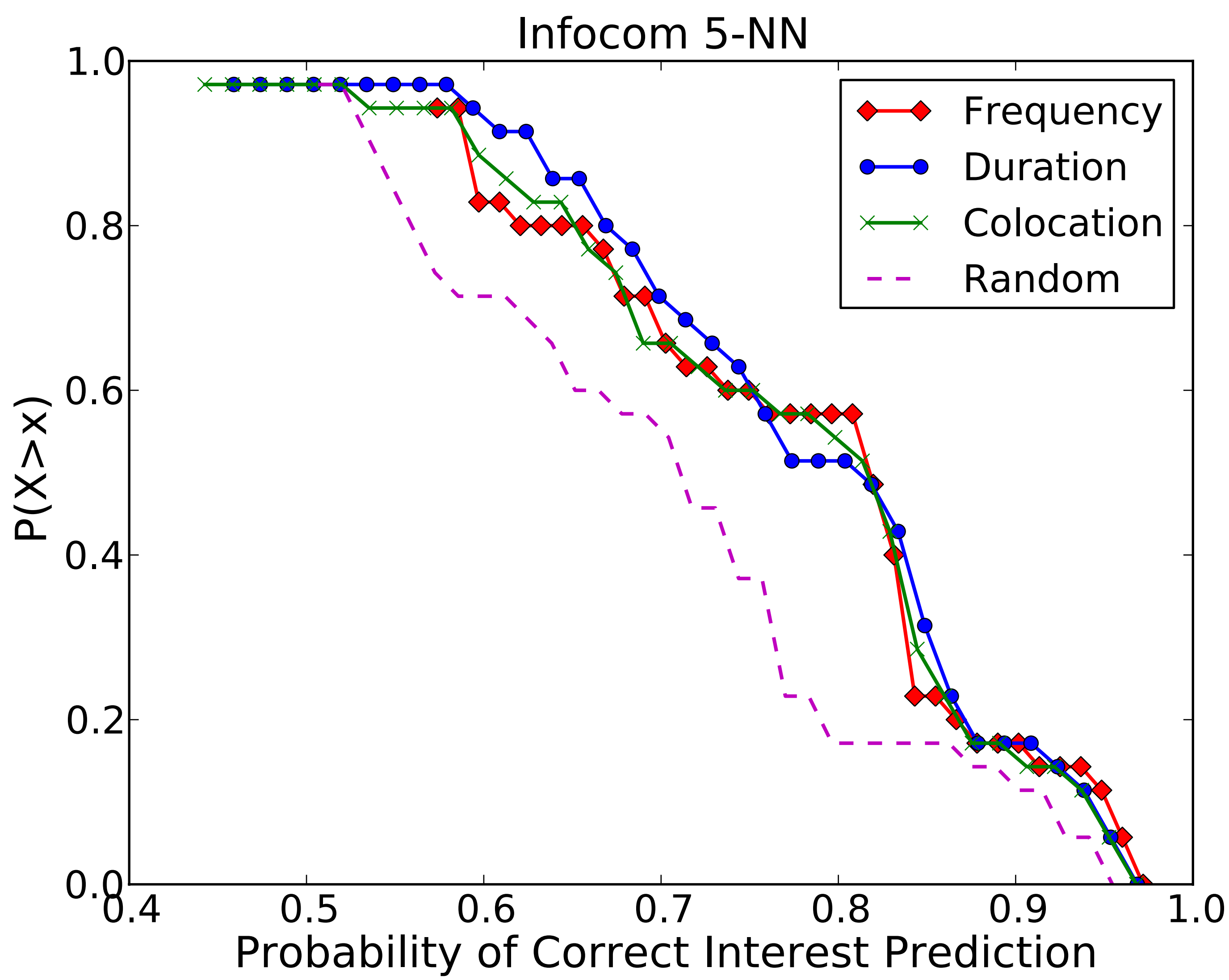
Insight : "Homophily"

Techniques : k-Nearest Neighbours and Leave-One-Out Error evaluation

Results, Supervised Learning (1) :



Results, Supervised Learning (2) :

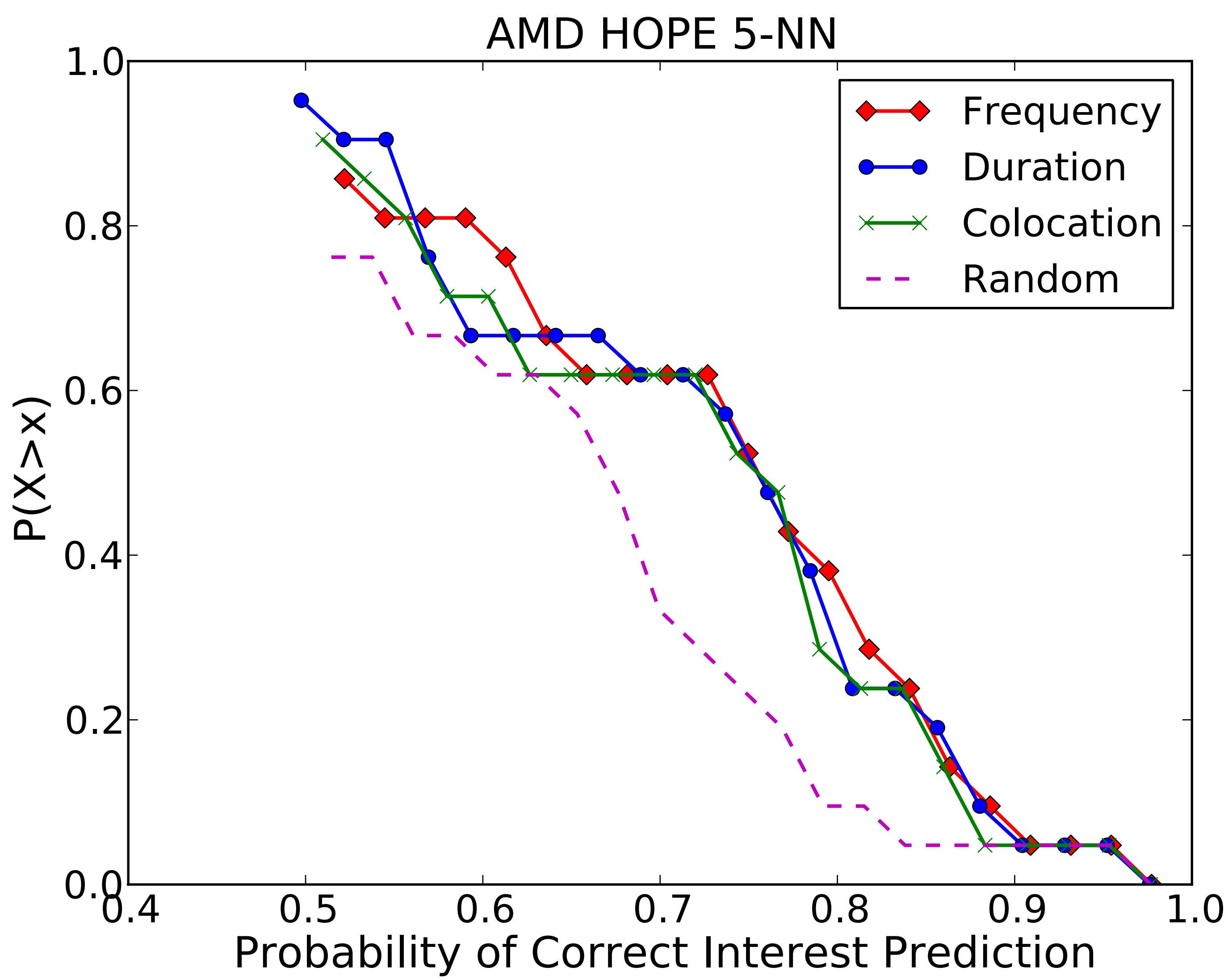


Random interest prediction: $P_{random}(i) = \left(\frac{t}{n}\right)^2 + \left(1 - \frac{t}{n}\right)^2$

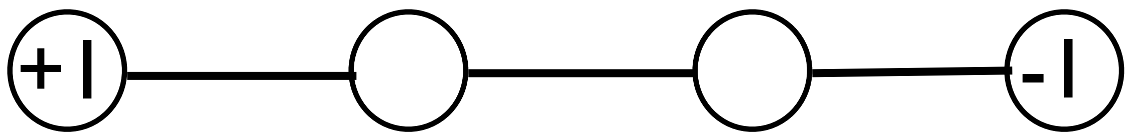
Results, Supervised Learning (3) :

Interest:	”Ad Hoc Nets”	”Multimedia”	”Sensor Nets”	”Security”	”Traffic Analysis”	(mean)
Colocation	0.54	0.85	0.83	0.62	0.67	0.78
Frequency	0.62	0.85	0.83	0.70	0.60	0.78
Duration	0.72	0.86	0.85	0.70	0.63	0.79
Random	0.50	0.77	0.72	0.58	0.55	0.71

Results, Supervised Learning (4) :



Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

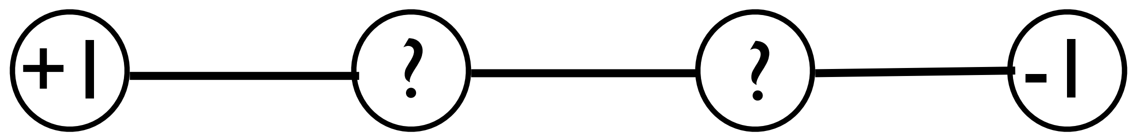


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

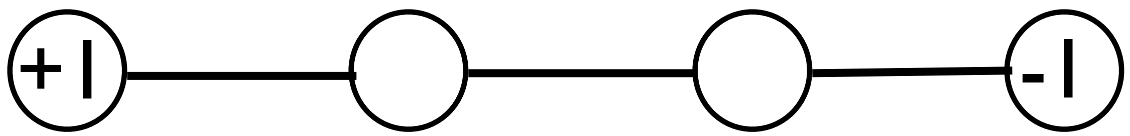


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

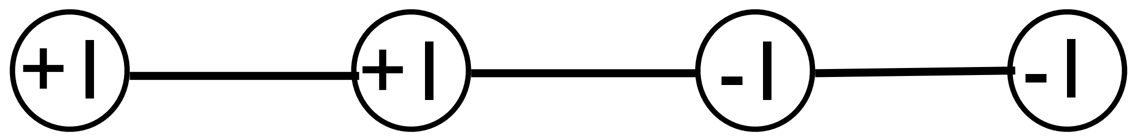


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

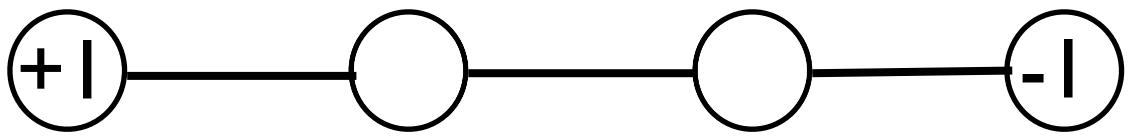


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

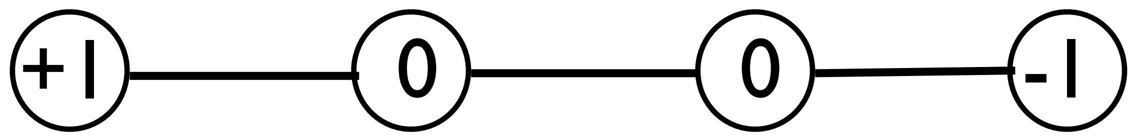


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

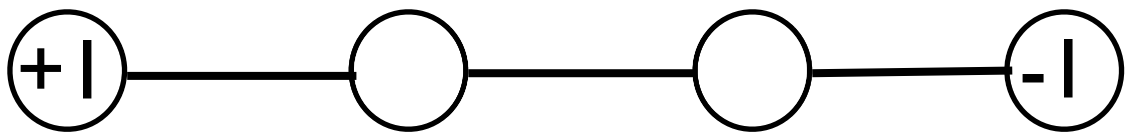


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

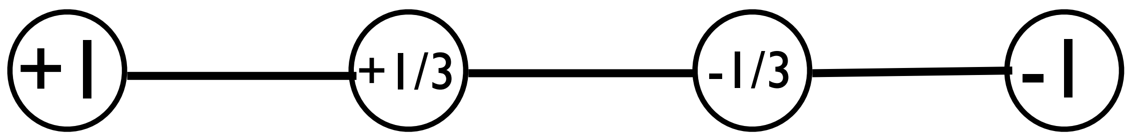


Goal : Predict interest of unlabelled users

Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Scenario 2 : Limited knowledge of user interests (Semi-Supervised Learning)

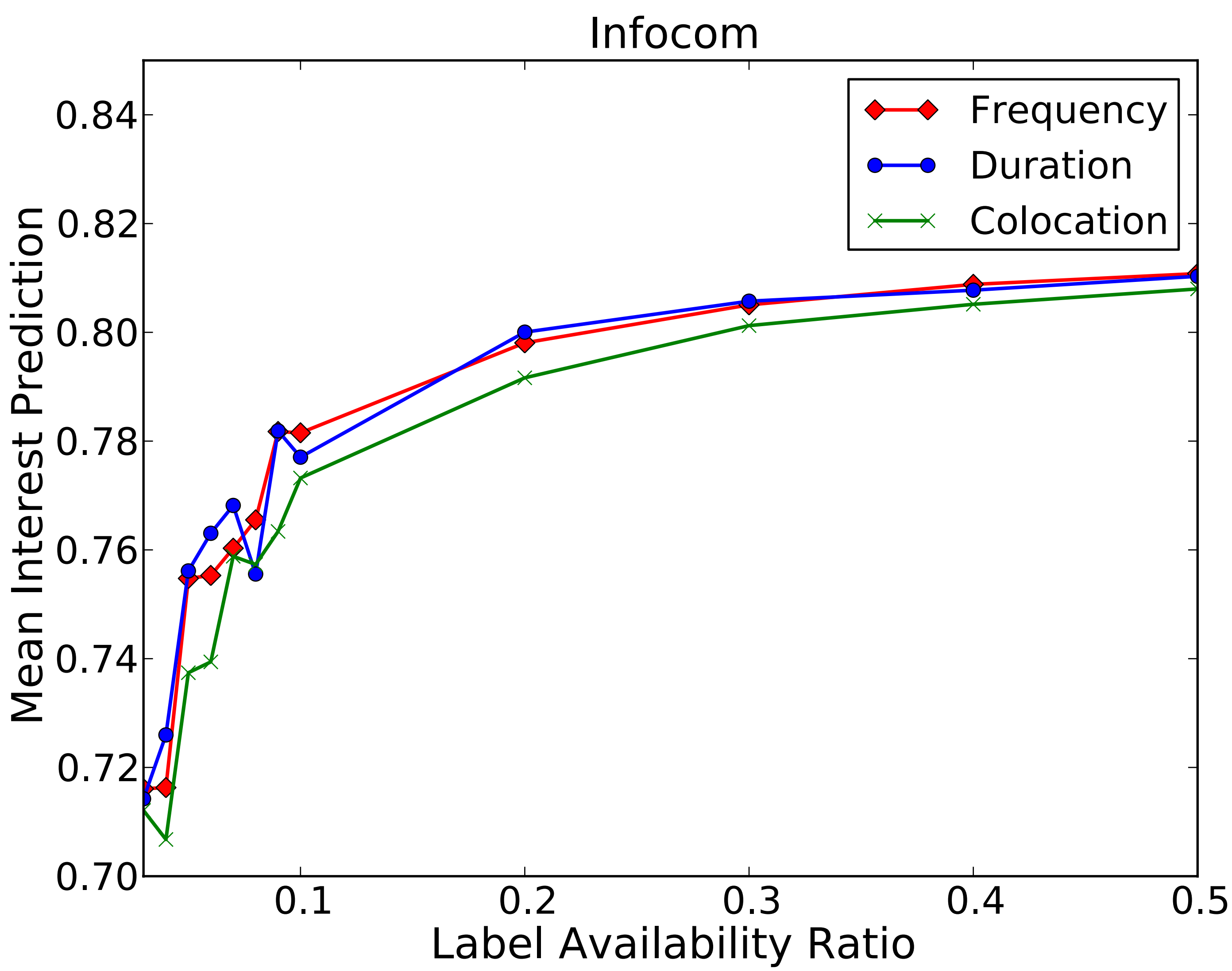


Goal : Predict interest of unlabelled users

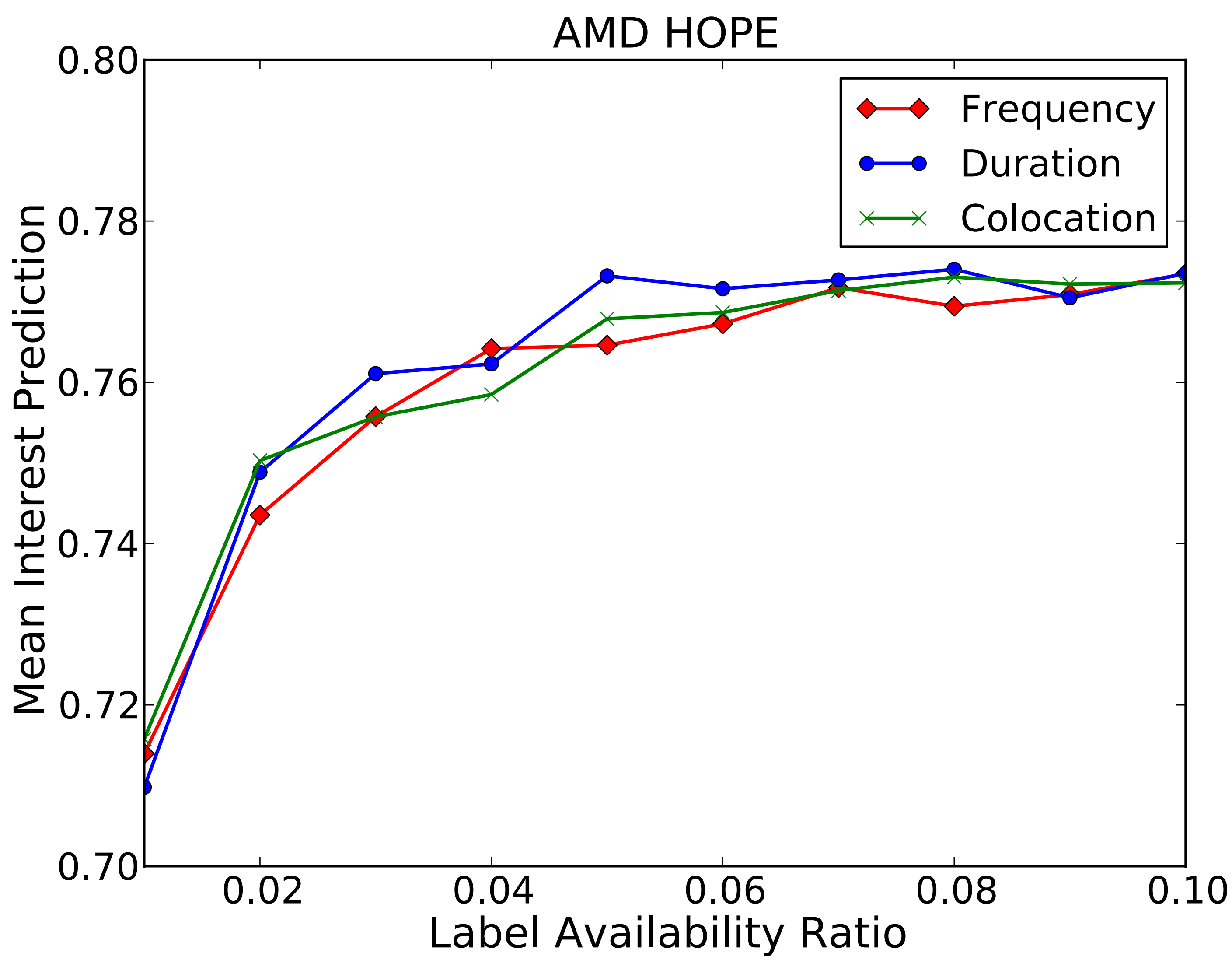
Techniques : Label propagation in a graph using Gaussian Random fields.

Intuition : Neighbouring nodes will have similar values

Results, Semi-Supervised Learning (I) :



Results, Semi-Supervised Learning (2) :



Conclusions and future work

- Mobility information can be used to predict user interests
- Little information is enough to achieve good prediction results
- Can we achieve better accuracy on the prediction task? Other representations or ML algorithms

Thanks!

Download url : www.cl.cam.ac.uk/~an346/

Selection of interests :