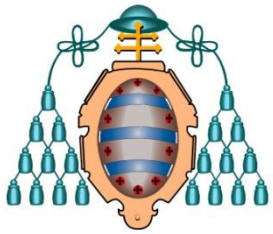
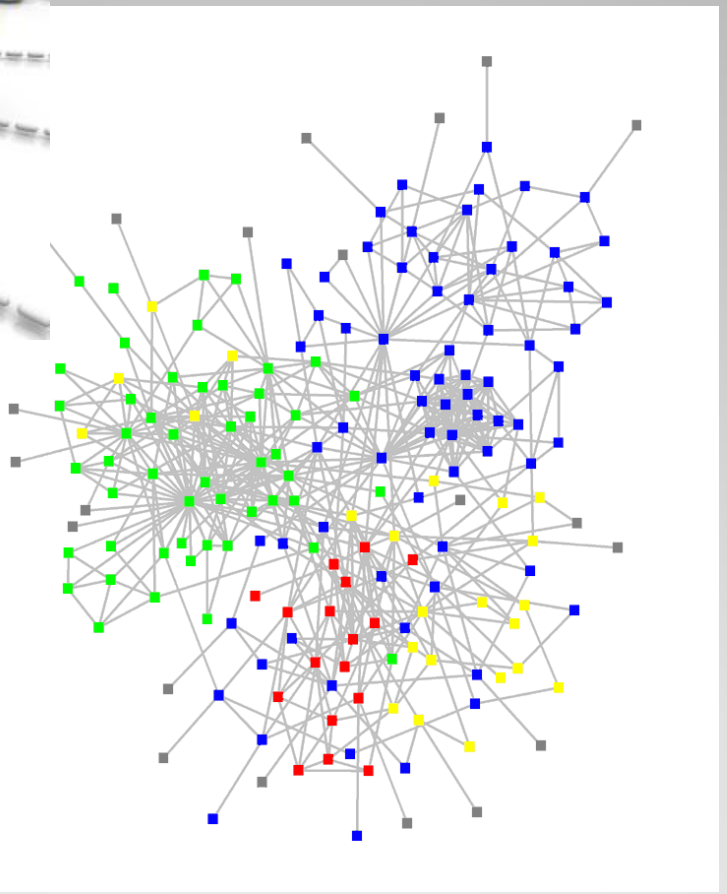
Ranked Tag Recommendation Systems based on Logistic Regression

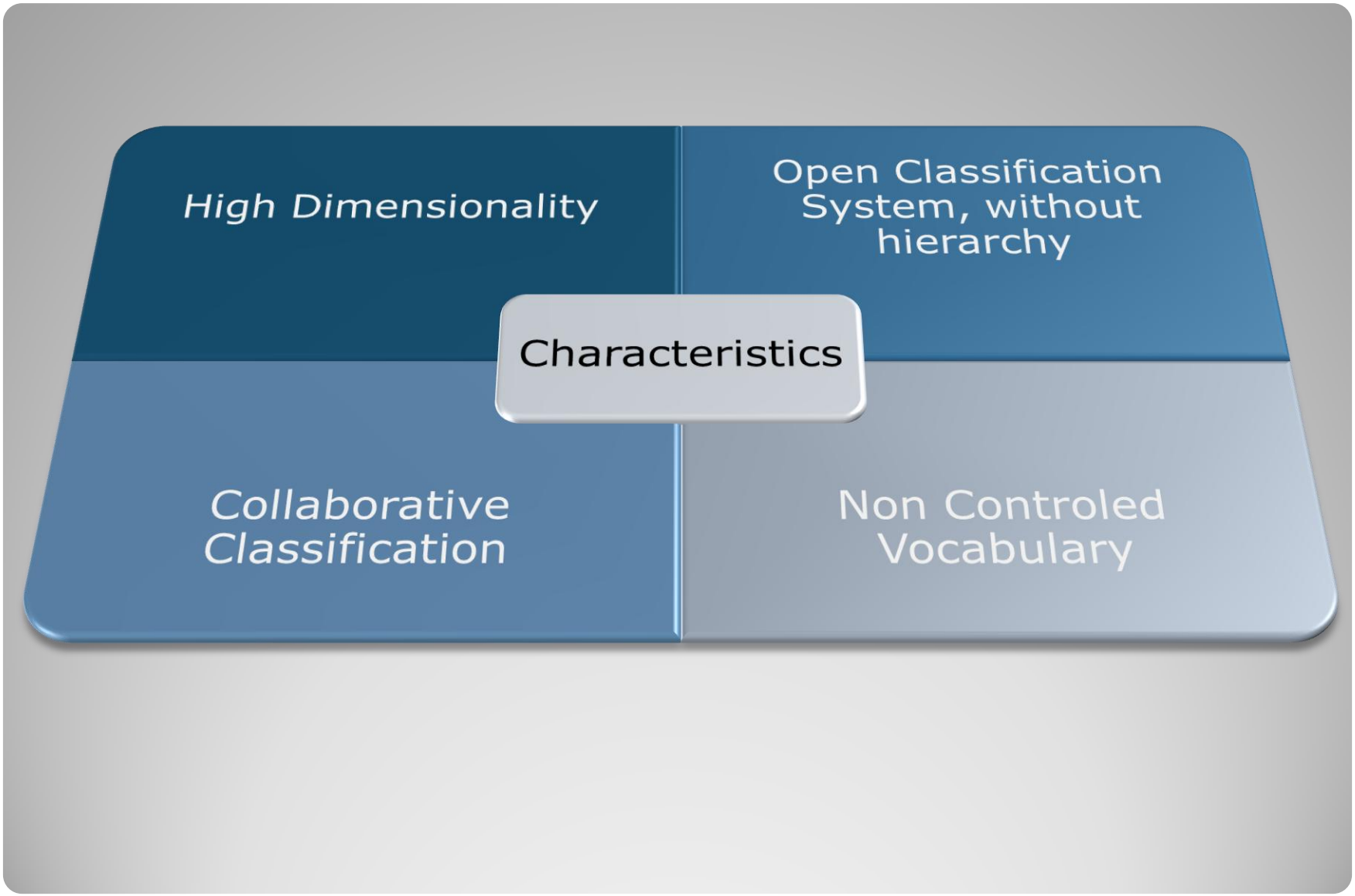
José Ramón Quevedo, Elena Montañés, José Ranilla e **Irene Díaz**
University of Oviedo



San Sebastián, June 25th, 2010



Social Networks



Social Networks

Example: Bibsonomy

- Folksonomy

Users: $\mathbb{U}$

Resources: $\mathbb{R}$

Tags: $\mathbb{T}$

Posts $\mathbb{Y} \subseteq \mathbb{U} \times \mathbb{T} \times \mathbb{R}$

Tag Recommendation System

MPT (Most Popular *Tags):*

The tags occuring most often are recommended

MPTR (Most Popular Tags by Resource):

The tags occurring most often together with r_i are then proposed as recommendations.

MPTU (Most Popular Tags by User):

The tags occurring most often together with u_i are taken as recommendations.

MPTRU (Most Popular Tags by Resource or User):

The tags occurring most often together with either r_i or u_i are taken as recommendations.

Most commonly used TRS

Our Approach

SVM based

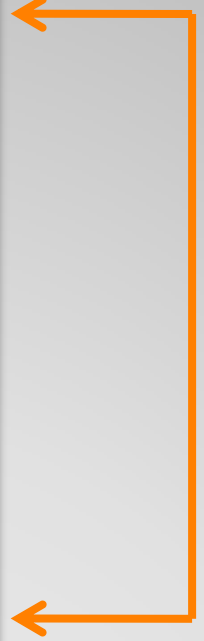
- LibLinear: It provides a probabilistic distribution before the classification, used to infer an order over tag set

This probability distribution is exerted to rank the tags, taking as most suitable tag the one with highest probability value.

One different training set for each post to label.

Post	User	Resource	Tag
P1	u1	r2	??????
P2	u1	r2	t3
P3	u3	r1	t4, t2
----	-----	-----	------
ps	u2	r3	t4
Pj	u1	r3	t4
Pi	u5	r2	t4
Pm	uk	rl	td

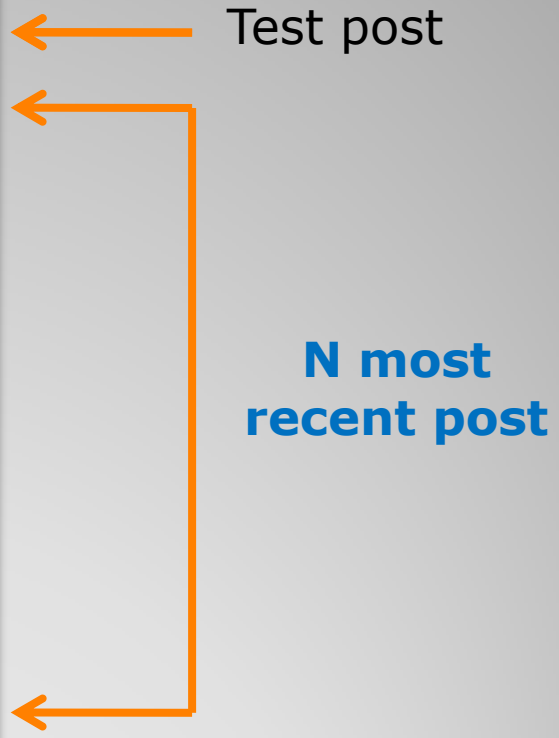
Test post



N most recent post

How to construct training set

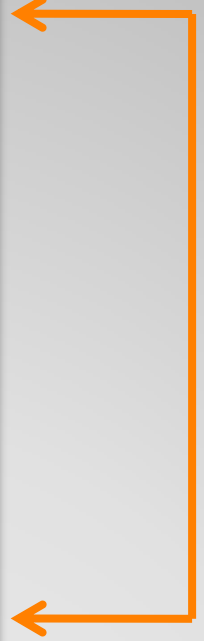
Post	User	Resource	Tag
P1	u1	**r2**	??????
P2	u1	**r2**	t3
----	-----	**-----**	------
Pi	u5	**r2**	t4



How to construct training set: TR

Post	User	Resource	Tag
P1	u1	r2	??????
P2	u1	r2	t3
P3	u3	r1	t4, t2
----	-----	-----	------
ps	u2	r3	t4
Pj	u1	r3	t4
Pi	u5	r2	t4
Pm	uk	rl	td

Test post



N most recent post

How to construct training set

Post	User	Resource	Tag
P1	**u1**	r2	??????
P2	**u1**	r2	t3
----	-----	-----	------
Pj	**u1**	r3	t4

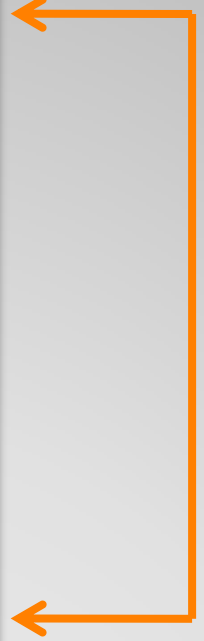
Test post

N most recent post

How to construct training set: TU

Post	User	Resource	Tag
P1	u1	r2	??????
P2	u1	r2	t3
P3	u3	r1	t4, t2
----	-----	-----	------
ps	u2	r3	t4
Pj	u1	r3	t4
Pi	u5	r2	t4
Pm	uk	rl	td

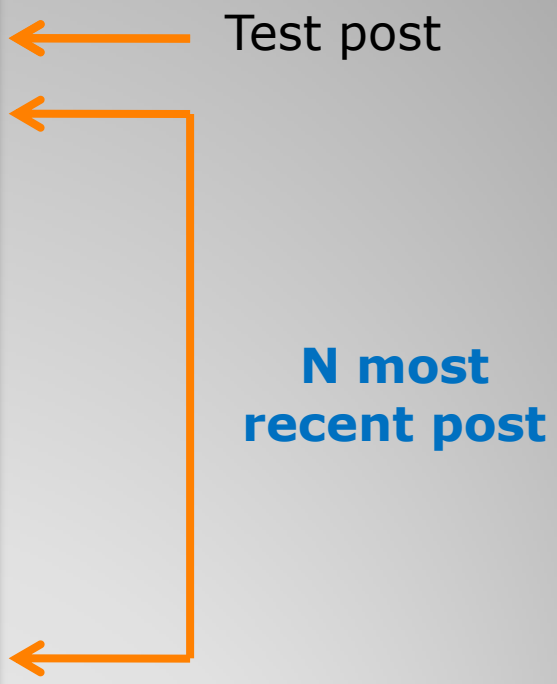
Test post



N most recent post

How to construct training set

Post	User	Resource	Tag
P1	**u1**	**r2**	??????
P2	**u1**	**r2**	t3
----	**-----**	-----	------
Pj	**u1**	r3	t4
Pi	u5	**r2**	t4



How to construct training set: TRU

Representing examples

With the tags used previously to label the resource of the post

Filtering : To represent each pots with the tags with which the resource test is labeled (TRUTR)

The category of the class is the post of the tag

Example	Resource	Feature	Class
u1	r2	t2	t2
u2	r1	t1	t3
u1	r1	t1, t2, t3	t4

Example	Resource	Feature	Class
test	r2	t1, t2	?

Example	Resource	Feature	Class
U1	r2	t2	t2
U2	r1	t1	t3
u3	r1	t1, t2	t4

Example

bt08

- posts *bibtex, Bibsonomy*
- ECML PKDD Discovery Challenge 2008

Users	*Tags*	*Resources*	*Posts*
1206	29739	96616	278008

- **1000 test post randomly selected**

- **N=*i**500, *i*=1, 2, …, 50**
- **Influence of the size of training set**

Results

MPT (Most Popular *Tags*):

The tags occuring most often are recommended

MPTR (Most Popular Tags by Resource):

The tags occurring most often together with r_i are then proposed as recommendations.

MPTU (Most Popular Tags by User):

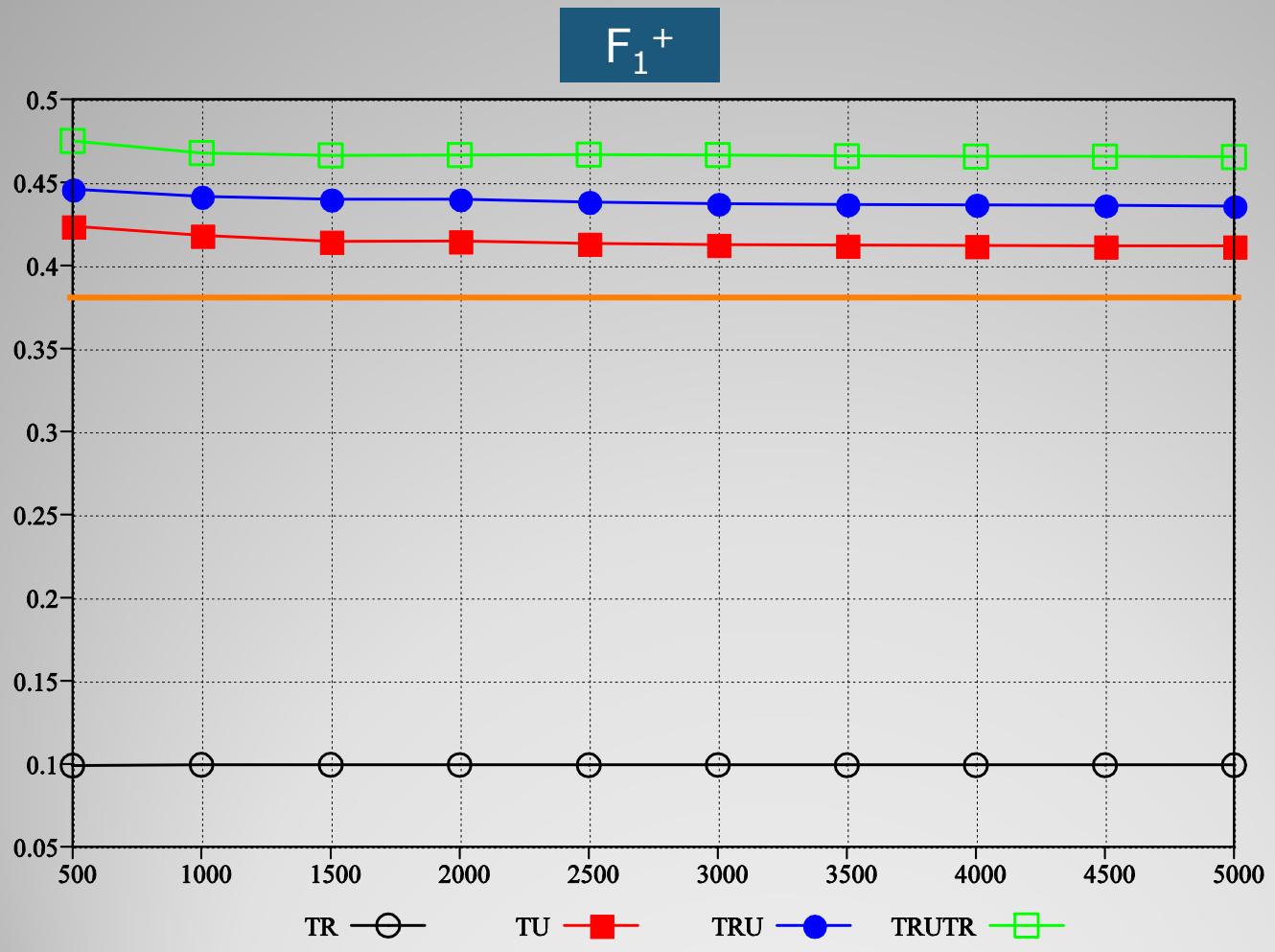
The tags occurring most often together with u_i are taken as recommendations.

MPTRU (Most Popular Tags by Resource or User):

The tags occurring most often together with either r_i or u_i are taken as recommendations.

	F1+
MPT	6.7%
MPTR	7.8%
MPTU	37.2%
MPTRU	38.2%

Most commonly used TRS



Results of our approach

Thank you for your attention