le cnam

A TRUST BASED RECOMMENDATION SYSTEM USING SELF ORGANIZED MAP FOR SERVICE SELECTION

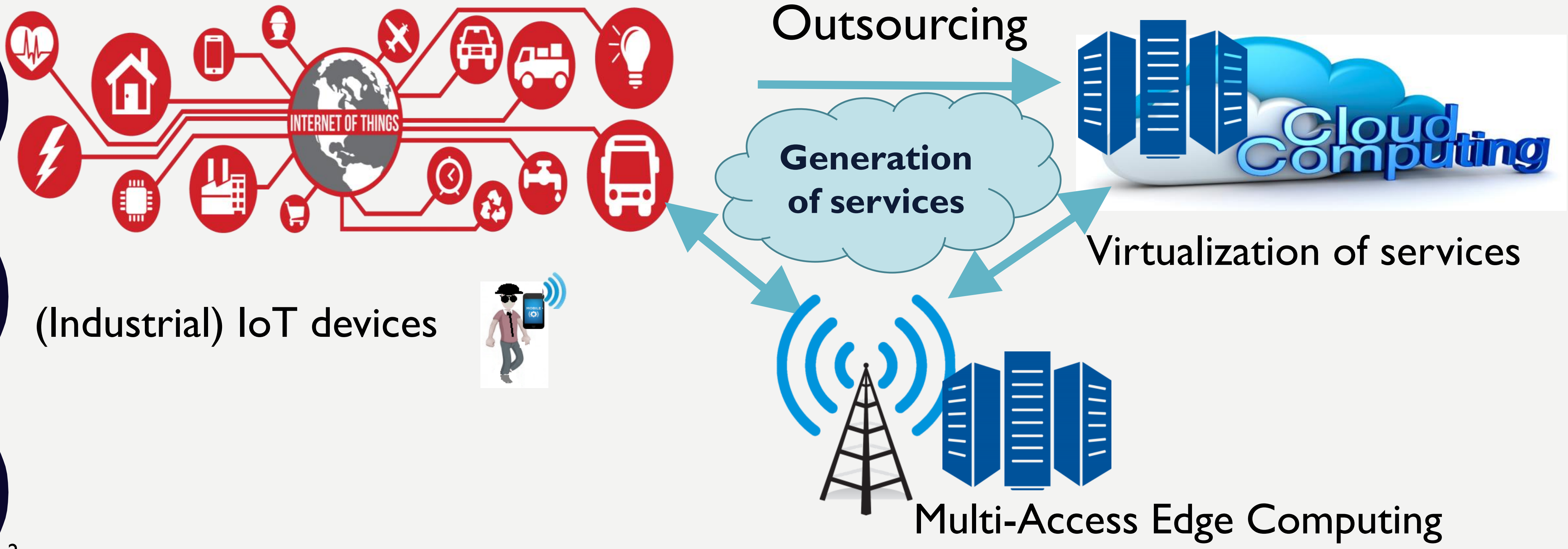
SAMIA BOUZEFRANE

samia.bouzefrane@cnam.fr

CEDRIC LAB – CNAM- PARIS

Youcef Ould-Yahia, Meziane Yacoub, Samia Bouzefrane, Hanifa Boucheneb, " Self Organized Map and trust-aware-based quality of service prediction for reliable services selection in distributed computing environment", to appear in *Int. J. Advanced Intelligence Paradigms.*

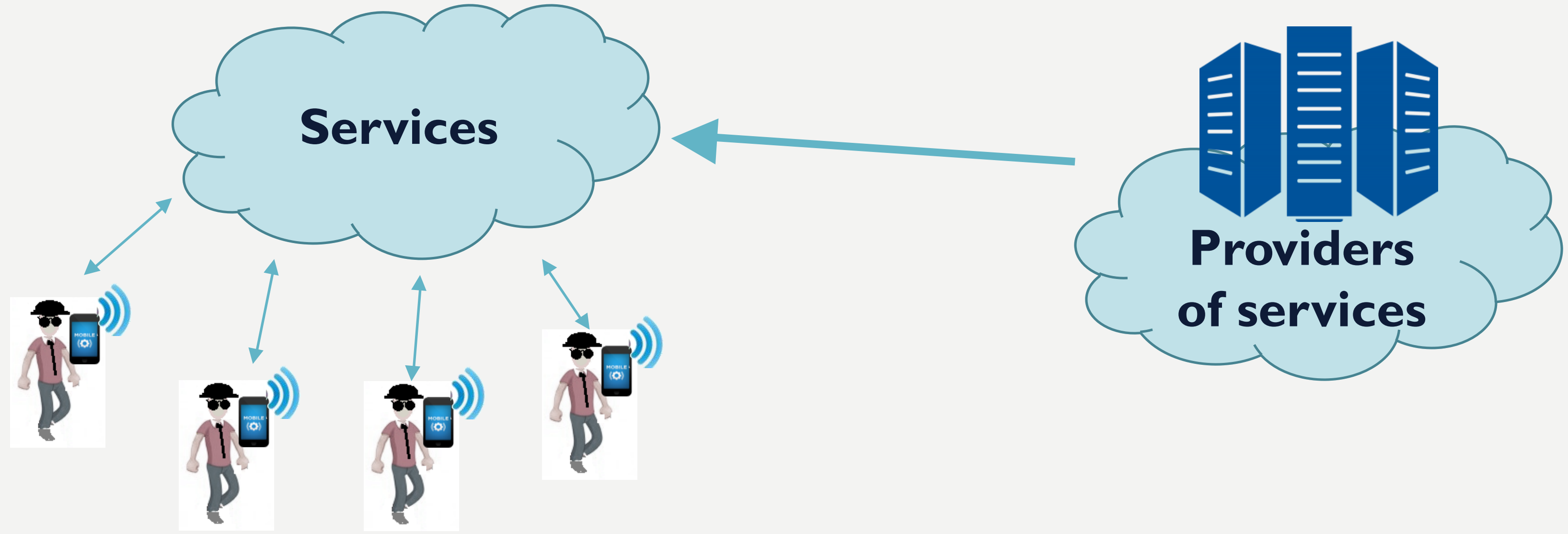
NEW ARCHITECTURE MODELS & PARADIGMS



TRUST OF SERVICES

A trusted service means what?

- A trusted service provider ?
- A secure service ?
- A good QoS to access the service ?
- A good reputation of the service as given by the consumers ?



Consumers of services

TRUST RELATIONSHIP

- Trust relationship can be built based on many indicators:
  - QoS
  - Security
  - Reliability
  - Reputation (given by the requestor or by others)
  - etc.
- The trust can be built efficiently if many criteria can be collected to evaluate the trustworthy of the service
- The issue is how to trust a service requested by a user for which he does not have any trust indicator ?

OBJECTIVE OF THIS WORK

- **Assumption**:
  - The reputation is our indicator of trust.
- **Goal**:
  - Prediction of the trust value of a service that is requested by a user for which he does not have any previous evaluation.

TERMINOLOGY USED

- **Evaluation**: is a numerical data value which is a user's appreciation for a service
  - Based on attributes of the service/service provider: QoS, security, **reputation**, etc.

- **Prediction of the evaluation**: computes the most likely evaluation that a user would have assigned to a given service.

- **Recommendation**: uses prediction to evaluate a list of services and then to propose to the user the most appropriate service.

RECOMMENDATION SYSTEMS

- **Content-based filtering**: recommends new services based on the services already consumed by the user
  - Problem with a new user (the system does not know the preferences of the user)
  - The system will recommend only the services that are similar to the user profile.

- **Collaborative filtering**: can predict a specific service to a user based on the evaluations done by users.
  - This approach assumes that if several users have similar preferences for a group of services, they may have the same preferences for another group of services (if users agree for the evaluation of services in the past, they are likely to agree in the future).

COLLABORATIVE FILTERING

- Collaborative filtering is a part of ML approaches.
- Based on the experiences of the users

- **Neighborhood Based** (memory) **Method**
  – The reputation of a service s is predicted thanks to the feedbacks given by similar (neighbors) users to user u / The reputation is given by user u to similar services as service s
  – Similarity ratings may be resource-intensive (to store the feedbacks of the users)

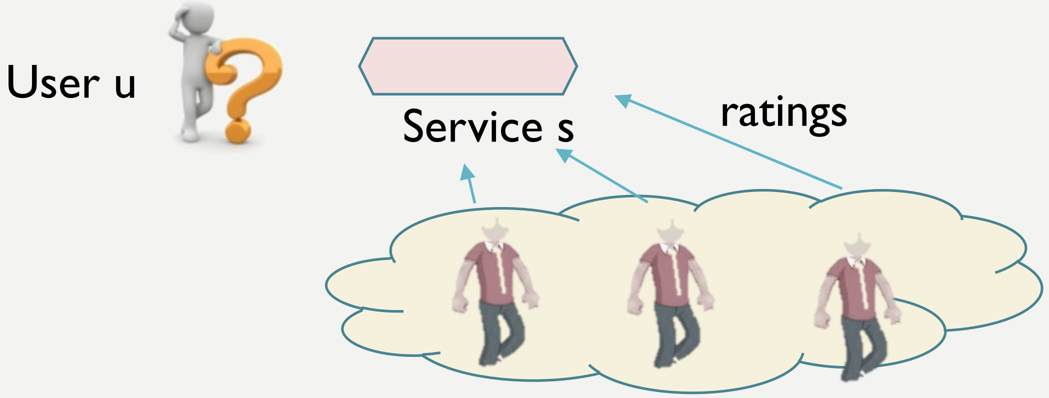
- **Model-Based Recommendation Method**
  – Off-line models (bayesian classification, K-means, PCA, etc.) for prediction.
  – Sensitive to missing data.

NEIGHBORHOOD BASED/1

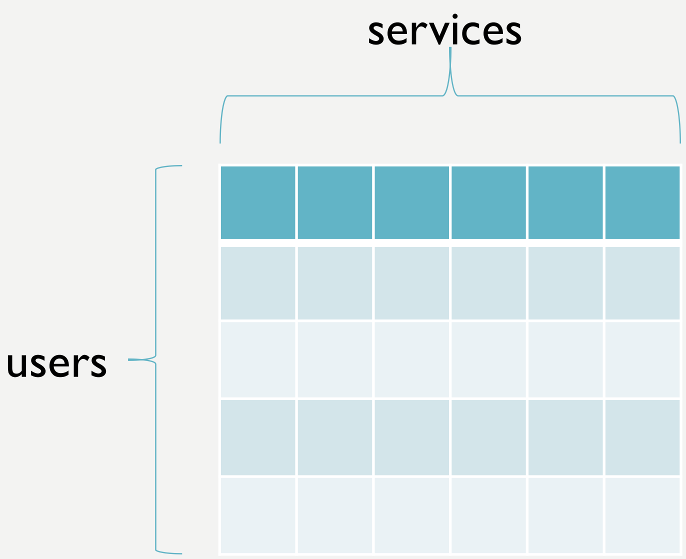
Aim: to predict classifications for new services by two ways:

- **User-based approach**: evaluates the interest of user u for a service s using the ratings given by neighbours of u (who have similar ratings patterns as user u)

- **Service-based approach**: evaluates the interest of user u for a service s using the ratings of u for services that are similar to s



Users with similar profiles



Ratings in the memory

k = argmax f(u , s)
f: regression/classification

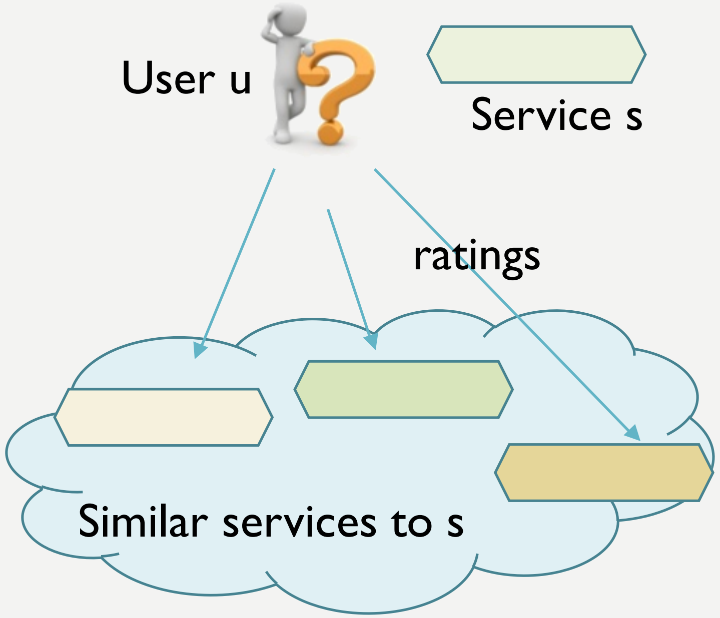
NEIGHBORHOOD BASED/2

Aim: to predict classifications for new services by two ways:

- **User-based approach**: evaluates the interest of user u for a service s using the ratings given by neighbours of u (who have similar ratings patterns as user u)

- **Service-based approach**: evaluates the interest of user u for a service s using the ratings of u for services that are similar to s

  - The model is trained with a dataset
  - The model is then used to predict ratings of users for new services



Model: Support Vector Machine, Bayesian Clustering, etc.

PREDICTING EVALUATIONS IN CF

Similarity between user u and user v
(cosine similarity/Pearson correlation coef.)

Evaluation of service s by user v

$$r'_{us} = \overline{r}_u + \frac{\sum_{v \in U_s} Sim(u,v) \times (r_{vs} - \overline{r_v})}{\sum_{v \in U_s} Sim(u,v)}$$

Predictive value

Users who evaluated service s

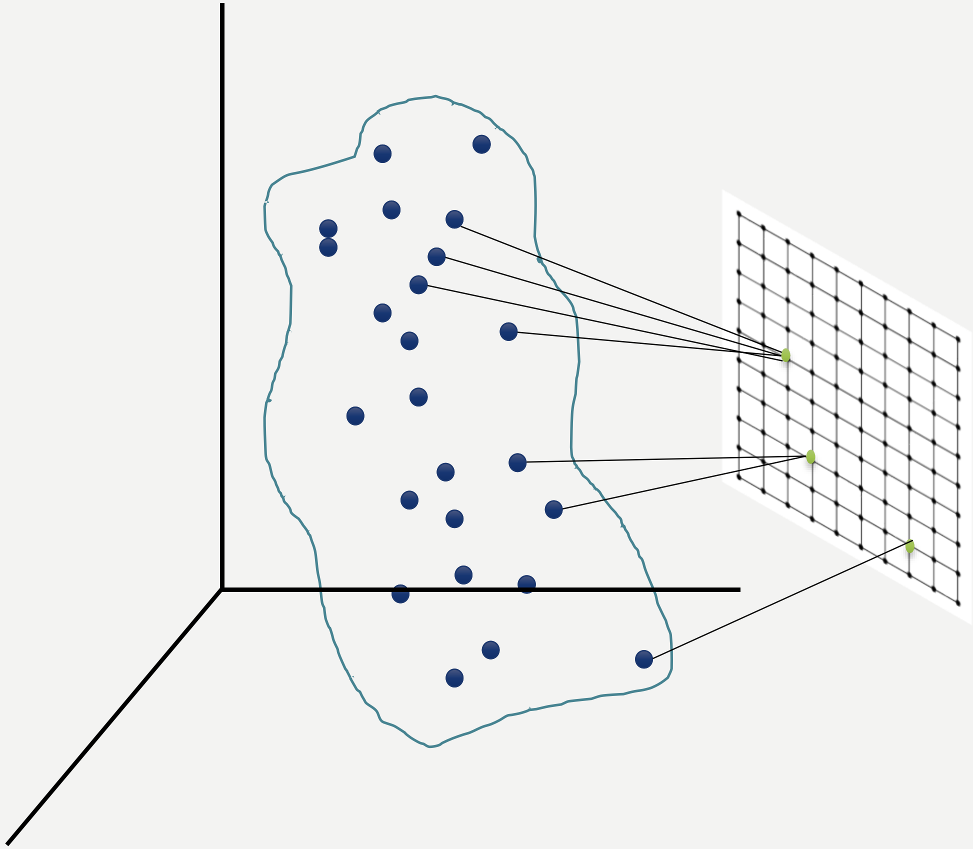
The mean of evaluations given by user u

OUR APPROACH

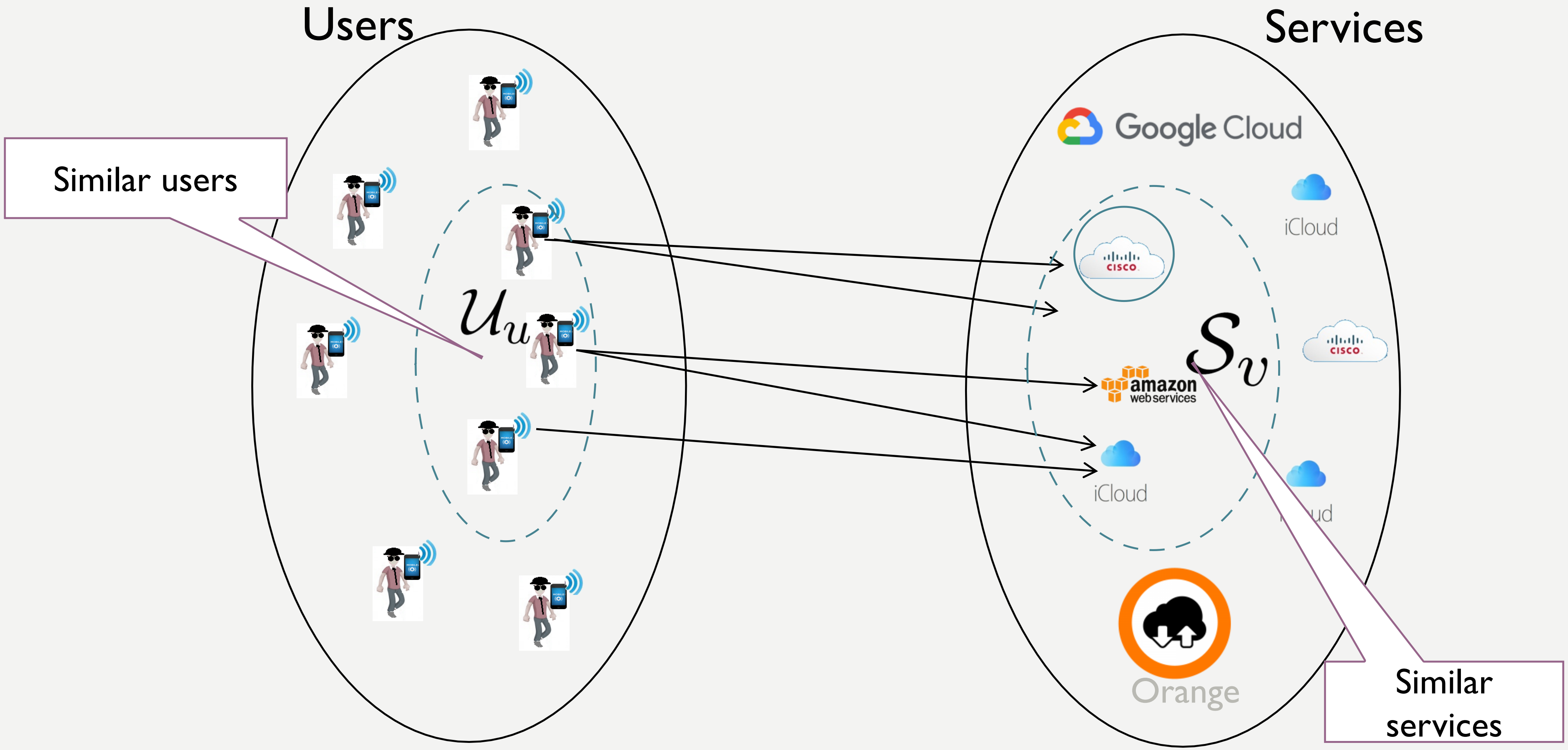
- Based on SOM (Self Organizing Map)
- Recommendation system to predict the trust/reputation of a service
  - How to predict a trusted service which is not already evaluated/known by a user?

SOM: SELF-ORGANIZING MAP

- ✓ Unsupervised learning (neural networks)
- ✓ based on vector quantization
- ✓ Reduction of dimensions
- ✓ Preserve the topological structure of the entry space (neighborhood between classes)
- ✓ Allows data analysis including non-linear relations
- ✓ Resiliency for missing data



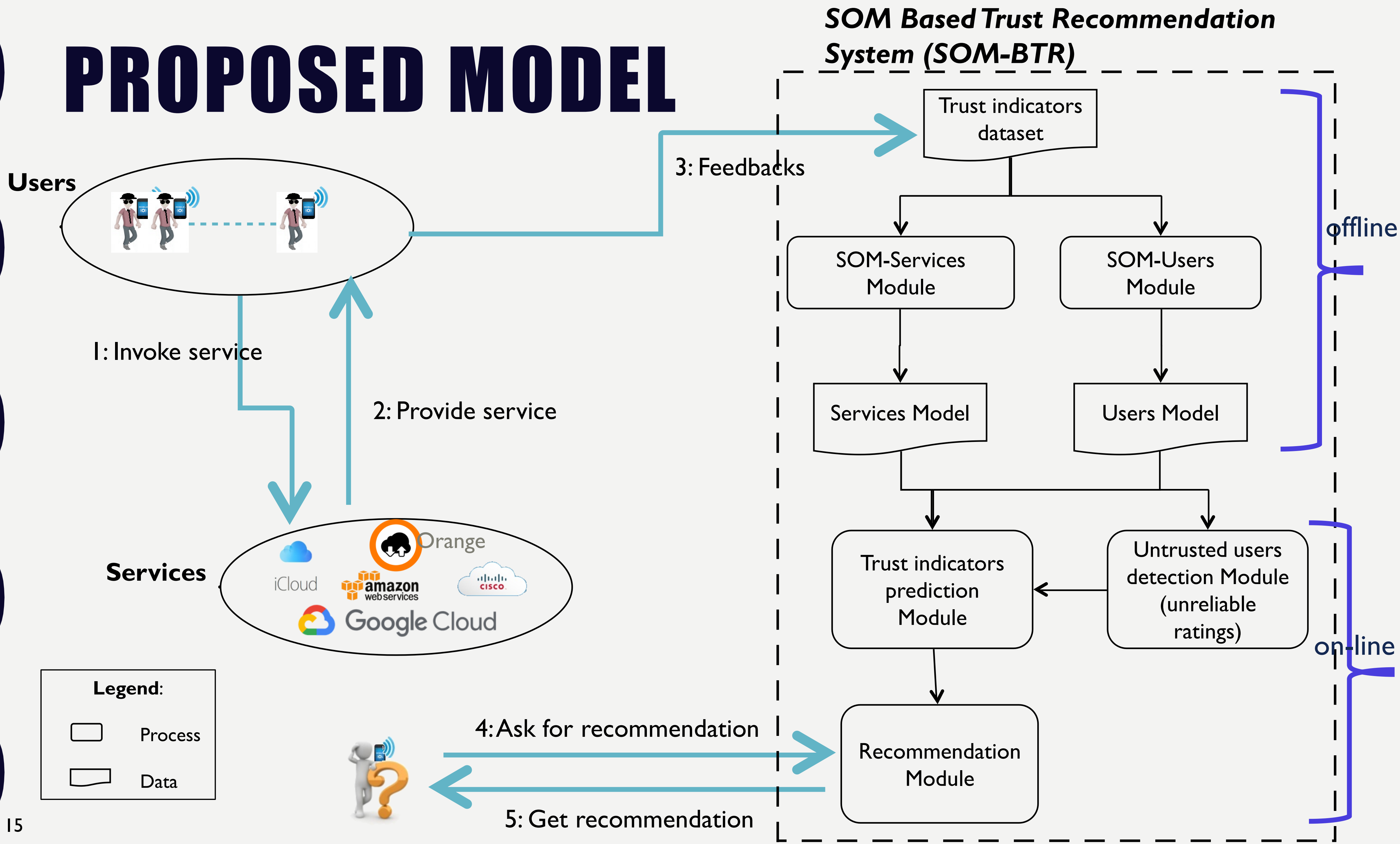
TRUST EVALUATION USING SOM



We consider a set of similar users (to user u) evaluating a set of similar services (to service v)
Similar users will behave like user u when ranking similar services

14

PROPOSED MODEL



15

DESCRIPTIVE AVERAGE

$$N = \{1.8, 1.2, 1.7, 2, 1.7, 47, 75, 1.8, 7, 8\}$$

The mean of set *N is equal to* 14.72

While :

- 60% of the values have a mean equal to 1.7
- 80% of the values have a mean equal to 3.1

47, 75

7, 8

1.8, 1.2, 1.7, 2, 1.7, 1.8

K-means (k=3)

The descriptive average is based on the classification with K-means

DESCRIPTIVE AVERAGE

$$N = \{n_1, n_2, ..., n_i\}$$

We apply K-means on the set *N* $\longrightarrow$ K clusters C_N^i

Ordered like the following : $|C_N^1| \geqslant |C_N^2| \geqslant ... \geqslant |C_N^k|$

We fix the parameter α between 0 and 1 :

$$\frac{|C_N^1| + |C_N^2| + ... + |C_N^j|}{|N|} \geqslant \alpha$$

$$j = \underset{1 \leq m \leq k}{\operatorname{argmin}} \sum_{m=1}^{j} \frac{|C_N^m|}{|N|} \geqslant \alpha, 0 < \alpha \leqslant 1$$

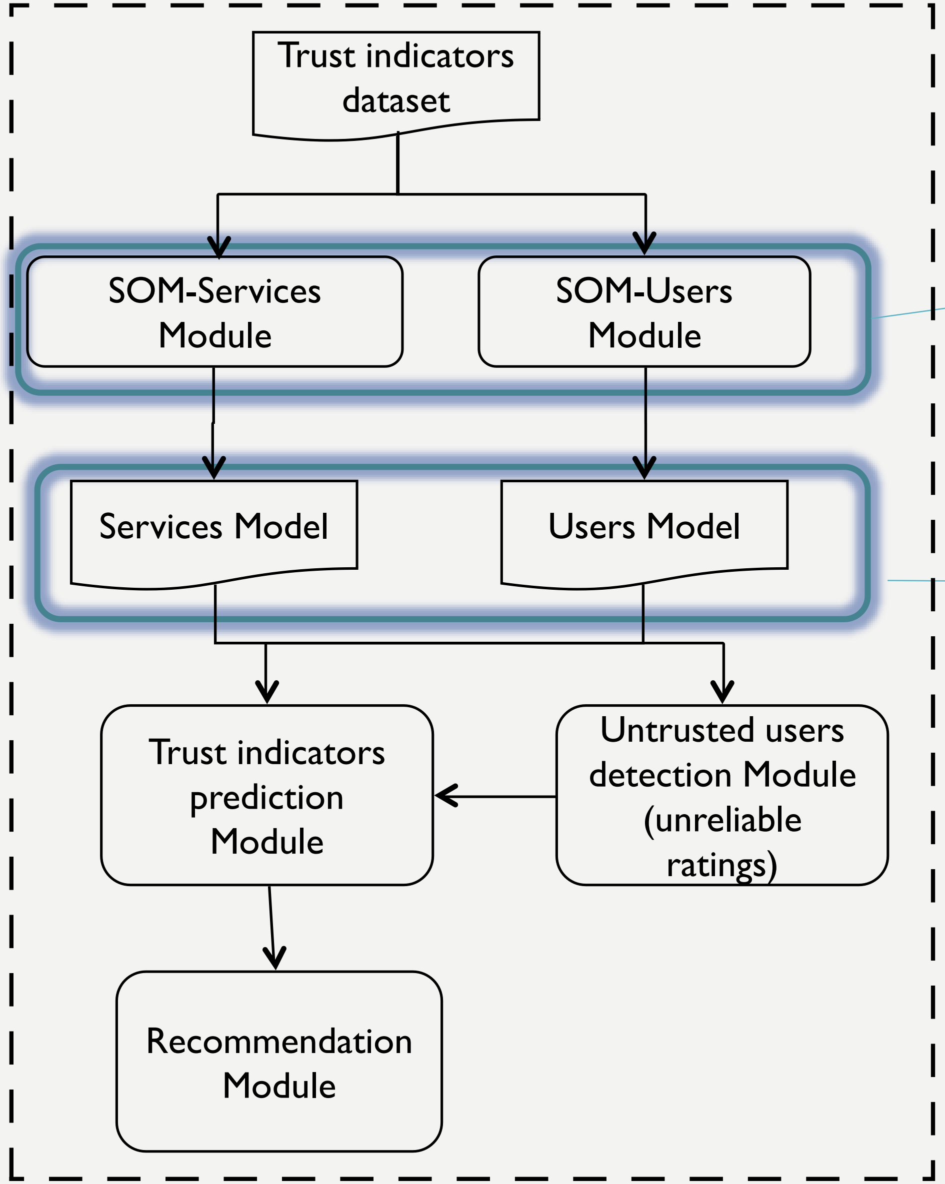
1.8, 1.2, 1.7, 2, 1.7, 1.8

K=3, α=0.6
DA=1.7

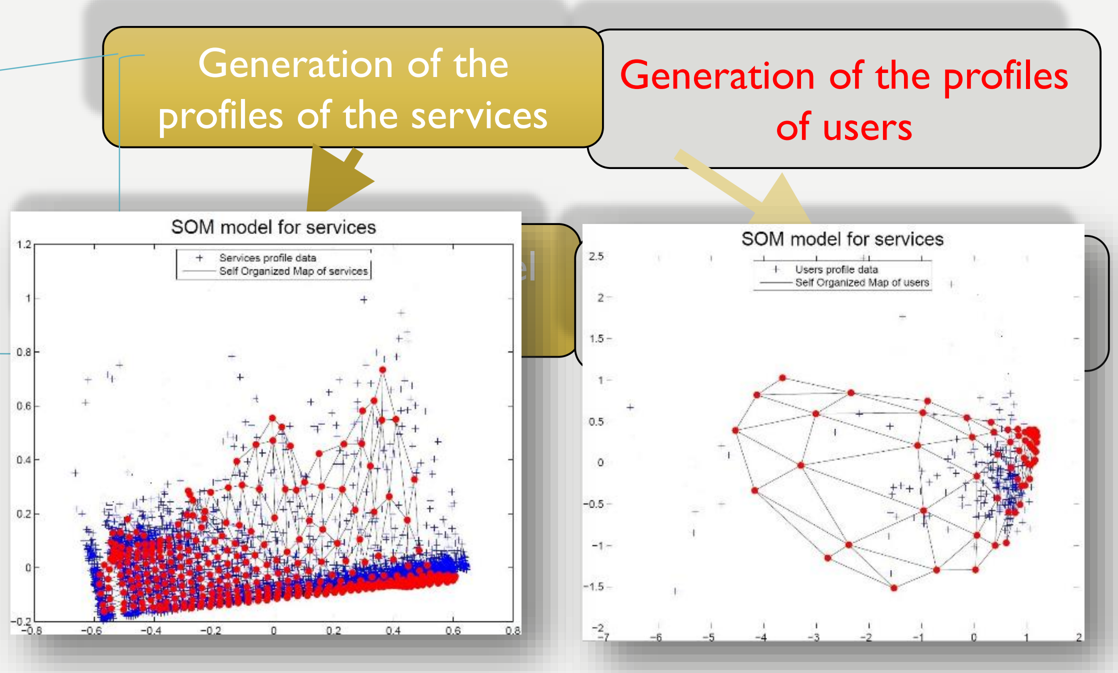
$$DAv_k^{\alpha}(N) = \frac{\sum_{i=1}^{|\xi|} n_i}{|\xi|}, n_i \in \xi$$

$$\xi = \bigcup_{i=1}^{j} C_N^i$$

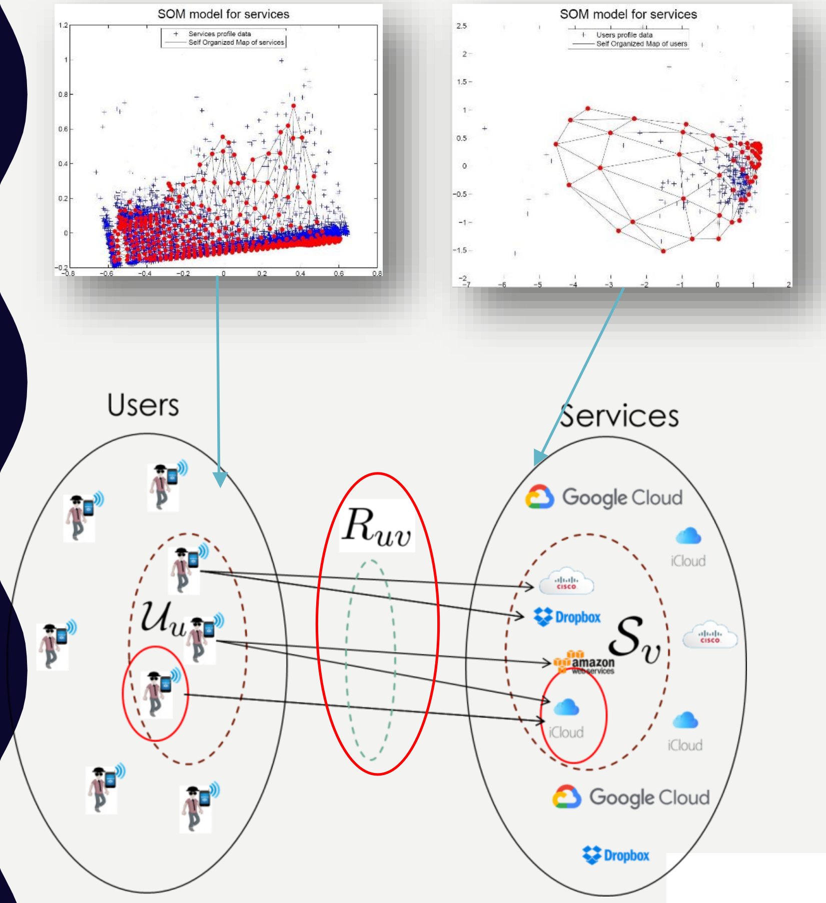
PHASE 1 : DESIGNING THE MODEL



Offline Process



PHASE 2 : RECOMMENDATION OF TRUST



On-line Process

1. Compute similarity using *BMN: Best Matching Neuron* (based on the distance)

$$\mathcal{S}_v = \{j \in S \mid BMN(j) = BMN(v)\}$$

2. Predict the reputation/trust of service s using Descriptive Average (using K-means)

$$r'_{uv} = DAv_k^{\alpha}(R_{uv}) = \frac{\sum_{r_{xy} \in \xi} r_{xy}}{|\xi|}$$

$$\text{Avec} \quad \xi = \bigcup_{k=1}^{j} C_{R_{uv}}^{k} \quad \text{et} \quad j = \underset{1 \leq m \leq K}{\text{argmin}} \sum_{m=1}^{j} \frac{|C_{R_{uv}}^{m}|}{|R_{uv}|} \geqslant \alpha \quad , \quad 0 < \alpha \leqslant 1.$$

PHASE 3 : UPDATE OF SOM MODELS

Algorithm 1 High-level description of the update algorithm **Offline Process**

Input: ΔT and ΔR
Output: Updated model
Initialisation: $t = 0$ time initialisation, $f = 0$ number of feedbacks, M numbers of neurons for services SOM-model

while $t < \Delta T$ and $f < \Delta R$ **do**
  if n new feedbacks are reported **then**
    $f = f + n$
  end if
  increment(t)
end while
// execute phase 1:
for all $s \in S$ **do**
  perform expressions (7) to (10)
end for
SOM update for services
for $1 < i < M$ **do**
  perform expressions (12) and (13)
end for
SOM update for users
return SOM_U and SOM_S

ΔR: window to update the users feedbacks
ΔT : time window to update the SOM models

IDENTIFICATION OF UNTRUSTED USERS

1. Invocation density: computes the density of calls done by a specific user *u* compared to others (*u* may have malicious behavior)
2. Coefficient of user Aberration : is the deviation of the feedbacks of user *u* compared with the feedbacks of other similar users
3. Credibility coefficient of the user *u.*

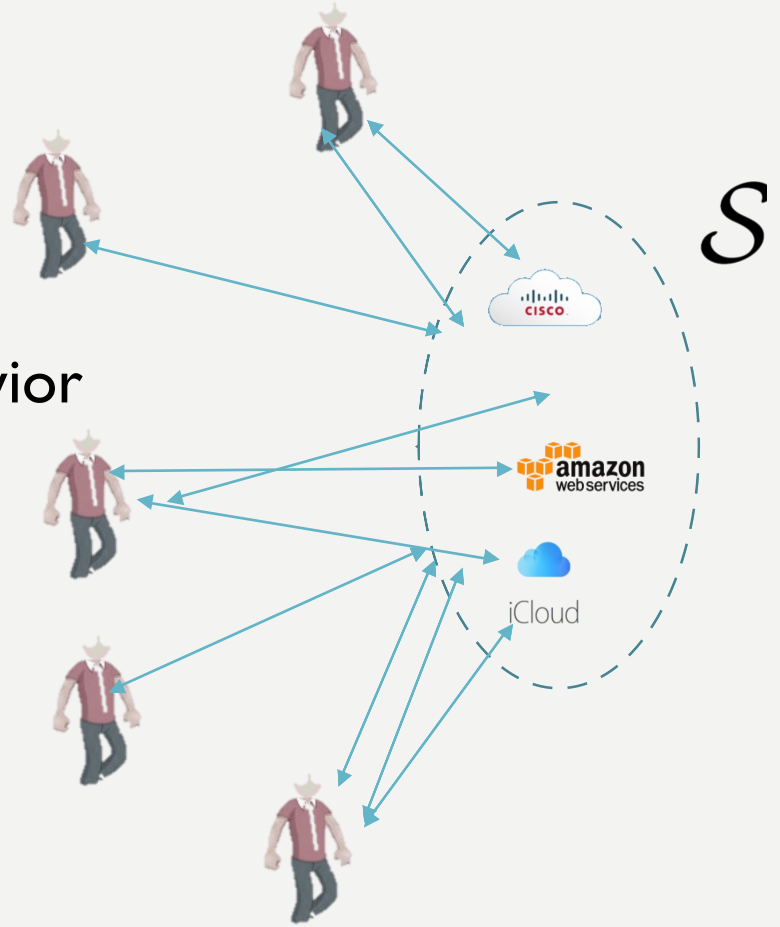
INVOCATION DENSITY OF THE USER

Number of times user *u* invokes the services of group S

$$D_{u,s} = \frac{n_{us} + 1}{DAv_k^{\alpha}(N_s) + 1}$$

≈ 1 means normal behavior

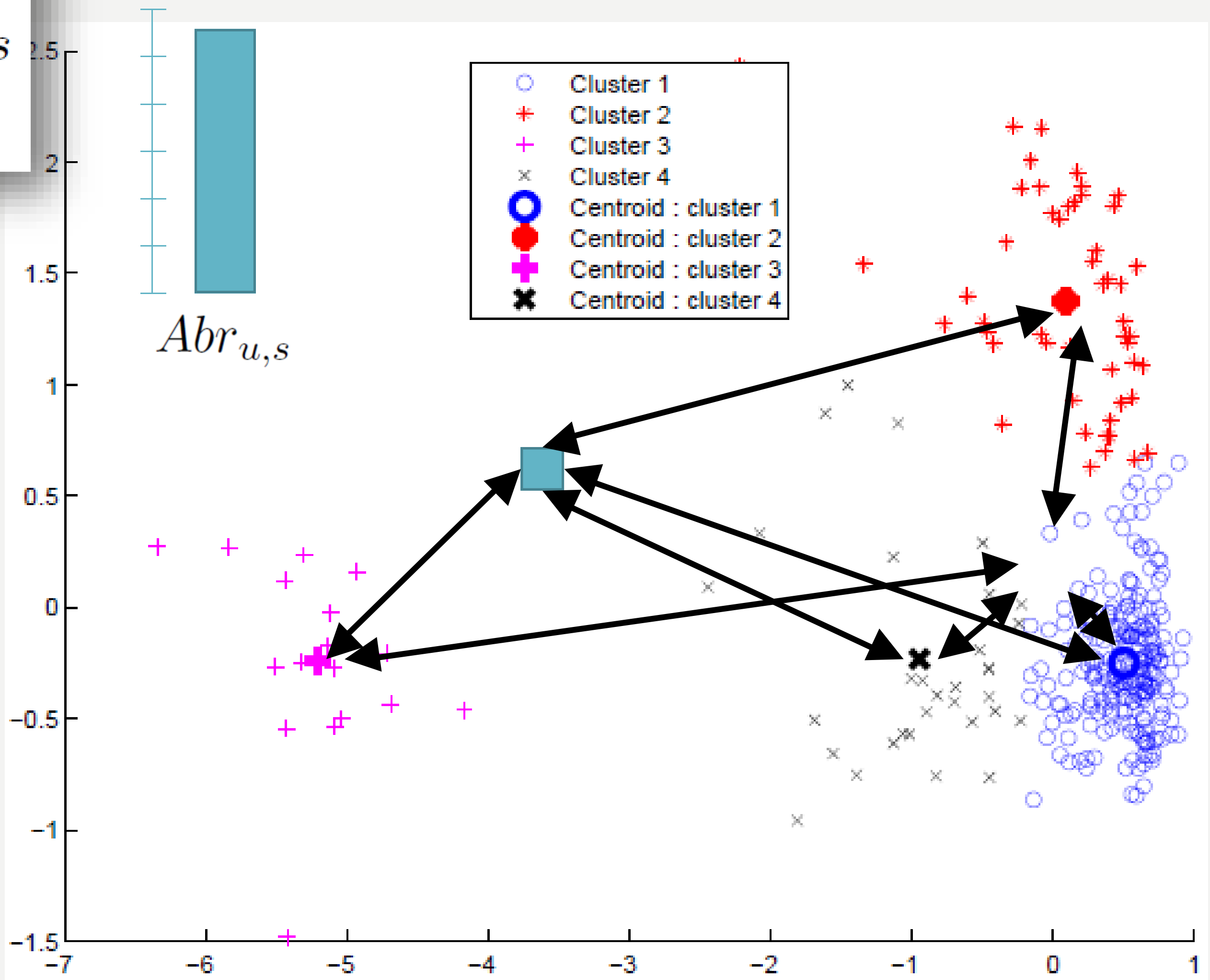
Distribution of the number of times that all the users invoke services of group S



COEFFICIENT OF USER ABERRATION

$$Abr_{u,s} = \sum_{i=1}^{k} |C_i| \times Dc_{i,s}$$

S_v



USER CREDIBILITY FACTOR

$$CrC_{u,s} = Abr_{u,s} \times e^{D_{u,s}}$$

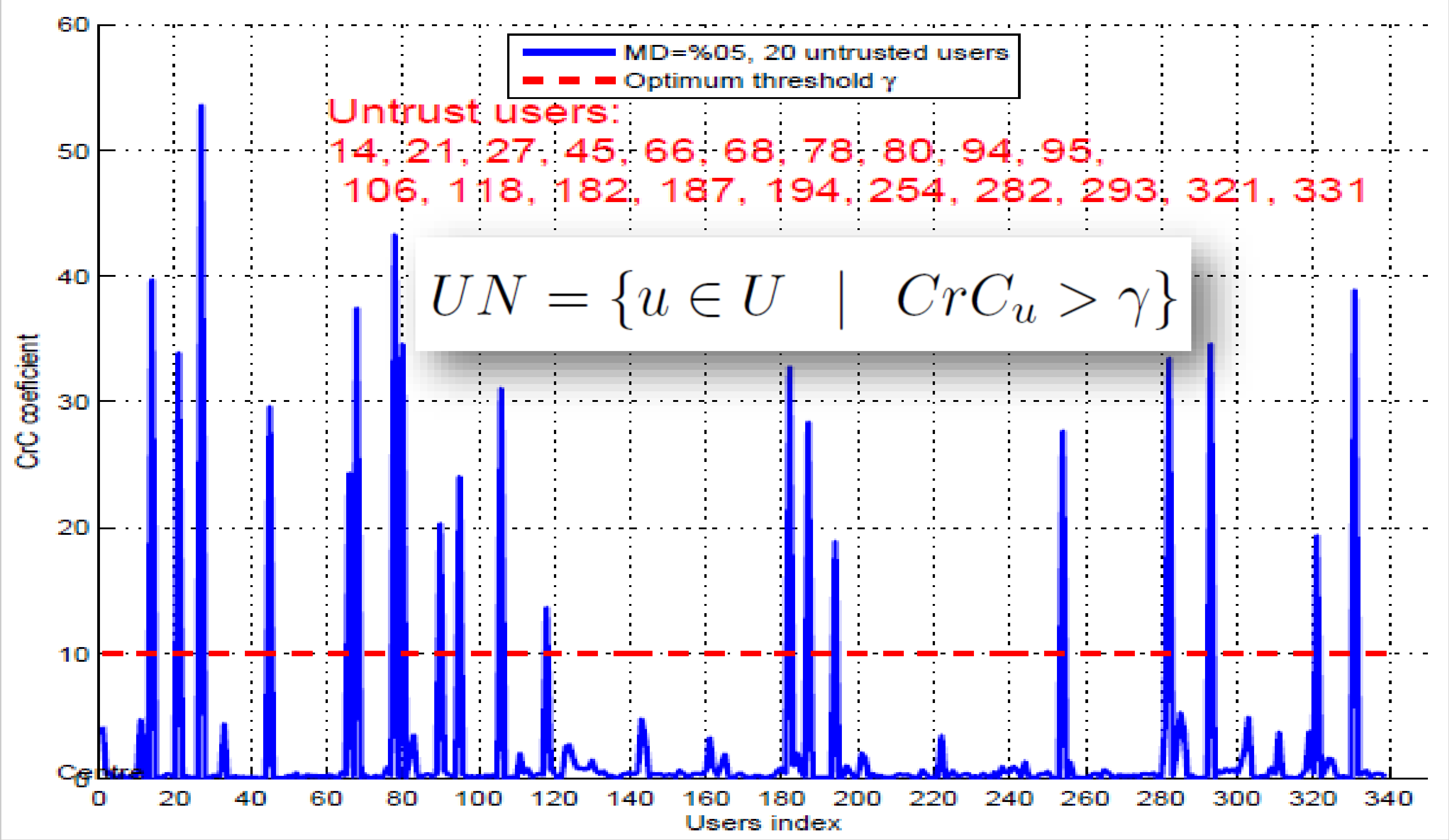
Invocation Density

Aberration coefficient

γ is a threshold

If CrC_u <= γ
user u is trusted

γ



$$UN = \{u \in U \mid CrC_u > \gamma\}$$

EXPERIMENTAL EVALUATION

WS-DREAM repository

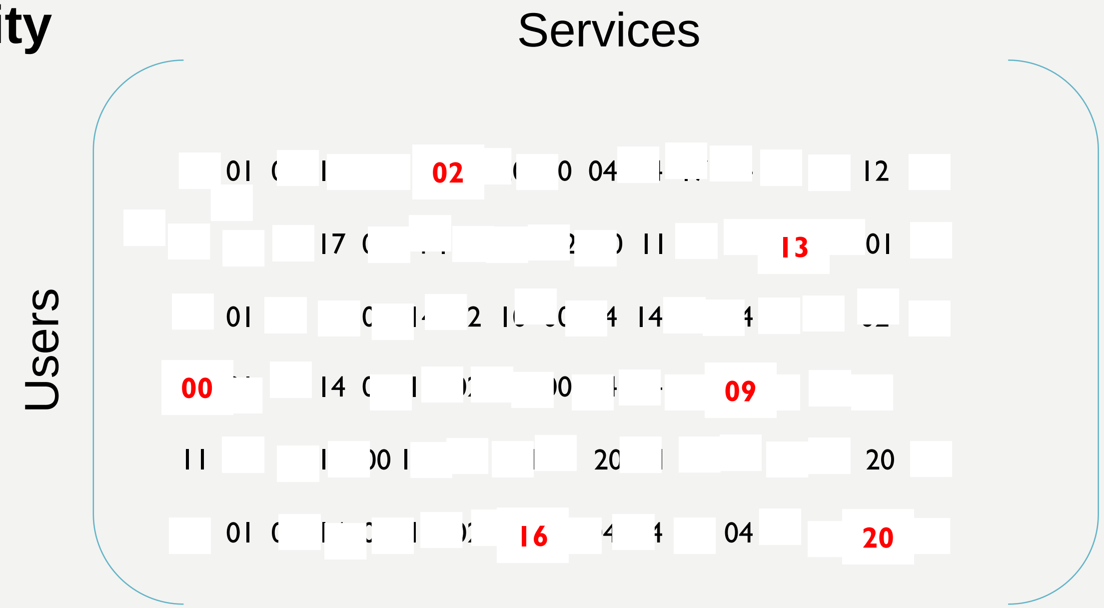
- ✓ 339 users (U) distributed over 30 countries
- ✓ 5825 Web services (S) distributed over 73 countries
- ✓ Matrix UxS has 1 974 675 evaluations
- ✓ Values of *Response Times* vary between 0 and 20 seconds.

Density of Matrix DM to simulate sparsity

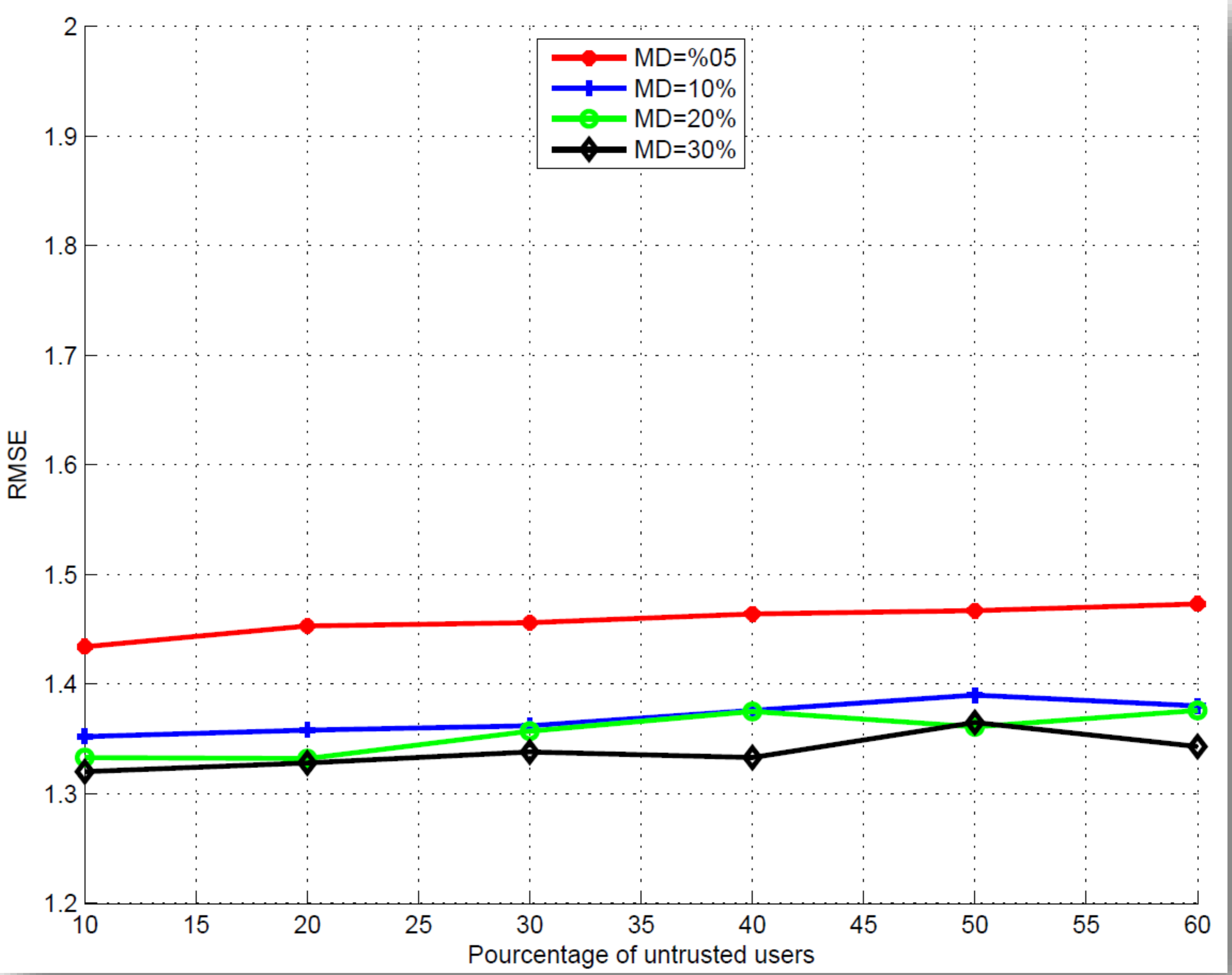
- ✓ MD = 30 % (70% missing data)
- ✓ MD = 10 % (90% missing data)
- ✓ MD = 05 % (95% missing data)

Untrusted users

- ✓ 2.9% users, 5.9% users
- ✓ 8.8% users, …, 17.7% users



ROBUSTNESS AGAINST UNTRUSTED USERS



RMSE: root-mean square error

Smaller is RMSE, greater is the prediction accuracy

The model is stable despite the increasing number of untrusted users.

COMPARATIVE ANALYSIS OF SOM-BTR

1. UPCC : similarity between users – PCC (Pearson Correlation Coefficient)
2. IPCC : similarity between services – based on cosine
3. WSRec : combines UPCC (User-based PCC) and IPCC (Item-based PCC)
4. GNMF : geographical neighbourhood - PCC - matrix factorization
5. TAP : IPCC - K-means for users
6. GURAP : geographical neighbourhood - reputation of users
7. SOM-BTR : similarity of services and users based on SOM model

COMPARATIVE ANALYSIS OF SOM-BTR

MD = (5%, 10%, 20% & 30%)
Untrusted users = 10 (over 339 users)

Methods	MD=5%	MD=10%	MD=20%	MD=30%
IPCC	2.305	2.183	2.149	2.079
UPCC	2.200	2.106	2.149	2.096
GNMF	2.275	2.078	1.703	1.655
WSRec	2.052	2.036	2.029	1.967
TAP	1.664	1.631	1.652	1.685
GURAP	1.494	1.433	1.332	1.284
SOM-BTR	**1.396**	**1.352**	**1.333**	**1.320**

CONCLUSION

- SOM-based approach for trust recommendation (first work that investigates SOM in this context)
- The selection is based on one trust indicator => needs extension to several trust indicators
  - Hydrid solution
- Adapt this approach to practical contexts (related to 5G or 6G for example).