Giancarlo Sperlì

Tutor: Antonio Picariello

XXX Cycle - III year presentation

Multimedia Social Networks

PHD Candidate

Graduation: MSc in Computer Engineering

Groups: MISLAB, PRIAMUS, CINI

Fellowship: MIUR research grant

Research Field: Big Data Analysis & Social Network

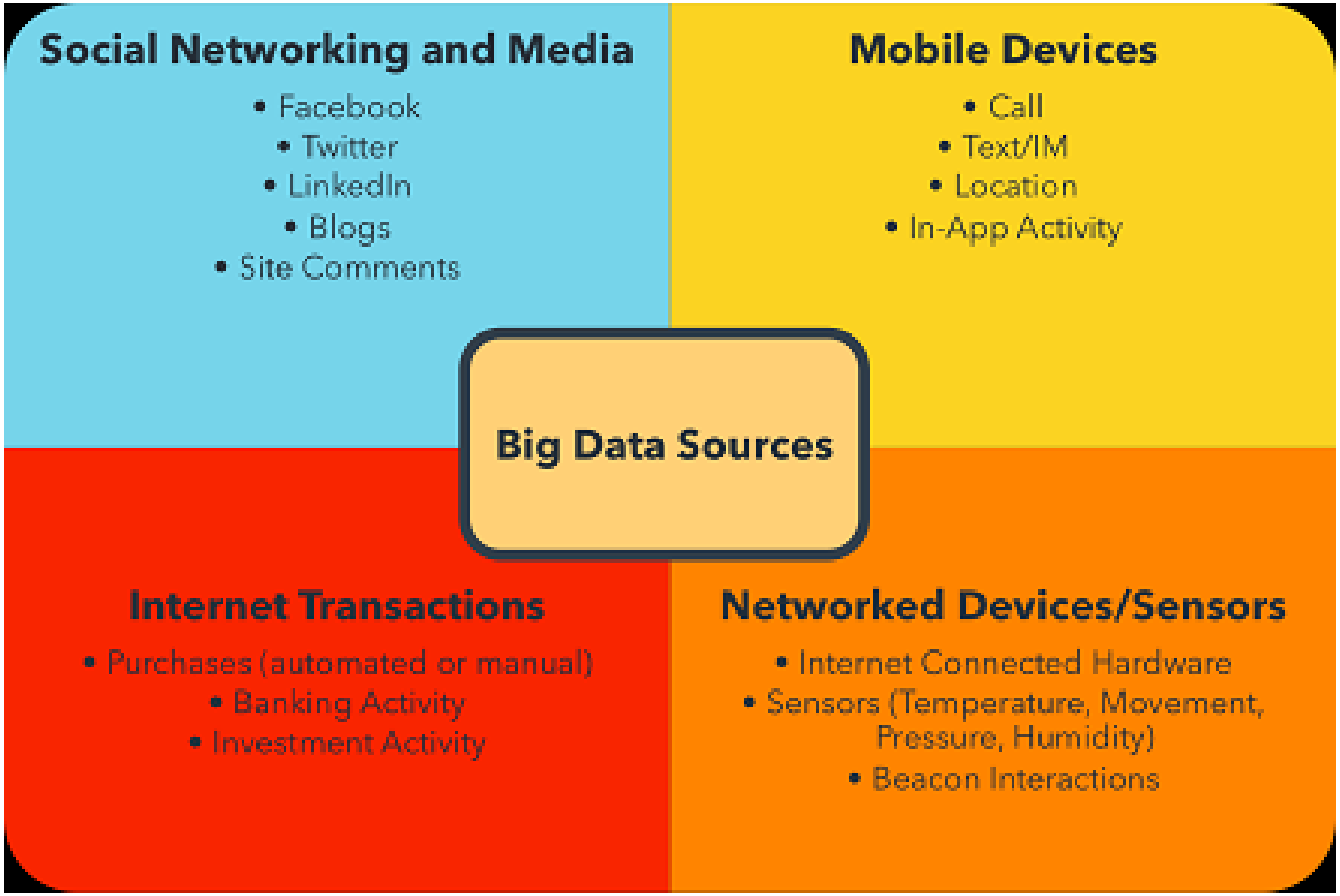
Study Activity:

	Credits year 1								Credits year 2								Credits year 3									
	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Total	Check
Modules	20	0,0	5,0	3,0	7,0	0,0	0,0	**15,0**	15	0,0	6,0	0,0	0,0	0,0	0,0	**6,0**	21	0,0	9,0	1,6	0,0	0,0	0,0	**10,6**	**31,60**	30-70
Seminars	8	1,0	1,5	3,8	5,2	0,0	0,0	**11,5**	6	6,0	0,0	0,0	0,0	0,0	0,0	**6,0**	12	0,0	0,0	0,0	0,0	6,0	0,0	**6,0**	**23,50**	10-30
Research	32	9,0	3,5	3,2	3,0	10,0	10,0	**38,7**	39	4,0	4,0	10,0	10,0	10,0	10,0	**48,0**	30	10,0	1,0	8,4	10,0	4,0	10,0	**43,4**	**130,10**	80-140
	60	10,0	10,0	10,0	15,2	10,0	10,0	**65,2**	60	10,0	10,0	10,0	10,0	10,0	10,0	**60,0**	63	10,0	10,0	10,0	10,0	10,0	10,0	**60,0**	**185,20**	180

Activity abroad: seven months (from May to December 2016) at the University of Maryland (UMD) under the supervision of the Prof. V.S. Subrahmanian (Department of Computer Science and Engineering).

Thesis: Multimedia Social Networks

Introduction (1/3)



Introduction (2/3)

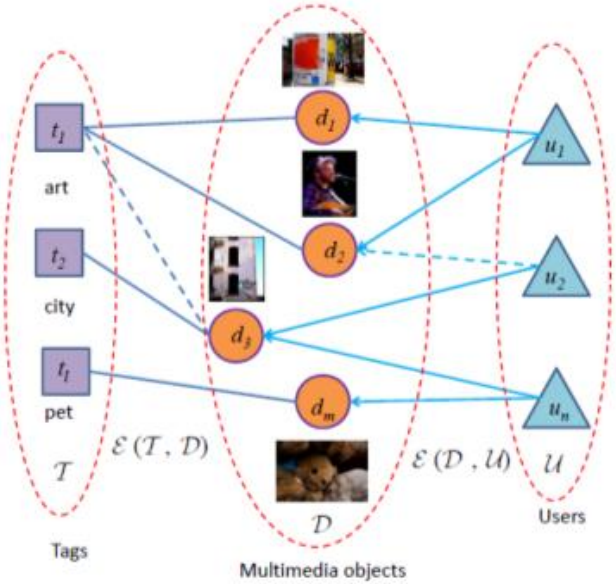
- …from a technological point of view
  - OSNs are enabled as Internet applications with a set of functionalities:
    - information sharing capabilities;
    - user generated content management;
    - support by means of several tools to different ways of communication and collaboration among users.

Introduction (3/3)

- …from a sociological perspective
  - OSNs are social structures constituted by a set actors (individuals or organizations), sets of dyadic ties and other social interactions, often instantiated through the shared information.

OSN modelling (1/2)

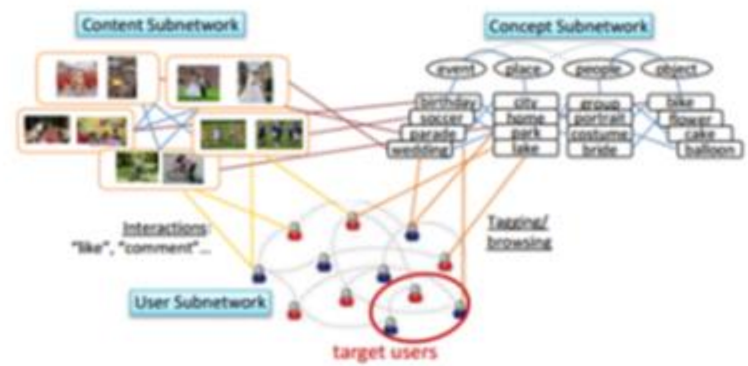
- A social network is represented as a directed graph, with each person (customer) as a node (Kempe et al.)

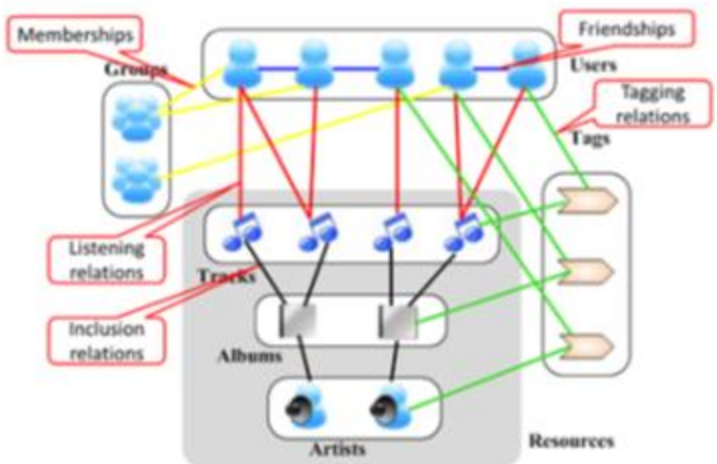


- A social network is represented as a tri-partite graph, composed by user, social media object and Tag (Qui et al.)

OSN modelling (2/2)

A social network is represented as a graph in which only the relationships established between users, tags and multimedia objects are modelled through hyperarc (Liu et al.)

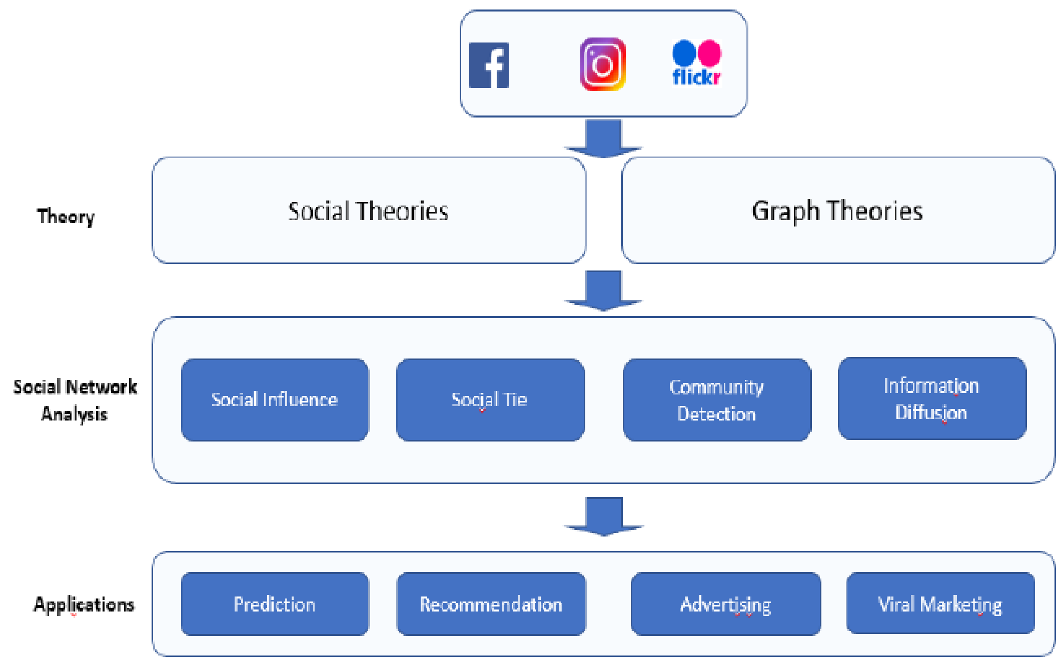




A social network is represented a unified hypergraph to model multi-type objects ant the high-order relations.(Bu et al.)

Proposed MSN Definition

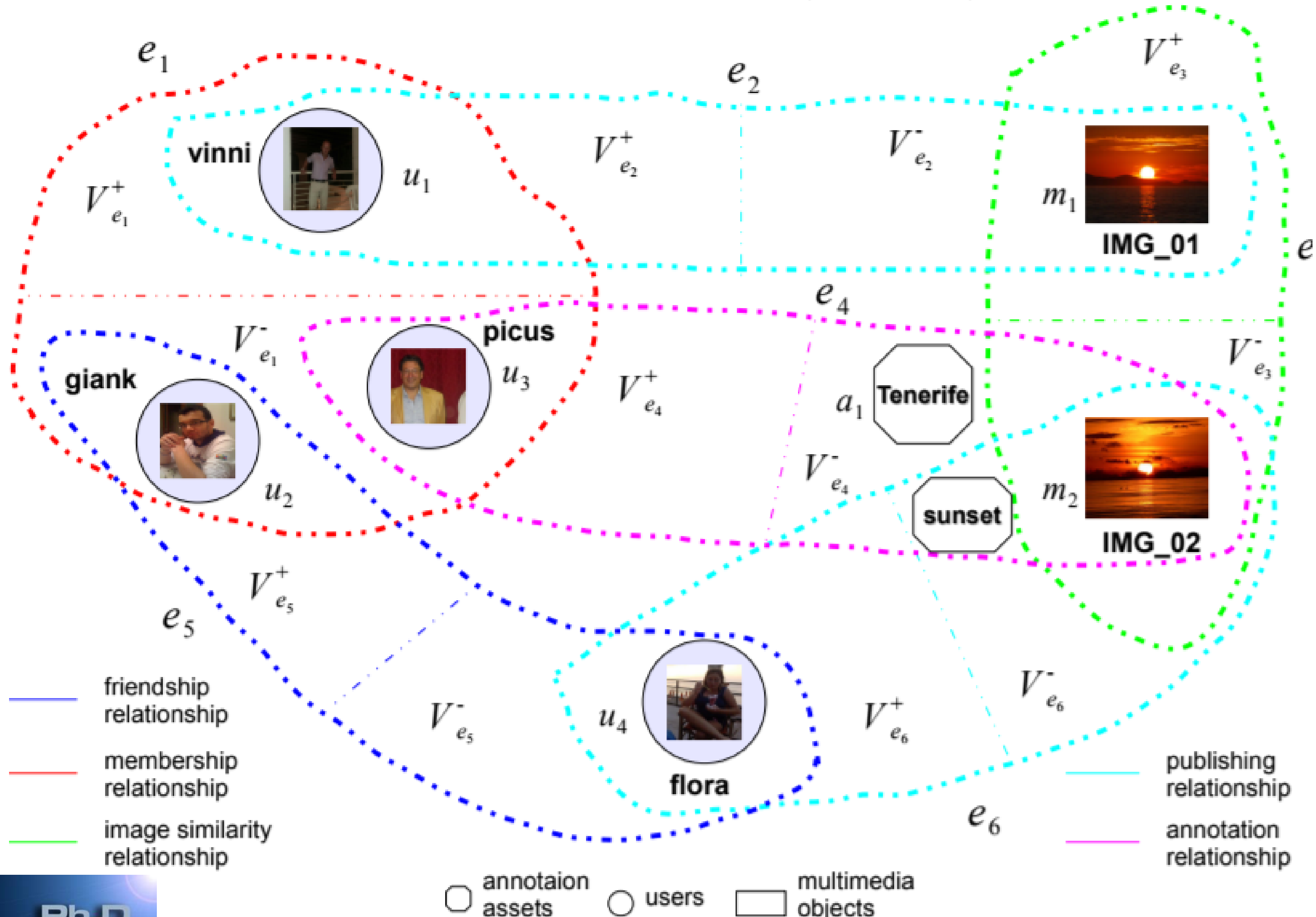
“integrated social media networks that combine the information on users, belonging to one or more social communities, together with all the multimedia contents that can be generated and used within the related environments”.



MSN Model (1/2)

- In our vision, a MSN is basically composed by two different entities:
  - Users - that correspond to the set of persons and organizations with related attributes belonging to one or more social communities.
  - Multimedia Objects - the set of multimedia resources that can be shared within a MSN community. High level (metadata) and low level information (features) can be properly and jointly used in our model.

MSN Model (2/2)



e_1
e_2
$V^+_{e_3}$
vinni
u_1
$V^+_{e_1}$
$V^+_{e_2}$
$V^-_{e_2}$
m_1
IMG_01
e
e_4
$V^-_{e_1}$
picus
u_3
giank
$V^+_{e_4}$
a_1
Tenerife
$V^-_{e_3}$
u_2
$V^-_{e_4}$
sunset
m_2
IMG_02
$V^+_{e_5}$
e_5
u_4
flora
$V^-_{e_5}$
$V^+_{e_6}$
$V^-_{e_6}$
e_6
friendship relationship
membership relationship
image similarity relationship
publishing relationship
annotation relationship
annotaion assets
users
multimedia objects

Influence analysis (1/2)

- Traditional communication theories state that a minority of users, called influentials, excel in persuading others.

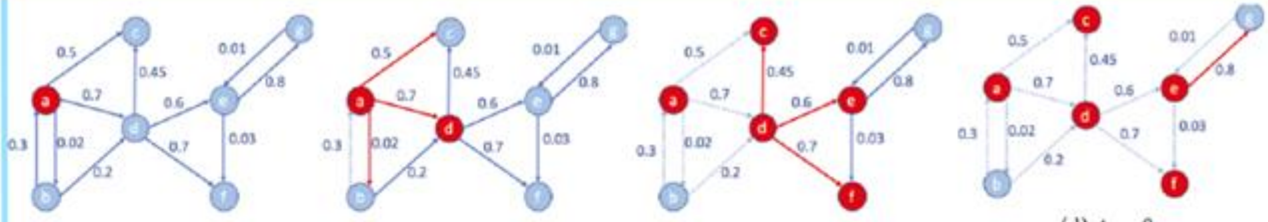
- An influence analysis problem can be faced using two steps:
  1. a diffusion model is defined with the aim of describing the influence spread in the network;
  2. a maximization algorithm is exploited to identify the set of nodes such that their activations maximize the diffusion or the propagation of influence.

- The selection of the most influence nodes is an optimization problem that has been proven by Kempe et al. to be NP-Hard.

$$S^* = argmax_{|S| \geq k} \sigma(S)$$

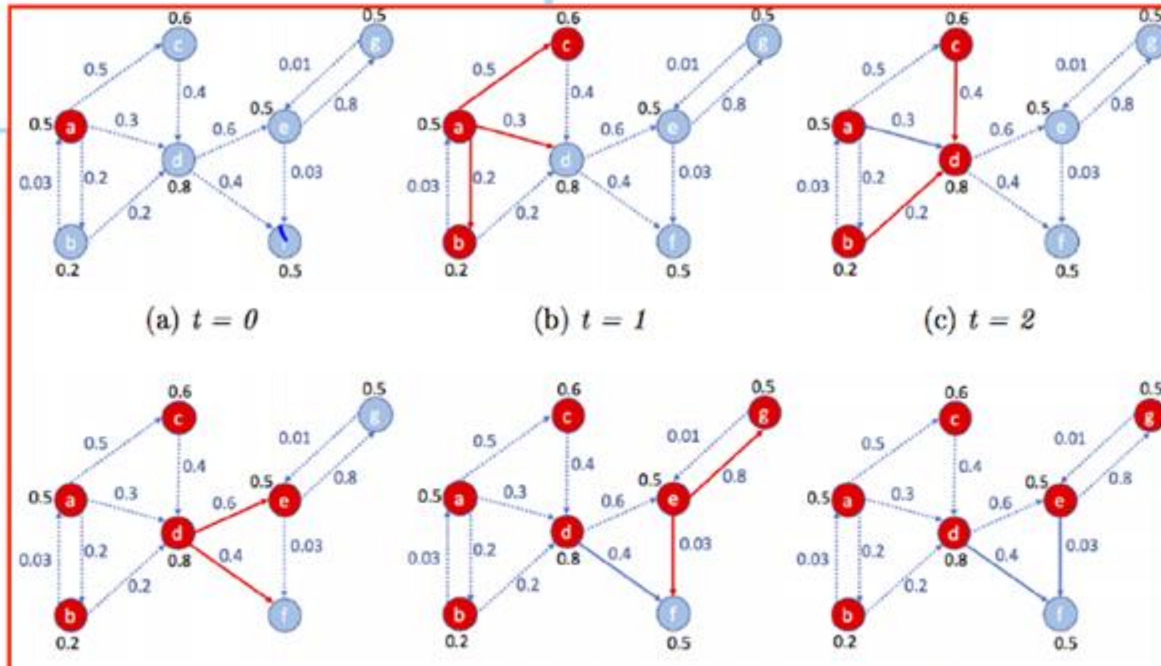
- Greedy strategies exploiting a non-negative, submodular and monotone influence function can obtain a solution that is no worse than (1-1/e) of the optimal one.

Influence analysis (2/2)



(a) $t = 0$ (b) $t = 1$ (c) $t = 2$ (d) $t = 3$

Independent Cascade (IC): on each edge (u,v) there is a probability $p_{u,v}$ to become active. Every node u has one attempt to activate a node v



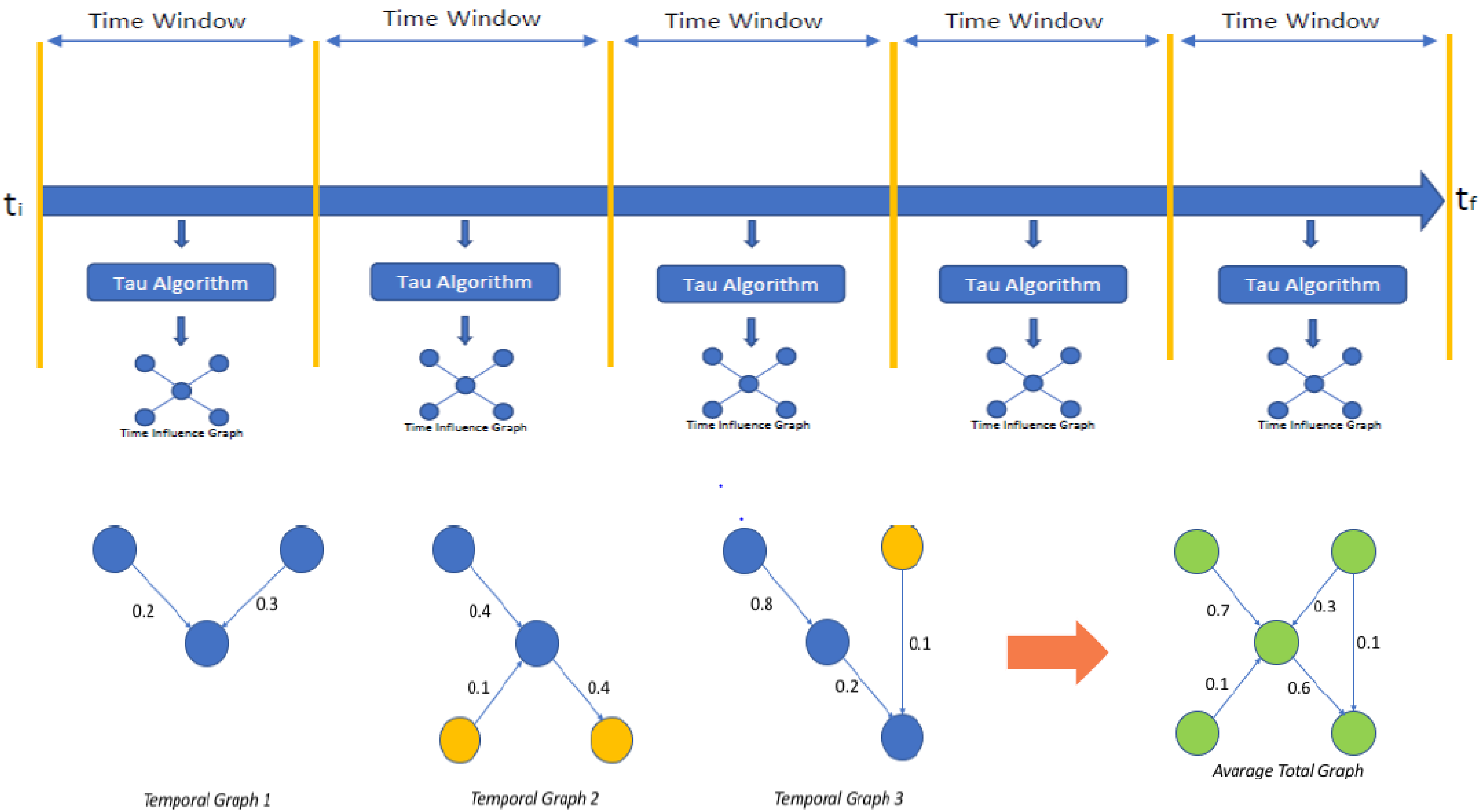
(a) $t = 0$ (b) $t = 1$ (c) $t = 2$

(d) $t = 3$ (e) $t = 4$ (f) $t = 5$

Linear Threshold (LT): each node has a threshold w_v and it becomes active only if the sum of the probabilities of its neighbors is greater than the threshold:

$$\sum_{u \in N(v) \cap S(t_i)} p_{u,v} \geq w_v$$

My Proposal: Influence Model (1/7)



My Proposal: Influence Model (2/7)

Basic assumptions

We assume the existence of a finite set of Actions (A) representing all the possible interactions among the set of Users (U) and the set of Objects (O) in one ore more online social networks, which can be properly captured during user browsing sessions exploiting log information.

Log tuple

A *log tuple* can be defined by the information $l = (a, u, o, \lambda_1, \cdots, \lambda_k)$, where $a \in A$, $u \in U$, $o \in O$ and $\lambda_1, \cdots, \lambda_k$, are particular attributes (e.g., timestamp, type of reaction, text and tags of a comment, etc.) used to describe an action.

My Proposal: Influence Model (3/7)

Reaction Operator

The *Reaction Operator* $reac^{\Delta t}(a_1, a_2)$ between two actions a_1 of user u_i and a_2 of user u_j – and both the actions are performed on the same object o (or on similar objects[a]) – returns the probability that a_2 occurs after a_1 within the interval Δt.

[a] The evaluation of such condition needs the defiinition of a similarity function between two objects.

My Proposal: Influence Model (4/7)

Influence Operator

Let $u_1, u_2 \in U$ be respectively two users, we say that $u_1 \underset{\tau}{\rightarrow} u_2$, if each action $a_{u_1} \in A_{u_1}(t)$ of user u_1 at time t determines an action $a_{u_2} \in A_{u_2}(t, \Delta t)$ of user u_2 in the interval $]t, \ldots, \Delta t]$ within a log L:

$$u_1 \underset{\tau}{\rightarrow} u_2 \iff \forall t_i \in T \exists a_1 \in A_{u_1}(t_i), a_2 \in A_{u_2}(t_i, \Delta t) \in L : \\ reac^{\Delta t}(a_1, a_2) \geq \tau$$

$T = \{t_1, t_2, \ldots, t_m\}$ being a sequence of temporal instants such that $t_1 < t_2 < \ldots t_m$ and $\tau \in [0, 1]$ a probability value.

My Proposal: Influence Model (5/7)

Influence Diffusion

Let u_1,u_2 and u_3 three users and L a given log,

$$\begin{cases} u_1 \underset{\tau^1}{\rightarrow} u_2 \\ u_2 \underset{\tau^2}{\rightarrow} u_3 \end{cases} \Rightarrow u_1 \underset{\tau^3}{\rightarrow} u_3 \qquad (1)$$

$\tau^3 \leq min(\tau^1, \tau^2)$.

My Proposal: Influence Model (5/7)

- **Definitions:**
  - Reactivity (p_1) of u w.r.t. v: $\dfrac{n_{u,v}^{R}}{\sum_{i\in N} n_{u,v_i}^{R}}$;
  - Shareability (p_2) of u w.r.t. v: $\dfrac{n_{u,v}^{R}}{n_{v}^{A}}$;

- **Proposed Influence Operators:**
  1) $\alpha * p_1 + \beta * p_2 = \alpha * \dfrac{n_{u,v}^{R}}{\sum_{i\in N} n_{u,v_i}^{R}} + \beta * \dfrac{n_{u,v}^{R}}{n_{v}^{A}}$
  2) $p_1 * p_2 = \dfrac{n_{u,v}^{R}}{\sum_{i\in N} n_{u,v_i}^{R}} * \dfrac{n_{u,v}^{R}}{n_{v}^{A}}$
  3) $Norm(p_1 + p_2) = Norm(\dfrac{n_{u,v}^{R}}{\sum_{i\in N} n_{u,v_i}^{R}} + \dfrac{n_{u,v}^{R}}{n_{v}^{A}})$

My Proposal: Influence Model (7/7)

Influence Graph

An *Influence Graph* is a labeled graph $G = (V, E, \tau)$ where:

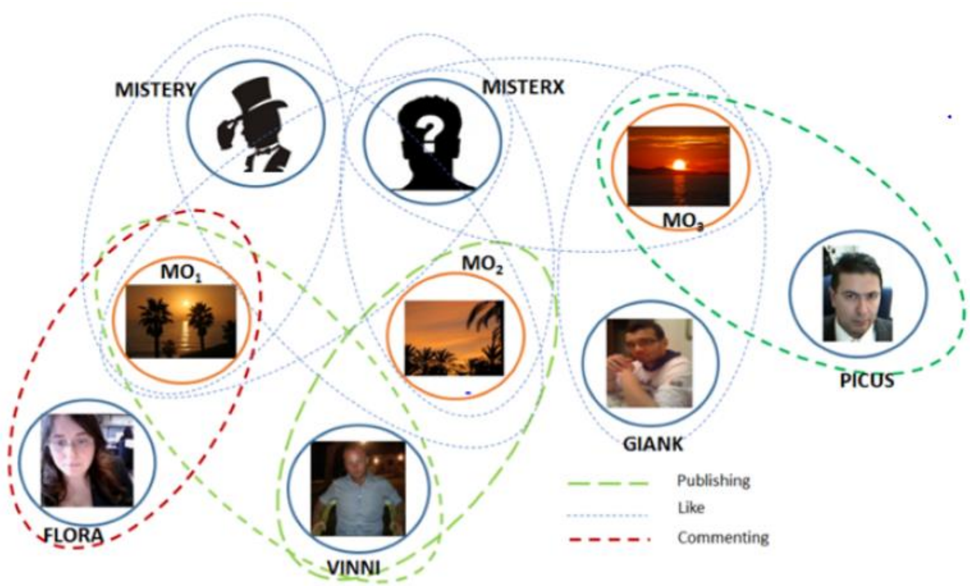
- V is the set of nodes such that each $v \in V$ corresponds to a user $u \in U$;
- $E \subseteq V \times V$ is the set of edge (with no self-loops);
- $\tau : V \times V \rightarrow [0, 1]$ is a function that assigns to each edge $e = (v_i, v_j)$ a label, representing the probability that user u_i can influece user u_j.

Stochastic Algorithm

User-Multimedia Object Matrix

Let U, M, $\mathbf{UP}$,$\mathbf{PM}$ be respectively the sets of users, multimedia objects and user-relevant path and relevant path-multimedia objects matrixes, we define the *User-Multimedia Object Matrix* as:

$$\mathbf{UM} = \{um_{u_i m_j}\} = \sum_{u_i \in U} (\prod_{p_k \in P} up_{u_i p_k} * pm_{p_k m_j})$$

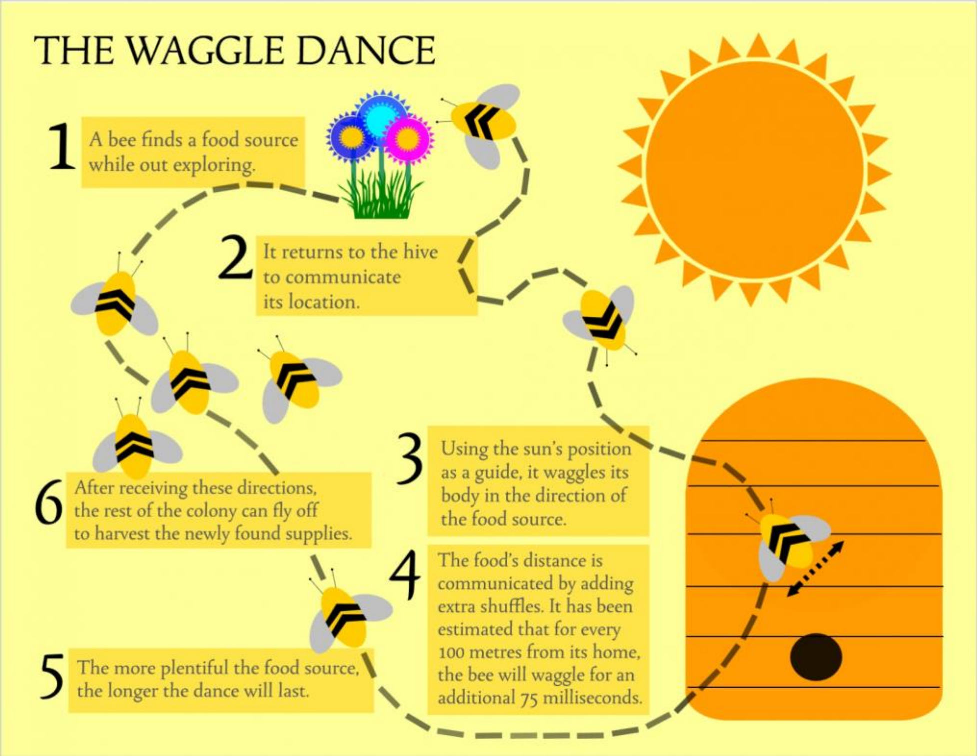


$$\begin{bmatrix} 3.0 & 2.0 & 0 \\ 0 & 0 & 2.0 \\ 0 & 0 & 0.33 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Bio-inspired Algorithm

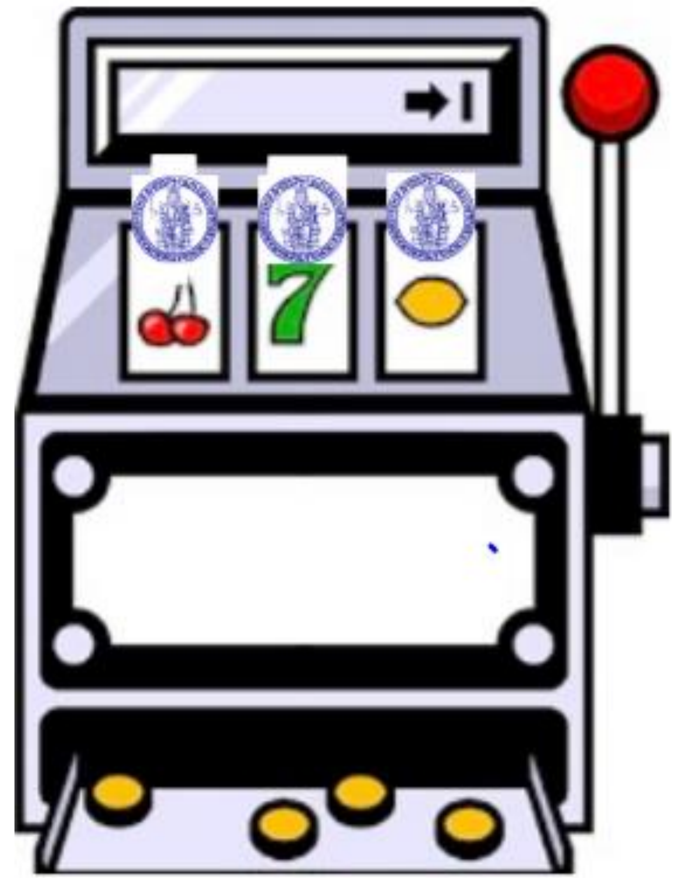
The ABC algorithm works in according to two fundamental steps:

1. an initial user ranking (based on their centrality in the Influence Graph) is performed (to determine the most suitable employer bees to lead the food search campaign together with a set of scouts represented by their neighbors in the network);
2. a top-k selection of the most influential users within the initial set (represented by the employer or scout bees that are effective leaders on the base of their waggle dance) is carried out in an iterative manner.

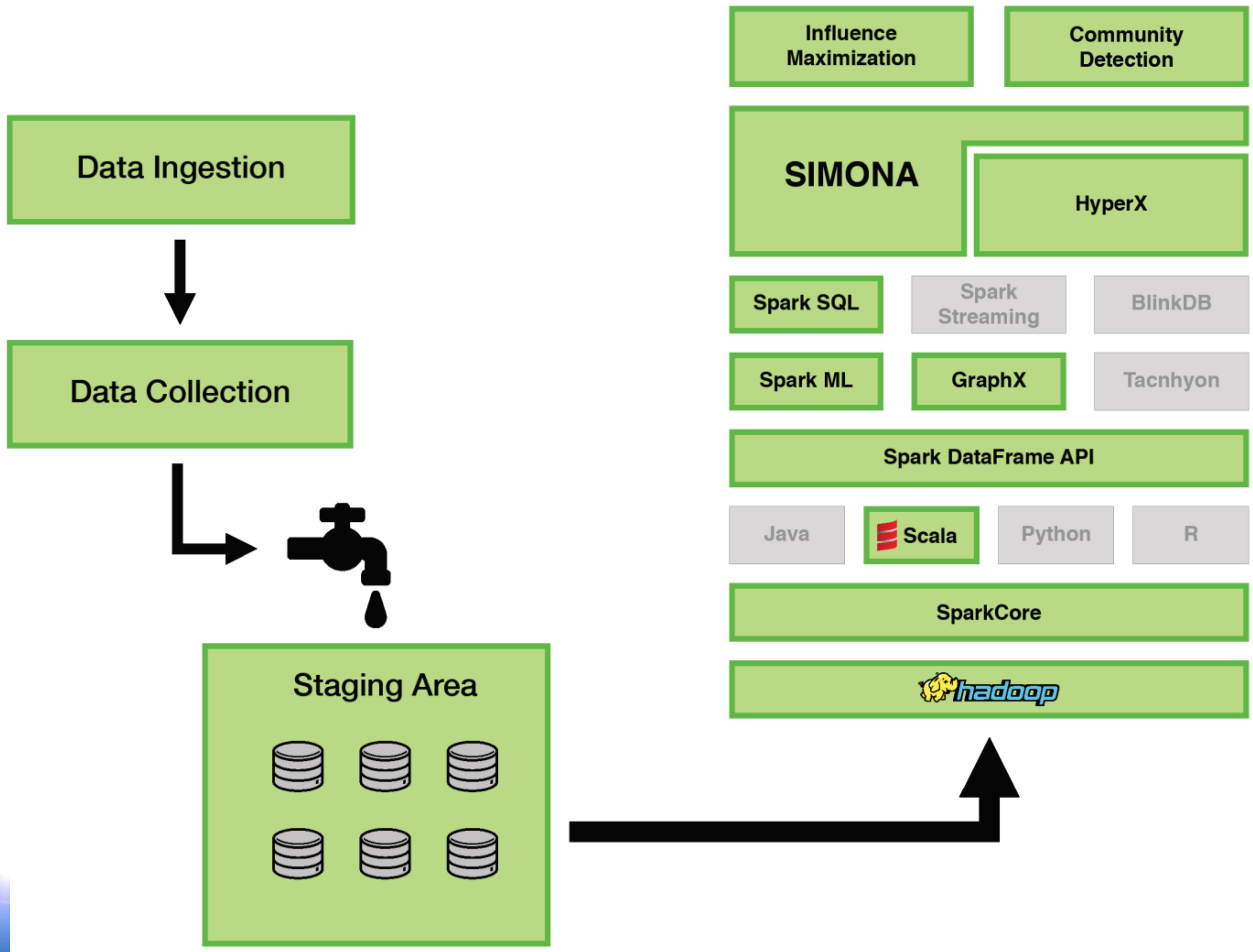


Game Theory Approach

- Our contributions:
  - Adopt the Game Theory and formulate a Combinatorial Multiarmed Bandit Problem
  - Aim to maximize the spread and to learn the influence probabilities simultaneously
    - Leads to classic exploration vs. exploitation trade-off
  - Consider node-level feedback: you only need to know who was influenced
- CMAB algorithm:
  - Each of m arms has reward distribution with unknown mean μ
  - Based on MAB, but the arms can trigger others arms
  - In each round t a subset of arms A is chosen and reward is function of these arms
  - Update and improve our knowledge



Proposed Architecture



Experimental Protocol

Dataset

YFCC100M	Pubblishing	Favorite	Comment
Complete	99.000.000	*	*
Analyzed	14.663.918	3.101.814	2.827.439

(*) Social features change on a day to day basis

Influence estimation methods:

1. Trivalency model
2. Weighted cascade
3. Different influence operators

Hardware details:

- Microsoft Azure with 2 compute optimized instances (2 x 8 CPU and 16 GB RAM)

Evaluation without Similarity

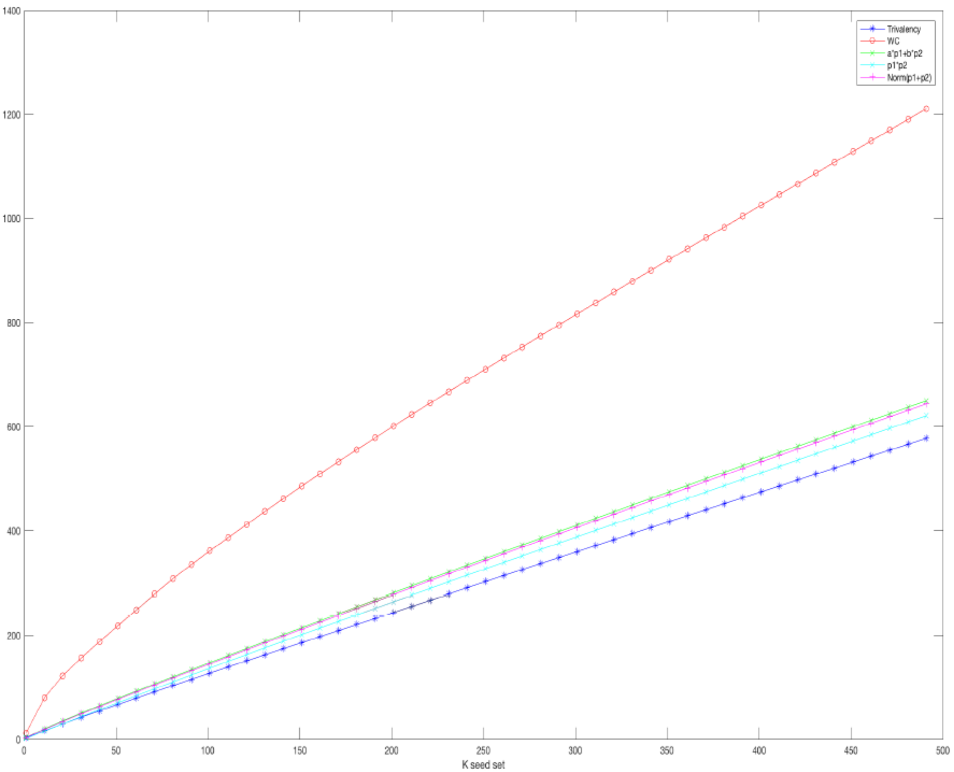


Figure a: Influence Spread with TIM on P1

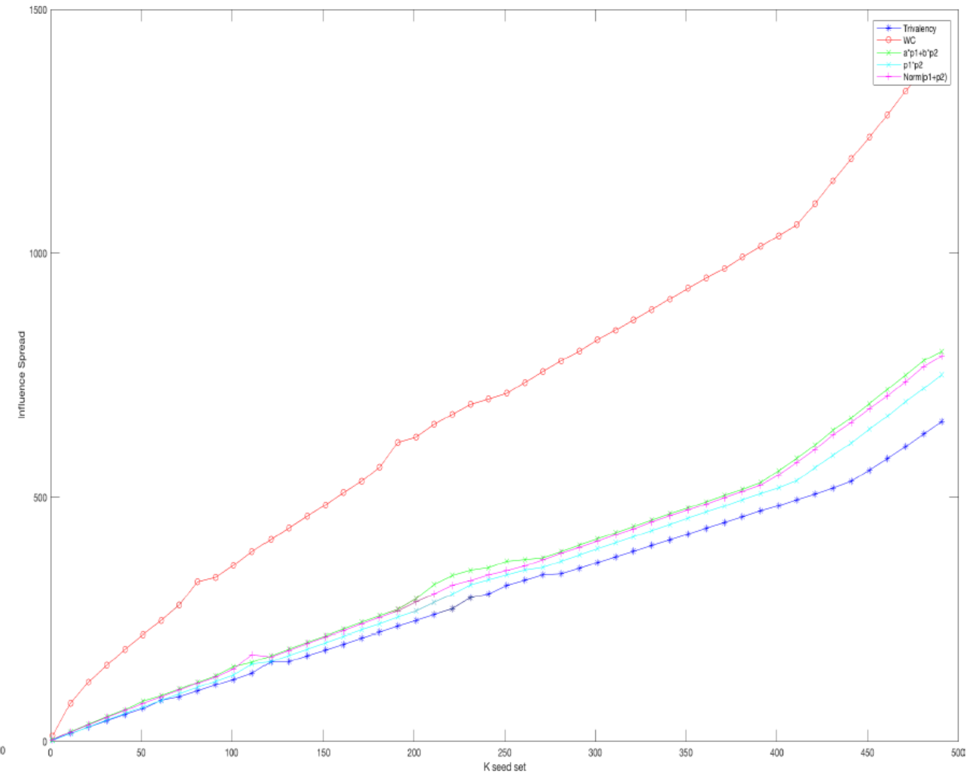


Figure b: Influence Spread with IMM on P1

Evaluation with Similarity

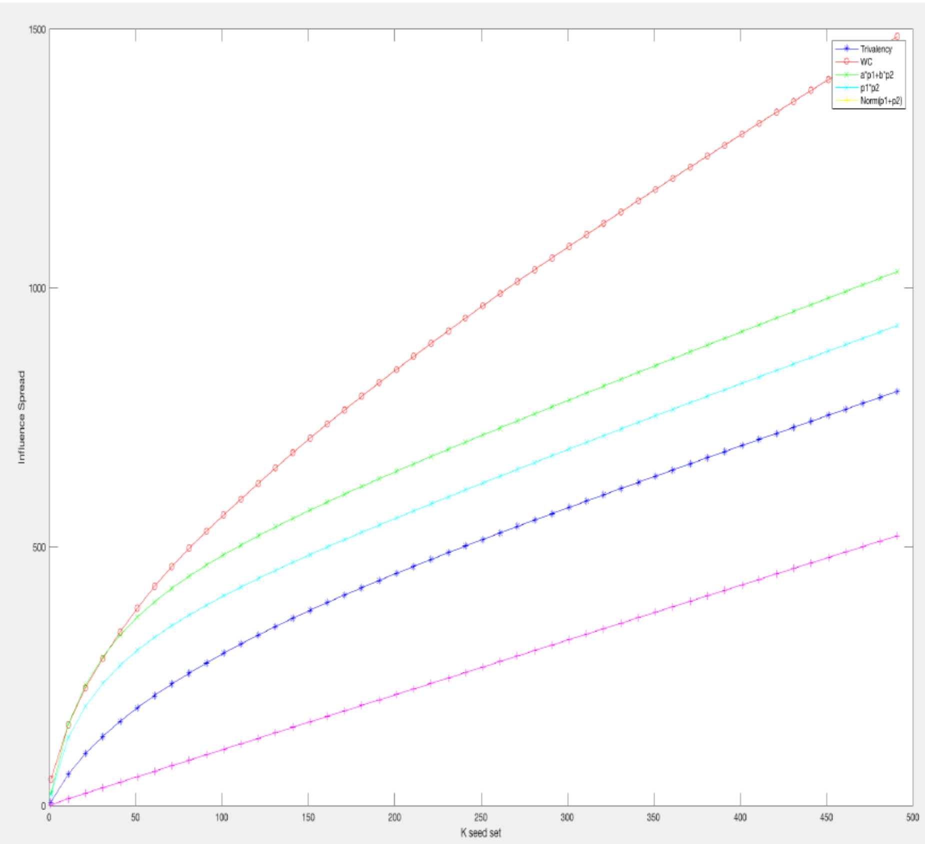


Figure (a):Influence Spread with TIM on P2

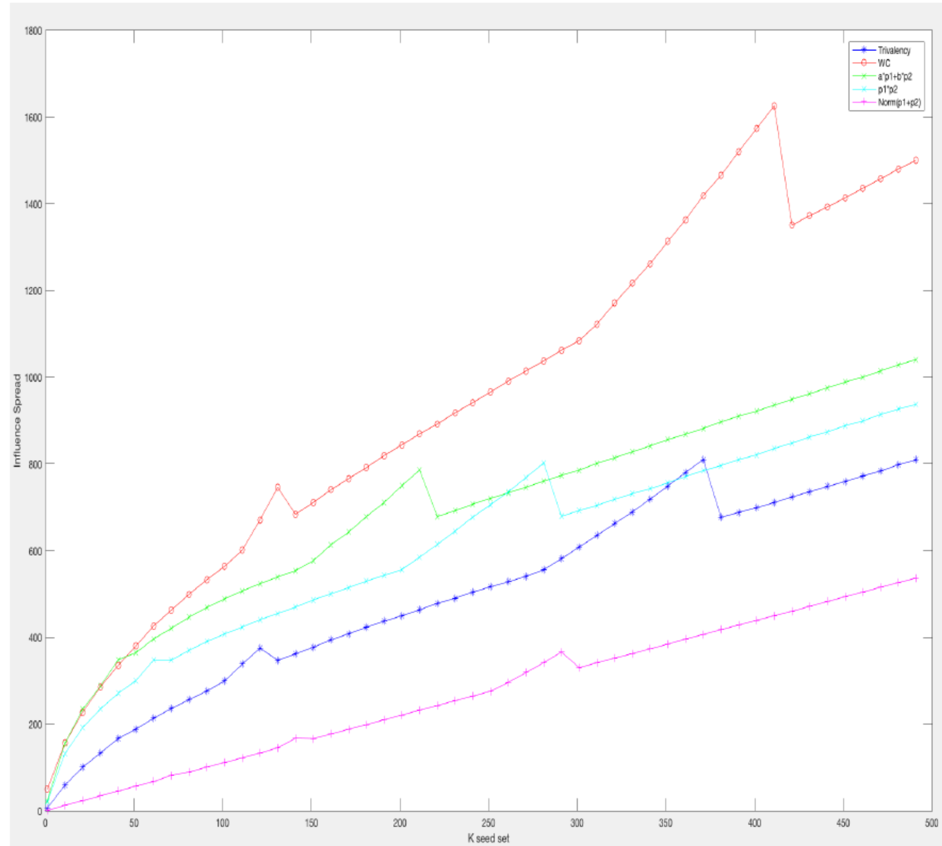


Figure (a):Influence Spread with IMM on P2

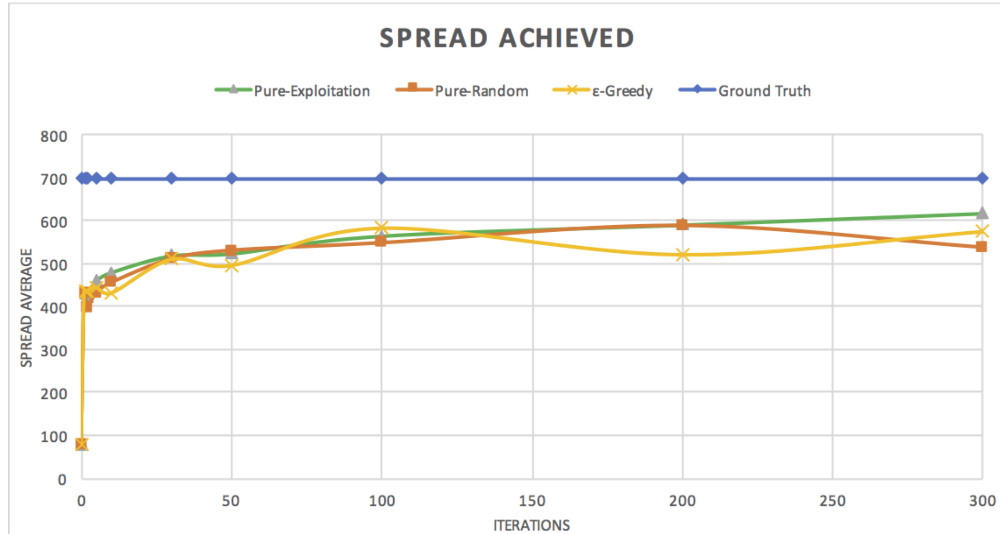
Evaluation

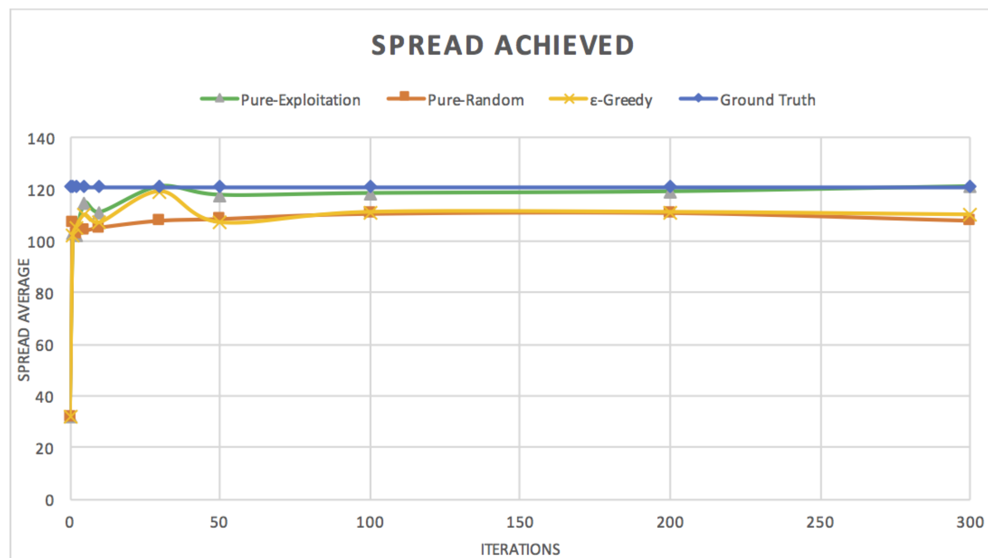
Reward Maximization Algorithms:

1. **Pure Exploitation:** performs exploitation in every round. TIM+ ($\varepsilon = 0.2$) as an oracle

2. **Pure Exploration:** performs exploration in every round.

3. **ε-Greedy:** exploration with probability ϵ) and exploitation with probability $1 - \epsilon$)

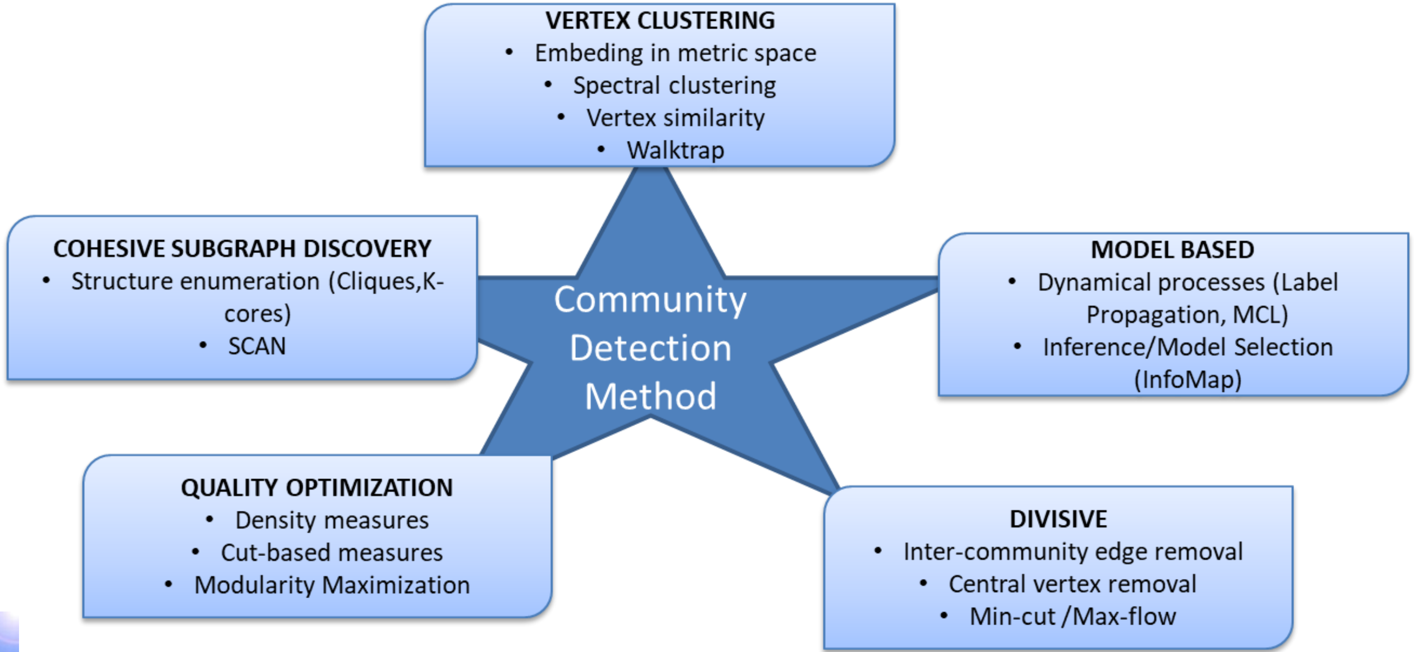




Community Detection

Community detection definitions:

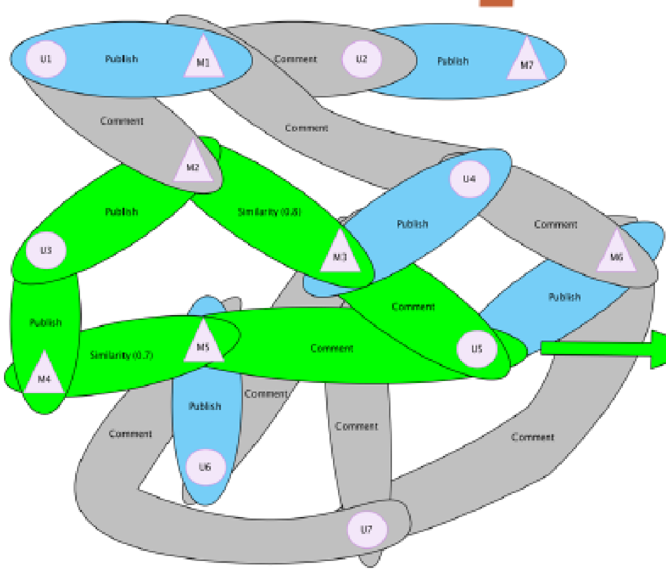
1. a densely connected subset of nodes that is only sparsely linked to the remaining network
2. groups of vertices that probably share common properties and/or play similar roles within the graph.
3. a community as a group of network nodes, within which the links connecting nodes are dense but between which they are sparse



My Proposal

***RelP*$_1$={Publish, Comment}**
***RelP*$_2$={Publish, Similarity, Comment}**

diffusion Pregel



W

	U1	U2	U3	U4	U5	U6	U7
U1	1	1	0	1	0	0	0
U2	0	1	0	0	0	0	0
U3	1	0	1	0	1,5	0,8	1,5
U4	0,8	0	0	1	1	1	1
U5	0	0	0	1	1	0	1
U6	0	0	0	0	1	1	1
U7	0	0	0	0	0	0	1

$W_{35} = \underbrace{(1 \times 0.8 \times 1)}_{RP_1} + \underbrace{(1 \times 0.7 \times 1)}_{RP_2} = 1.5$

WU = W x W^T

	U1	U2	U3	U4	U5	U6	U7	Sum
U1	3	1	1	1,8	1	0	0	7,80
U2	1	1	0	0	0	0	0	2,00
U3	1	0	7,14	4,6	3	3,8	1,5	21,04
U4	1,8	0	4,6	4,64	3	3	1	18,00
U5	1	0	3	3	3	2	1	13,00
U6	0	0	3,8	3	2	3	1	12,80
U7	0	0	1,5	1	1	1	1	5,50
Sum	7,80	2,00	21,04	18,04	13,00	12,80	5,50	

	Incremental expantion	conductance score
Iteration 1		
Seed 1	3	0,661
	3, 4	0,581
	3, 4, 6	0,495
	3, 4, 6, 5	0,427
	3, 4, 6, 5, 7	0,394
	3, 4, 6, 5, 7, 1	0,367
	3, 4, 6, 5, 7, 1, 2	0,358
Community 1	3, 4, 6, 5, 7, 1, 2	

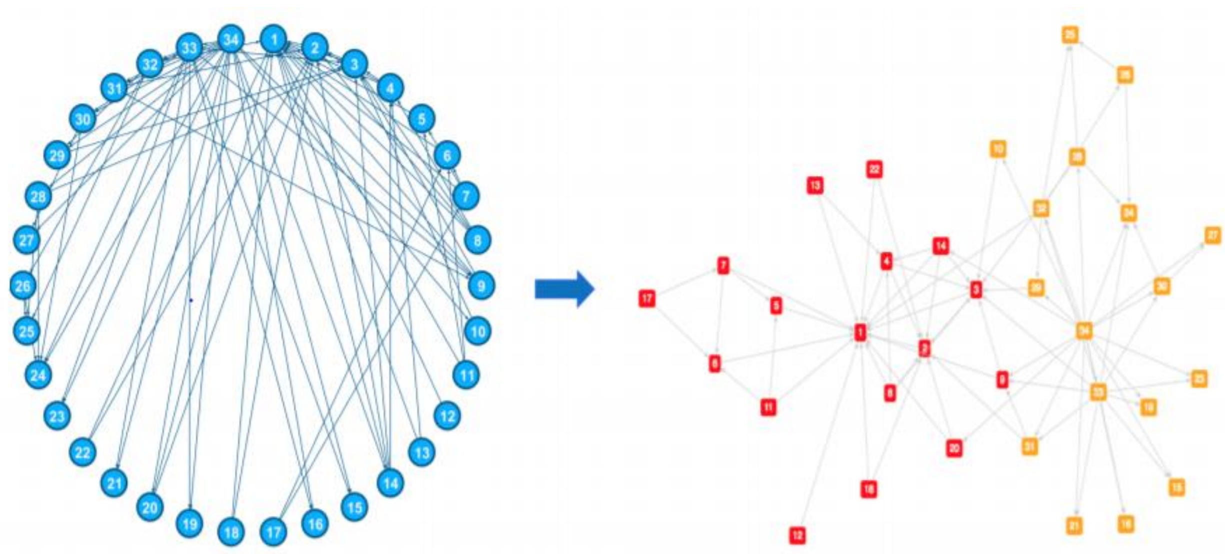
$$\phi(C_i) = \frac{cut(C_i)}{min\{\deg(C_i), \deg(\bar{C}_i)\}}$$

Effectiveness Evaluation (1/2)

Zackary’s karate club (*)

- (34 nodes, 78 pairwise links) → 2 communities

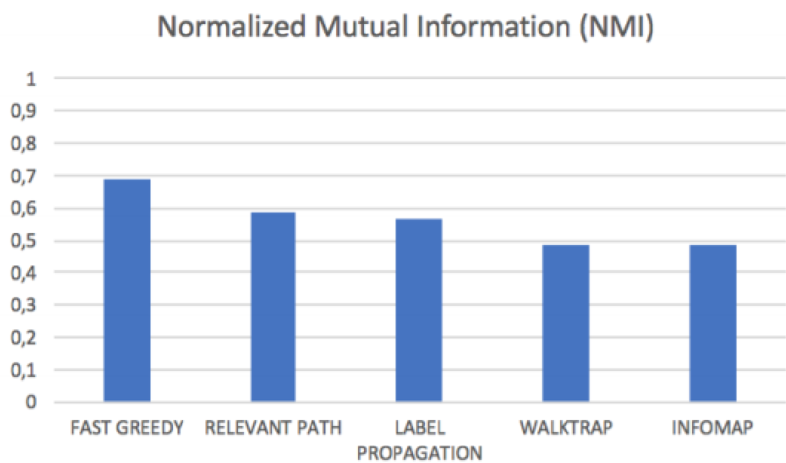
- **Compared algorithms:**
  - Fast Greedy
  - Label Propagation
  - InfoMap
  - Walktrap
- **Quality metrics:**
  - NMI
  - ARI
  - TP – FP

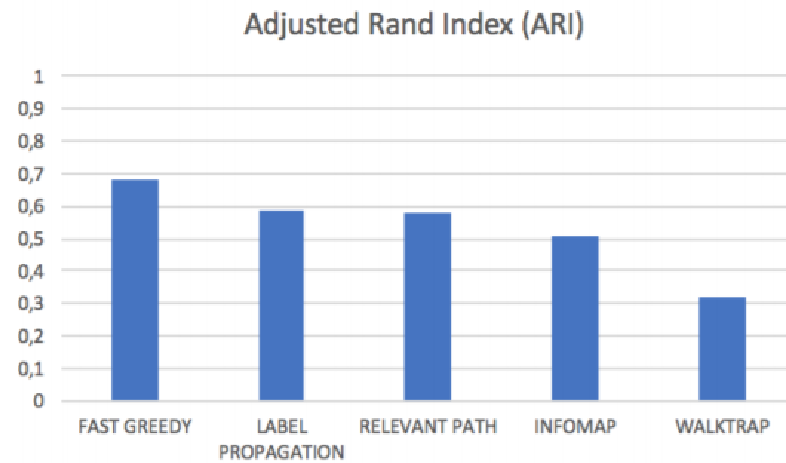


*https://networkdata.ics.uci.edu/data.php?id=105

Effectiveness Evaluation (2/2)

Experimental 1 Results

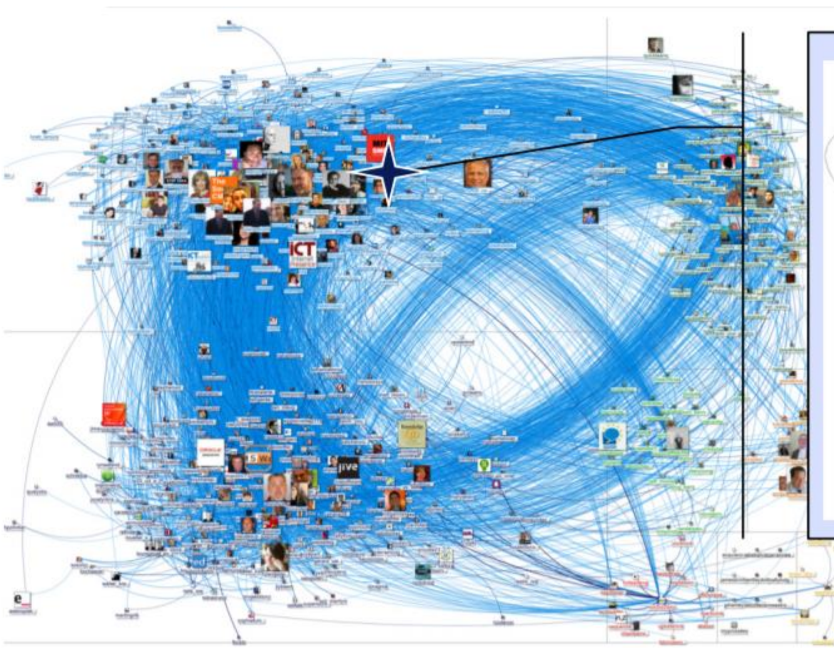


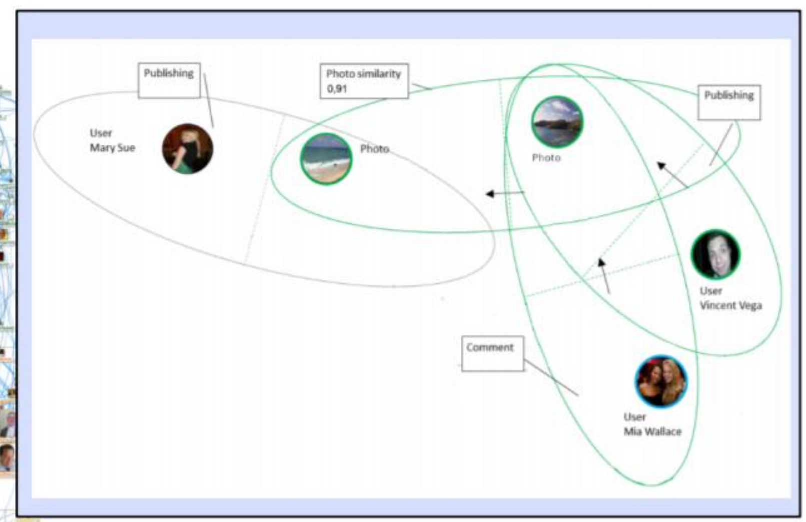


	Relevant Path		Infomap		Fast Greedy		Label		Walktrap	
	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
Community 1	88,24%	17,65%	94,12%	11,76%	100,00%	5,88%	94,12%	5,88%	52,94%	0,00%
Community 2	82,35%	0,00%	52,94%	5,88%	64,71%	0,00%	64,71%	5,88%	52,94%	0,00%
Community 3	0,00%	100,00%	0,00%	100,00%	0,00%	100,00%	0,00%	100,00%	0,00%	100,00%
Community 4									0,00%	100,00%
Community 5									0,00%	100,00%

Qualitative Evaluation

Dataset	N Communites	N hyperedge
Without Similarity	115	9585
Similarity ≥ 0.90	11	13452
Similarity ≥ 0.80	26	29700
Similarity ≥ 0.70	106	224412





Conclusions

- This dissertation is mainly focused on the design of novel data model relying on hypergraph data structure for representing MSN sufficiently general with respect to:
  1. A particular social information network;
  2. The different kinds of entities;
  3. The different types of relationships;
  4. The different applications
- The features of the proposed model have been used to deal with the following two challenges:
  – Influence Maximization
  – Community detection
- The evaluation, made on Flickr, shows:
  – how the proposed approaches can be properly faced IM problem leveraging the introduced model.
  – how the proposed community detection approach has similar performance w.r.t. the well-known algorithms
  – How the proposed approach can be properly faced with community detection problem leveraging the features of heterogeneous networks.

References (1/4)

Journal Papers

1. Sperlì, G., Amato, F., Moscato, V., & Picariello, A. (2016). Multimedia social network modeling using hypergraphs. International Journal of Multimedia Data Engineering and Management (IJMDEM), 7(3), 53-77.
2. Amato, F., Moscato, V., Picariello, A., Sperlí, G., D'Acierno, A., & Penta, A. (2017). Semantic summarization of web news. Encyclopedia with Semantic Computing and Robotic Intelligence, 1(01), 1630006.
3. Amato, F., Castiglione, A., De Santo, A., Moscato, V., Picariello, A., Persia, F., & Sperlí, G. (2017). Recognizing human behaviours in online social networks. Computers & Security.
4. Amato, F., Moscato, V., Picariello, A., Piccialli, F., & Sperlí, G. Centrality in heterogeneous social networks for lurkers detection: An approach based on hypergraphs. Concurrency and Computation: Practice and Experience.

References (2/4)

Accepted Journal Papers

1. Sperlì, G., Amato A., Mercorio F., Mezzazanica M., Moscato, V., & Picariello, A. (2017). A Social Media Recommender System. International Journal of Multimedia Data Engineering and Management (IJMDEM).
2. Amato F., Castiglione A., Moscato M., Picariello P., Sperlì G. Multimedia summarization using social media content, Multimedia Tools and Applications (MTAP).

Submitted Journal Papers

1. Chakraborty T., Jajodia S., Katz J., Picariello A., Sperlì G., Subrahmanian V.S., FORGE: A Fake Online Repository Generation Engine for Cyber Deception. Transaction on Information Forensics & Security.
2. Sperlì G. Big Data and Recommendation System. Encyclopedia of Big Data Technologies.

References (3/4)

Conference Papers

1. A. D'Acierno, F. Gargiulo, V.Moscato, A. Penta, F. Persia, A. Picariello, C. Sansone, G. Sperlì "A Multimedia Summarizer integrating Text and Images" (21-33) IIMSS 2015, Sorrento, Italy.
2. F. Amato, A. De Santo, F. Gargiulo, V. Moscato, F. Persia, A. Picariello, G. Sperli: "A Novel Approach to Query Expansion based on Semantic Similarity Measures" (344-353), DATA 2015:
3. F. Amato, A. De Santo, V. Moscato, A. Picariello, D. Serpico, G. Sperli: "A Lexicon-Grammar Based Methodology for Ontology Population for e-Health Applications"(521-526) CISIS 2015, Blumenau, Brazil
4. F. Persia, D. D'Auria, G.Sperlì, A. Tufano: "A prototype for Anomaly Detection in Video Surveillance Context"(517-528) Somet 2015, Naples, Italy:
5. Amato, F., Moscato, V., Picariello, A., & Sperlí, G. (2016, February). Multimedia social network modeling: A proposal. In Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on (pp. 448-453). IEEE.
6. Amato, F., Moscato, V., Picariello, A., & Sperlí, G. (2016, March). Modelling multimedia social network for topic ranking. In Advanced Information Networking and Applications Workshops (WAINA), 2016 30th International Conference on (pp. 81-86). IEEE.
7. Amato, F., Moscato, V., Picariello, A., & Sperlí, G. (2016, March). Modelling multimedia social network for topic ranking. In Advanced Information Networking and Applications Workshops (WAINA), 2016 30th International Conference on (pp. 81-86). IEEE.
8. Flora, A., Vincenzo, M., Antonio, P., & Giancarlo, S. (2016, November). Modeling User-Content Interaction in Multimedia Social Networks Using Hypergraphs. In Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on (pp. 343-350). IEEE.
9. Amanto, F., Moscato, V., Picariello, A., & Sperlı, G. Recommender Systems and Social Networks: an application in Cultural Heritage.
10. Amato, F., Moscato, V., Picariello, A., & Sperlí, G. (2017, January). Kira: a system for knowledge-based access to multimedia art collections. In Semantic Computing (ICSC), 2017 IEEE 11th International Conference on (pp. 338-343).

References (4/4)

Conference Papers

11. Amato, F., Moscato, V., Picariello, A., & Sperlí, G. (2017, April). Recommendation in Social Media Networks. In Multimedia Big Data (BigMM), 2017 IEEE Third International Conference on (pp. 213-216). IEEE.
12. Amato, F., Cozzolino, G., Moscato, V., Picariello, A., & Sperlí, G. (2017, March). Automatic Personalization of Visiting Path Based on Users Behaviour. In Advanced Information Networking and Applications Workshops (WAINA), 2017 31st International Conference on (pp. 692-697). IEEE.
13. Amato, F., Moscato, V., Picariello, A., & Sperlí, G. (2017, May). Influence Maximization in Social Media Networks Using Hypergraphs. In International Conference on Green, Pervasive, and Cloud Computing (pp. 207-221). Springer, Cham.
14. Cozzolino G, Amato F., Moscato V., Picariello A., Sperlì G "Sentiment Analysis on yelp social network" KES 2017
15. Amato, F., Moscato, V., Picariello, A., & Sperlı, G. Diffusion Algorithms in Multimedia Social Networks: a preliminary model. ASONAM 2017
16. Flora Amato, Vincenzo Moscato, Antonio Picariello, Giovanni Ponti, Giancarlo Sperlì: Influence Analysis in Business Social Media. MIDAS@PKDD/ECML 2017: 43-54
17. Amato, F., Castiglione, A., Moscato, V., Picariello, B. A., & Sperlı, G. (2017, November). Detection of Lurkers in Online Social Networks. In Cyberspace Safety and Security: 9th International Symposium, CSS 2017, Xi'an China, October 23–25, 2017, Proceedings (Vol. 10581, p. 1). Springer.