di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

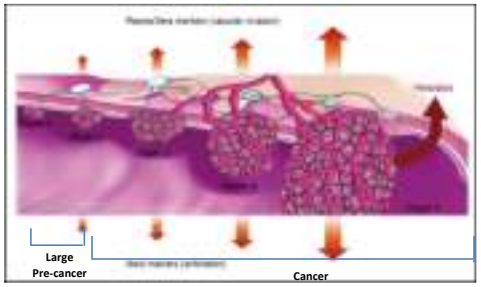
Smart e-Health Systems - Prima parte

15:10	15:30	**Analisi dati di stili di vita alimentari**	***Francesca Cordero***, *Alessandro Mazzei, Luca Anselma, Rossano Schifanella*
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

Diet and lifestyle characteristics

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Colorectal Cancer



Results of miRNA and gut microbiome analyses will be further evaluated in relation to diet and lifestyle factors collected by the questionnaire in the study.

Characteristics		All (n=48)	Controls (n=24)	Polyps (n=9)	Inflammations (n=12)	CRC (n=3)
Alcohol consumption	No	18	10	2	5	1
	Yes	28	13	7	6	2
Alcohol consumption (wine, (glass per week)	Low(< 3)	6	1	2	3	0
	Medium (3-6)	4	3	1	0	0
	High (>6)	18	9	4	3	2
Wine consumption (glass per week)	No	20	10	3	6	1
	Moderate (<5)	9	5	3	1	0
	Frequent (>5)	17	8	3	4	2
Coffee consumption (Yes/No)	No	6	1	1	3	1
	Yes	40	22	8	8	2
Coffee consumption	1	10	4	3	3	0
	2-3	22	12	3	5	2
	>3	8	7	1	0	0
Milk consumption	No	28	14	4	7	3
	yes	18	9	5	4	0
Sugar	No	9	3	4	2	0
	Yes	37	20	5	9	3
(number of teaspoon for day)	1	7	3	2	1	1
	2-4	20	11	1	7	1
	>5	10	6	2	1	1

Characteristics		All (n=48)	Controls (n=24)	Polyps (n=9)	Inflammations (n=12)	CRC (n=3)
Fruit consumption	Yes/No	46/2	22/2	9/0	12/0	3/0
Fruit intake	Portions/week		6	8	6	7
Vegetables consumption (raw+cooked)	Yes/No	48/0	24/0	9/0	12/0	3/0
Vegetables intake (raw+cooked)	Portions/week		2,22	3,5	1,9	3,6
Pasta/Bread consumption	Yes/No	47/1	24/0	9/0	12/0	2/1
Pasta/bread intake	Portions/week		4,8	5,3	3,7	2,7
Cheese	Yes/No	45/3	23/1	9/0	9/3	3/0
Cheese intake	Portions/week		0	3,1	3,8	3,7
Total Meat consumption	Yes/No	47/1	24/0	9/0	11/1	3/0
Meat intake	Portions/week		6	5	4	5
Fish consumption	Yes/No	37/11	5/3	6/3	8/4	3/0
Fish intake	Portions/week		1	1	1	2
Yogurt	Yes/No	26/22	23/11	6/3	7/5	1/2

Benvenuti nel sistema AcQUE!

Dr Alessio Naccarati
Dr Mario Trompetto
Dr Carlo Senore

LILT
prevenire è vivere

IIGM
Italian Institute for Genomic Medicine

Clinica Santa Rita Vercelli

CPO

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Smart e-Health Systems - Prima parte

15:10	15:30	**Analisi dati di stili di vita alimentari**	*Francesca Cordero,* ***Alessandro Mazzei, Luca Anselma****, Rossano Schifanella*
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

MADiMan: Multimedia Application for Diet Management

Intelligenza artificiale per la gestione della dieta

Luca Anselma -> anselma@di.unito.it

Alessandro Mazzei -> mazzei@di.unito.it

The diet

-> In a diet it is necessary to consider **total energy** requirements and the specific required amounts of macronutrients such as **proteins**, **carbohydrates** and **lipids.**

-> In the literature systems of Dietary Reference Values (**DRV**s) recommended to be followed for significant amounts of time

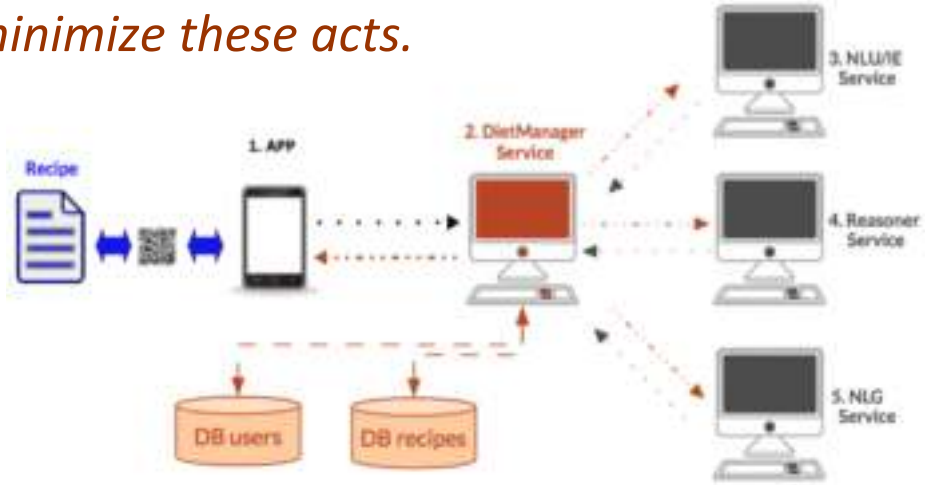
The diet transgression problem

What are the consequences of diet transgressions over the week?

Anselma, L., Mazzei, A., and Michieli, F. D. (2017). **An artificial intelligence framework for compensating transgressions and its application to diet management.** Journal of Biomedical Informatics, 68:58–70.

MADiMan: Multimedia Application for Diet Management

... a virtual dietitian that is able: (1) to recover the nutritional information directly from a specific recipe, (2) to reason over recipes and diets with flexibility, i.e. by allowing some forms of diet disobedience, (3) to support users in compensating such acts of disobedience, and (4) to persuade the user to minimize these acts.



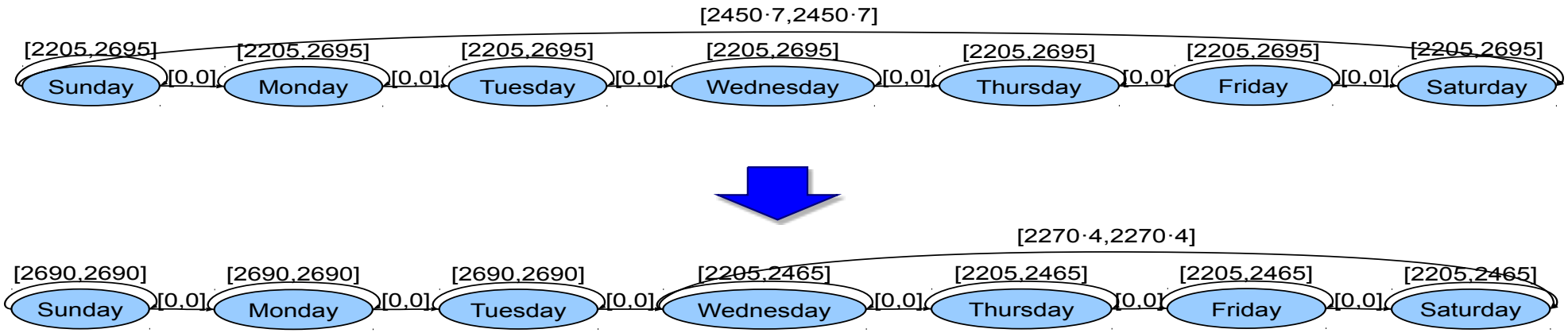
MADiMan has been partially funded by Regione Piemonte, Innovation Hub for ICT, 2011-2014, POR-FESR 07-13

Mazzei, A. et al (2015). **Mobile computing and artificial intelligence for diet management**. In - ICIAP 2015 Workshop: MADiMa, Genoa, Italy, September 7-8, 2015, Proceedings, number 9281 in Lecture Notes in Computer Science, pages 342–349. Springer.

Reasoning about the diet with Simple Temporal Problems (STP)

-> Let us assume that John ate 2690 kcal on Sunday, Monday and Tuesday

-> We add to the STP such new information and propagate the constraints



Anselma, L., Mazzei, A., and Michieli, F. D. (2017). **An artificial intelligence framework for compensating transgressions and its application to diet management.** Journal of Biomedical Informatics, 68:58–70.

Natural Language Generation

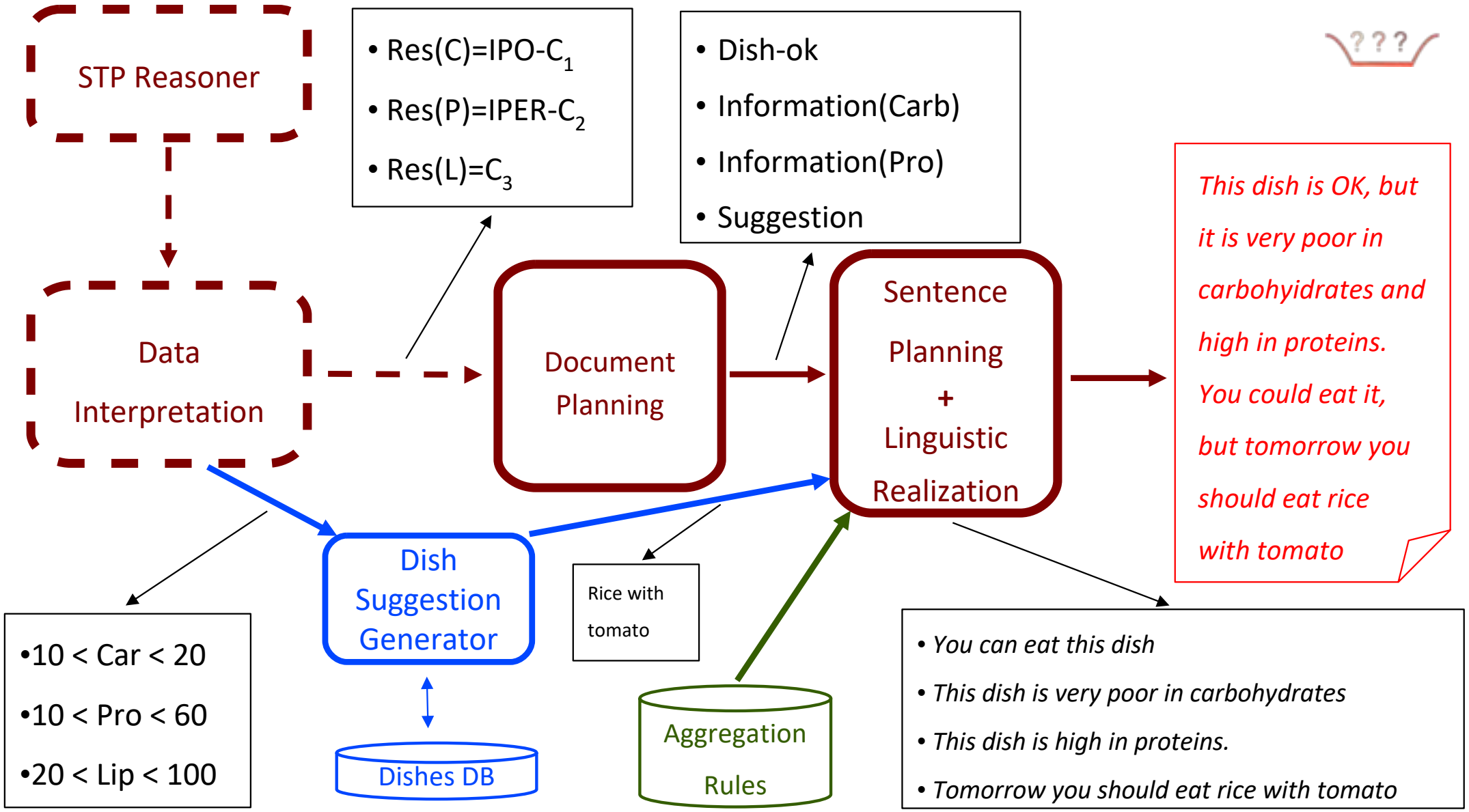
answer +

explanation +

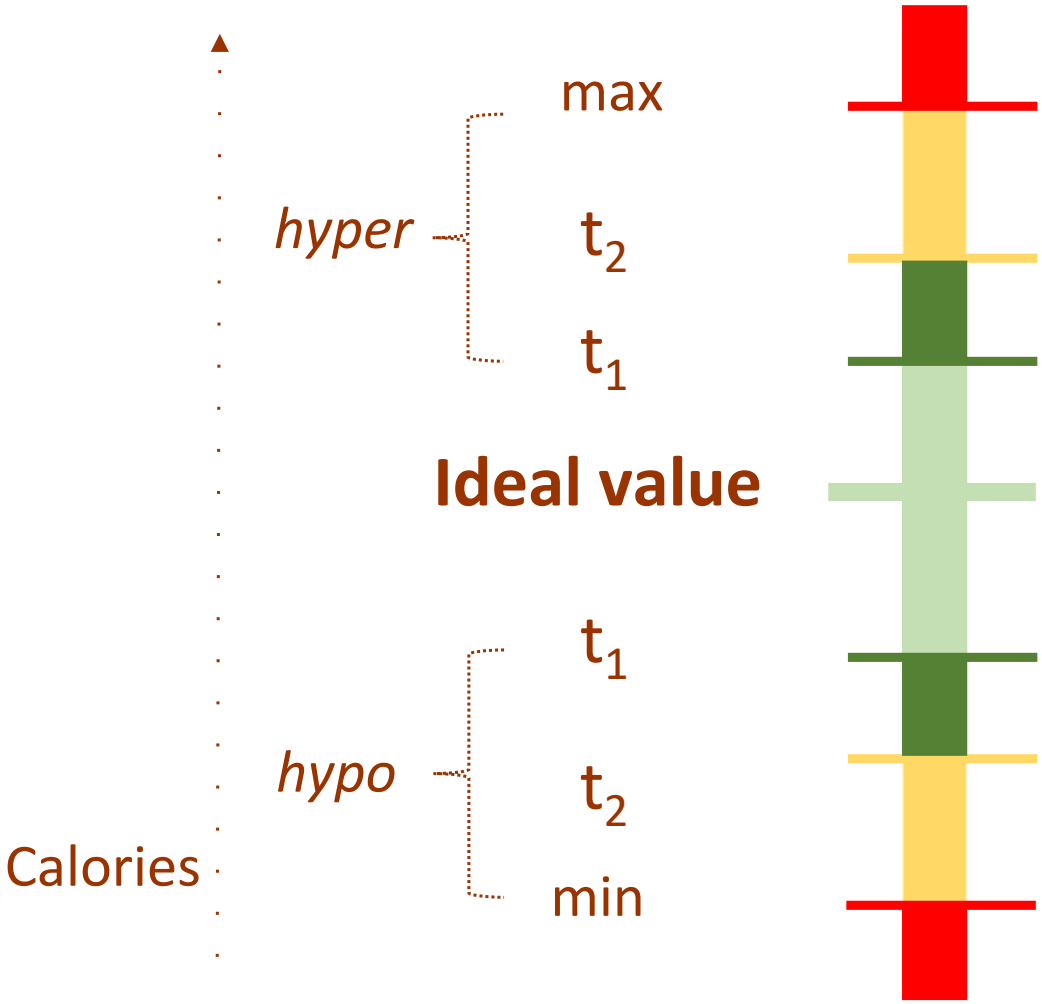
suggestion =

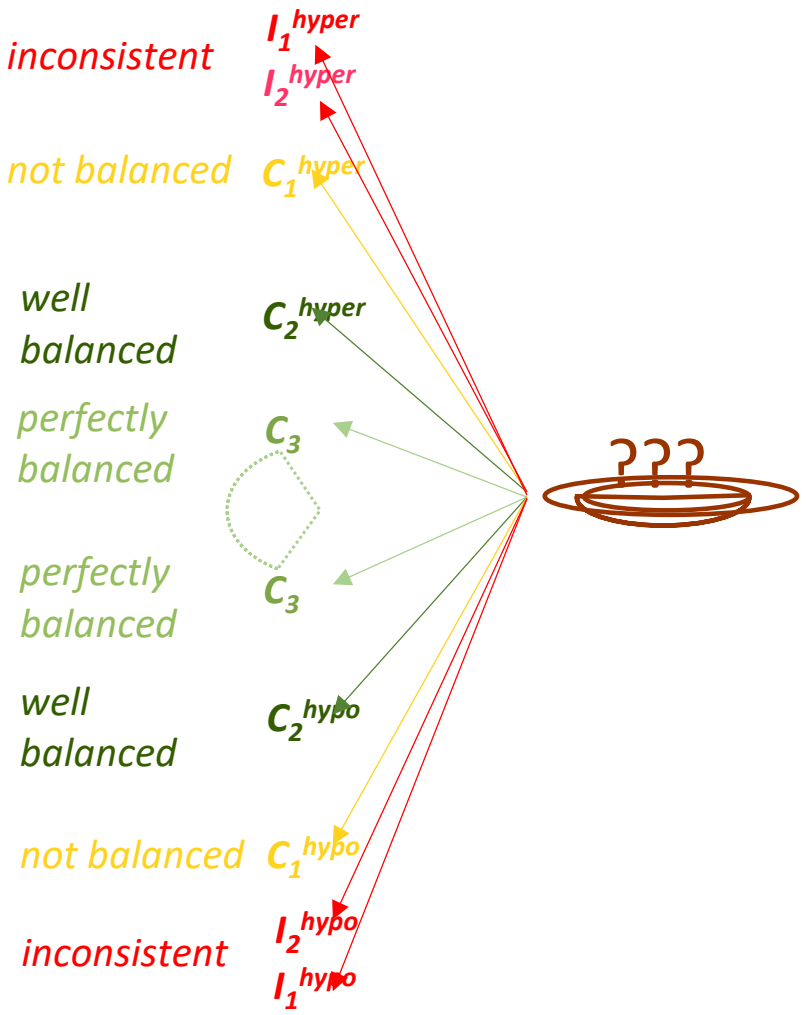
Message

Anselma, L. and Mazzei, A. (2017). **An approach for explaining reasoning on the diet domain**. In Proc. of the 1st Workshop on Natural Language for Artificial Intelligence co-located with 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017)


STP Reasoner
Data Interpretation
• Res(C)=IPO-C1
• Res(P)=IPER-C2
• Res(L)=C3
Document Planning
• Dish-ok
• Information(Carb)
• Information(Pro)
• Suggestion
Sentence Planning + Linguistic Realization
This dish is OK, but it is very poor in carbohyidrates and high in proteins. You could eat it, but tomorrow you should eat rice with tomato
•10 < Car < 20
•10 < Pro < 60
•20 < Lip < 100
Dish Suggestion Generator
Dishes DB
Rice with tomato
Aggregation Rules
• You can eat this dish
• This dish is very poor in carbohydrates
• This dish is high in proteins.
• Tomorrow you should eat rice with tomato

Symbolic interpretation of numerical reasoning





Templates with 1 macro-nutrient

-> I_1 Inconsistent : *This dish is not good at all, it's too poor in proteins!*

-> I_2 Inconsistent : *You cannot have this dish now because it doesn't provide enough proteins, but if you eat a nice dish of beans Sunday, you can have it on Monday.*

Templates with 1 macro-nutrient

-> C_1 **Not balanced** : *This dish is good, but it's poor in proteins. In the next days you'll have to eat far more proteins.*

-> C_2 **Well balanced** : *This dish is OK, but it's a bit poor in proteins. In the next days you'll need more proteins! :)*

-> C_3 **Perfectly balanced** : *Great choice! This dish is perfect for the proteins in your diet :)*

Work in Progress

-> Reasoning

Mediterranean diet

-> NLG

Causal Explanation

Time, Agrovoc Ontology

-> Evaluation

iOS app for dietitians

Thanks for your time!

-> Anselma, L., Mazzei, A., Piovesan, L., and De Michieli, F. (2014). **Adopting STP for diet management**. In Proc. of IEEE International Conference on Healthcare Informatics.

-> Mazzei, A. et al (2015). **Mobile computing and artificial intelligence for diet management**. In - ICIAP 2015 Workshop: MADiMa, Genoa, Italy, September 7-8, 2015, Proceedings, number 9281 in Lecture Notes in Computer Science, pages 342–349. Springer.

-> Anselma, L. and Mazzei, A. (2015)**. Towards diet management with automatic reasoning and persuasive natural language generation.** In Progress in Artificial Intelligence - 17th Portuguese Conference on Artificial Intelligence, EPIA 2015,

-> Mazzei, A. (2015). **Generare messaggi persuasivi per una dieta salutare**. In Proc. of CLiC-it 2015, Second Italian Conference on Computational Linguistics.

-> Anselma, L., Mazzei, A., and Michieli, F. D. (2017). **An artificial intelligence framework for compensating transgressions and its application to diet management.** Journal of Biomedical Informatics, 68:58–70.

-> Anselma, L. and Mazzei, A. (2017). **An approach for explaining reasoning on the diet domain**. In Proc. of the 1st Workshop on Natural Language for Artificial Intelligence co-located with 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017)

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Smart e-Health Systems - Prima parte

15:10	15:30	**Analisi dati di stili di vita alimentari**	*Francesca Cordero, Alessandro Mazzei, Luca Anselma,* ***Rossano Schifanella***
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

Healthy City

Large-scale Study of Food Consumption and Health Outcomes

Rossano Schifanella
Contact: rossano.schifanella@unito.it

Recent & ongoing activities

Who?

UniTO

NOKIA Bell Labs

TESCO

ISI

ISI Foundation
& ISI Global Science
Foundation

What?

Relationship between spatial environment, life style (food) and health

Health surveillance

Data-driven prevention and intervention strategies

How?

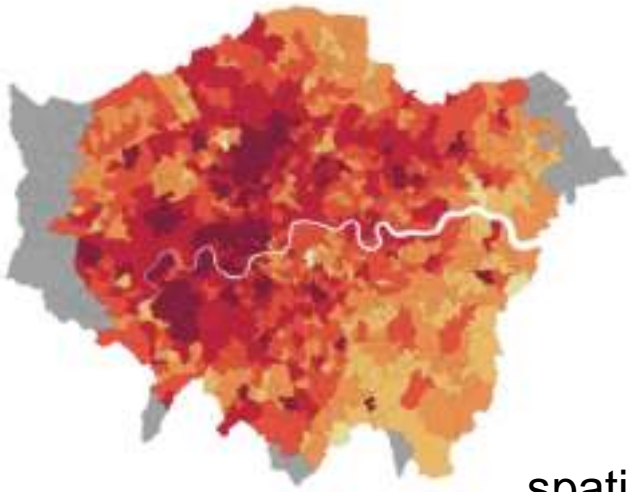
Large-scale data analytics
Machine learning
Spatial analysis and modeling
Network science

Modeling Food Consumption

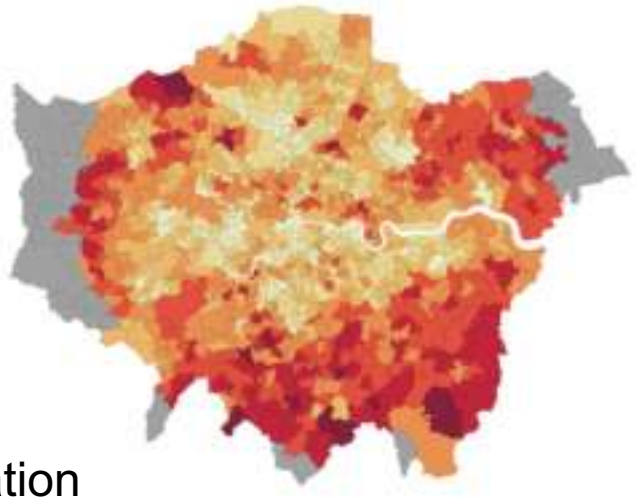
- **Motivation: nutritional routine is a major factor that impacts well-being**
- Tesco dataset:
  - 1.6B food items
  - 1.6M loyalty card owners
  - fine-grained spatial granularity (ZIP code)
  - fine-grained temporal granularity (2015)
- Food consumption footprint at different levels:
  - ingredients
  - macronutrients (e.g., carbohydrates, proteins, fats)
  - micronutrients
  - calories and other indicators of food healthiness

Spatio-temporal and socio-economic dimensions

fibers



nutrients diversity



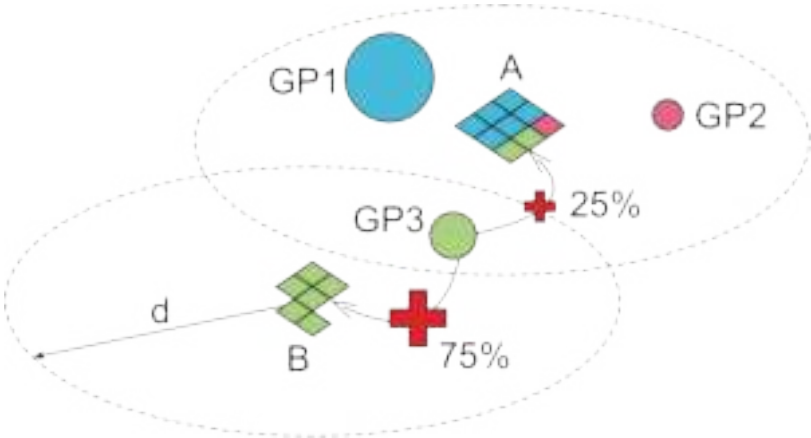
spatial autocorrelation
hot-cold spots
temporal trends
socio-economic factors

Modeling Drugs Prescriptions at scale

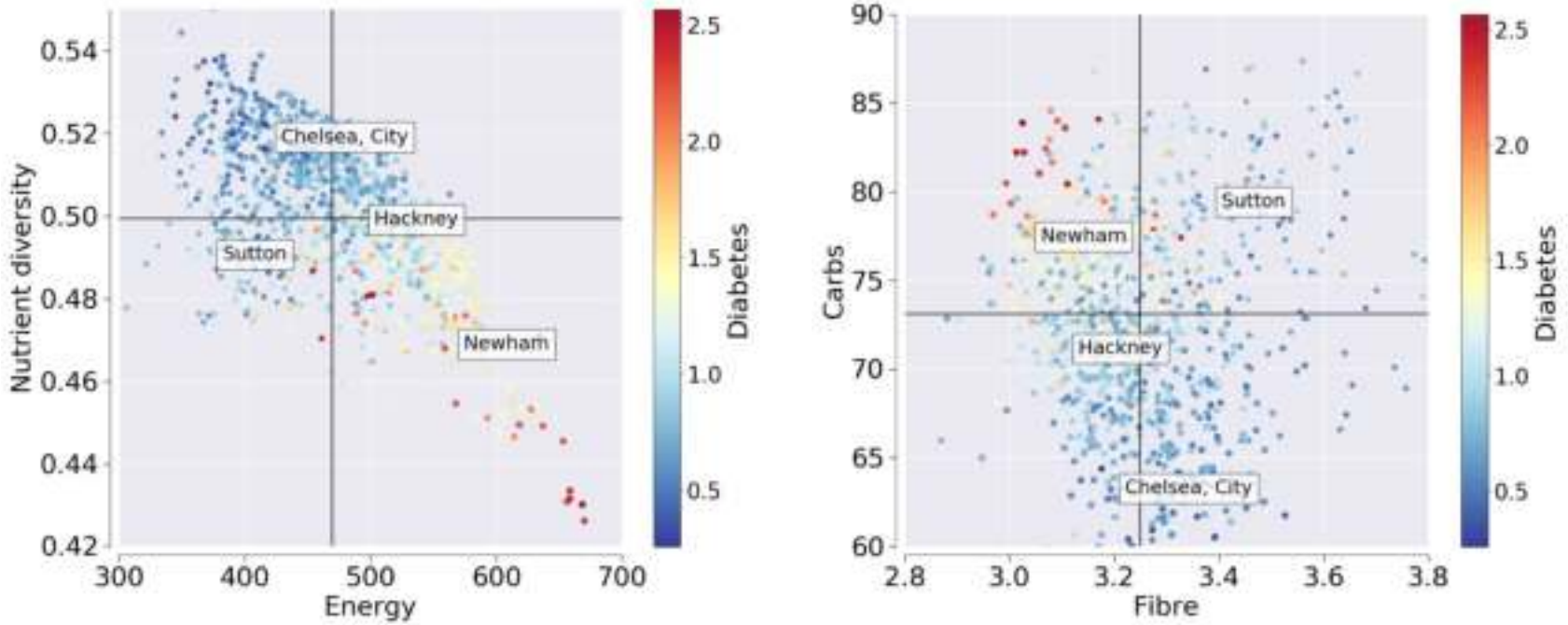
- NHS UK prescriptions (2010-2017)
  - 1.1B per year
  - GP (General Practice) granularity
  - LSOA patients origin, age/gender
- OpenPrescribing

- Focus on metabolic syndrome
  - diabetes, cholesterol, hypertension

- Focus on
  - antibiotics resistance determinants
  - opioids epidemics



Spatio-temporal and socio-economic dimensions



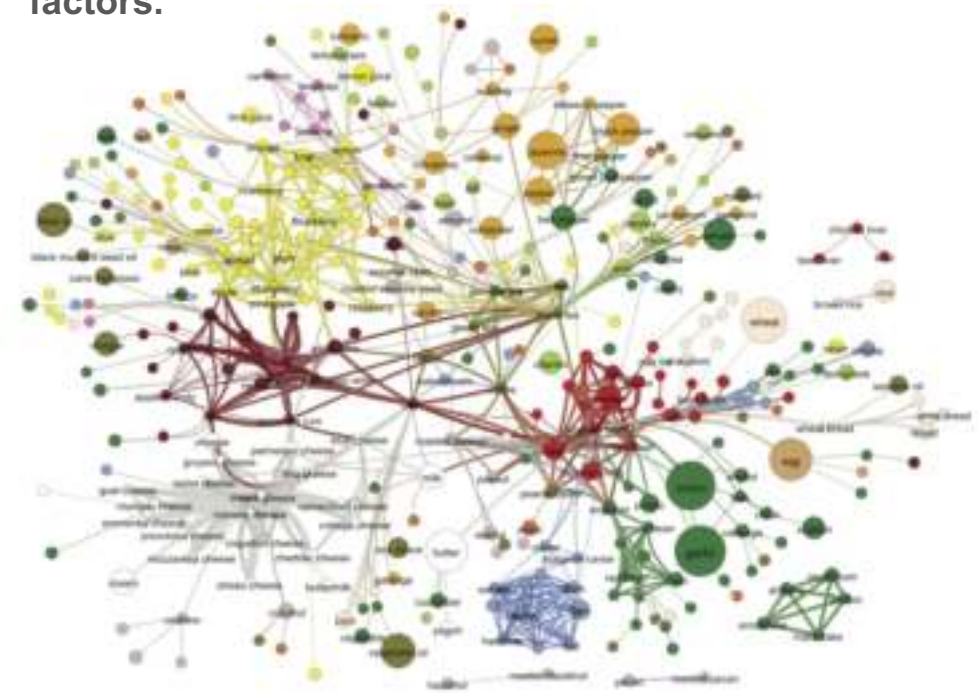
output

Regression models to explain health outcomes as a function of food daily habits
Predictive framework of unhealthy areas of a city

Modeling Daily Habits and Diets

- Network science
- Hypothesis: structural properties are connected with health indicators.
- Promote healthy diets
- Recommender systems based on explorative approaches

Diets evolve over time, being influenced by income, individual preferences, cultural traditions, environmental, geographical, social and economic factors.



Future works

- **Health surveillance and predictive analytics**
- **Prevention strategies**
  - **Public authorities** (inform policies)
  - **Education** (new educational programs that inform people about the dangers of not eating well)
  - **Food companies** (the food industry could reformulate their offering and elaborate plans to improve nutrition)
  - **Technology companies** (predictive analytics and wearable sensors will transform how people manage their health)
- **Design interventions**

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Smart e-Health Systems - Prima parte

15:10	15:30	**Analisi dati di stili di vita alimentari**	*Francesca Cordero, Alessandro Mazzei, Luca Anselma, Rossano Schifanella*
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

Personal Informatics & Assistive Technologies

Federica Cena - Amon Rapp

SIOS Research Group

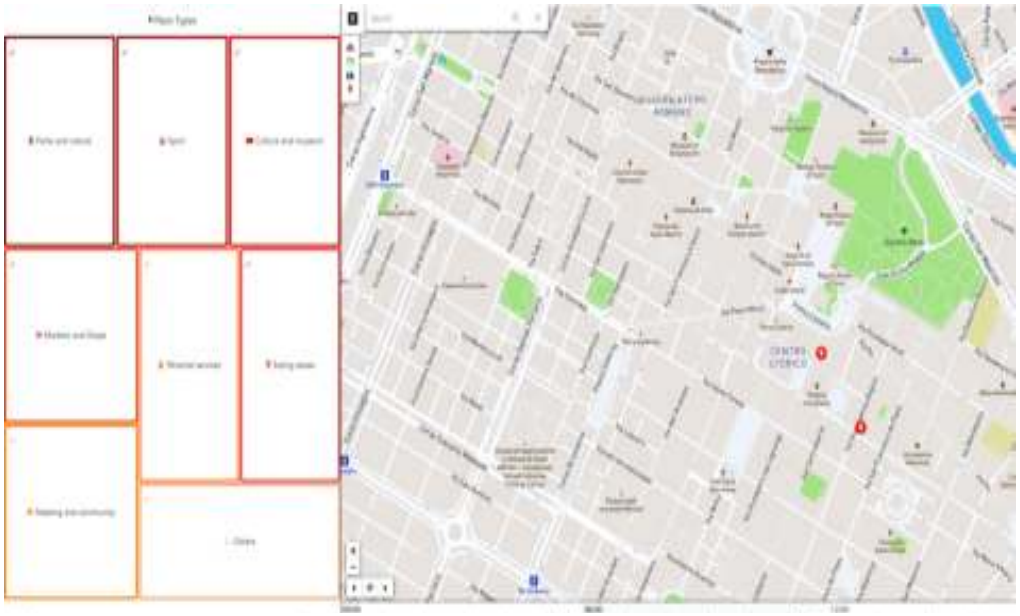
Responsabile: Luca Console

- Il gruppo di ricerca SIOS (Smart Interactive Objects and Systems) è attivo dal 1987
- Il gruppo ha competenze in Interazione Uomo-Macchina e sistemi di raccomandazione e modellazione dell'utente.

Progetto: Personalized Interactive Maps for Autism

Progetto: in corso con **Compagnia San Paolo** e **ASL Città di Torino, Ambulatorio Disturbi dello Spettro Autistico in età adulta (2016-2019)**

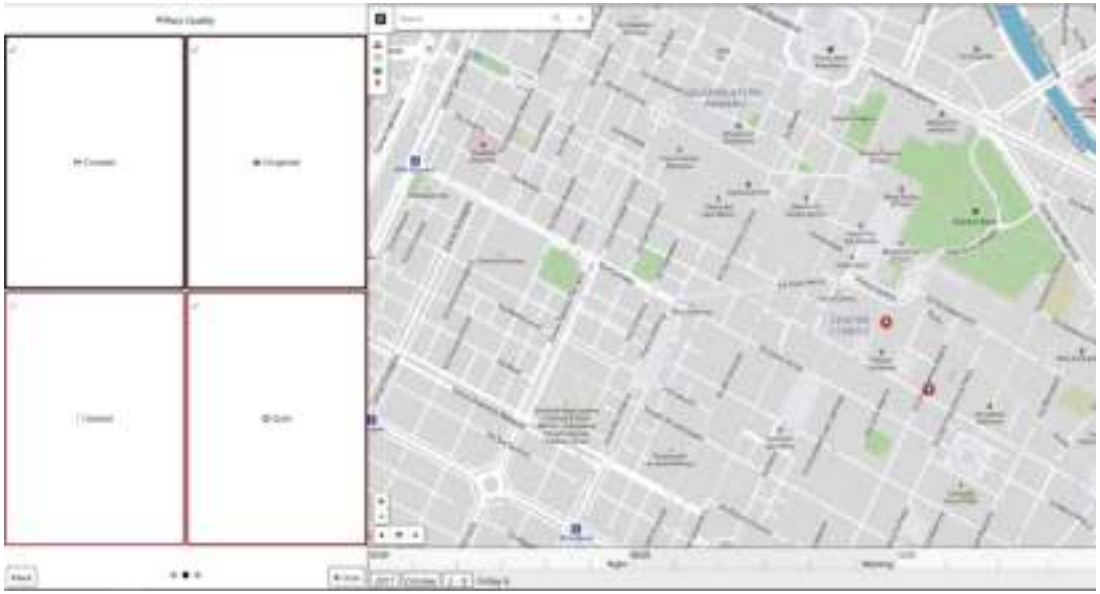
- **Problema**: problemi di persone con autismo in relazione allo spazio
- **Obt**: analisi delle caratteristiche in relazione all'orientamento spaziale e progettazione di strumenti personalizzati per supportarne gli spostamenti e l'esecuzione di attività quotidiane nella città



Progetto: Personalized Interactive Maps for Autism

- **Strumenti**: utilizzo di dispositivi personali e indossabili per raccogliere dati sulla persona e ricavare abitudini, idiosincrasie, interessi, livello di arousal ed utilizzarli per adattare il servizio

- **Sviluppi**: malati di Alzheimer, e individui affetti da agnosia spaziale.



Tema di ricerca: Self-Tracking for health

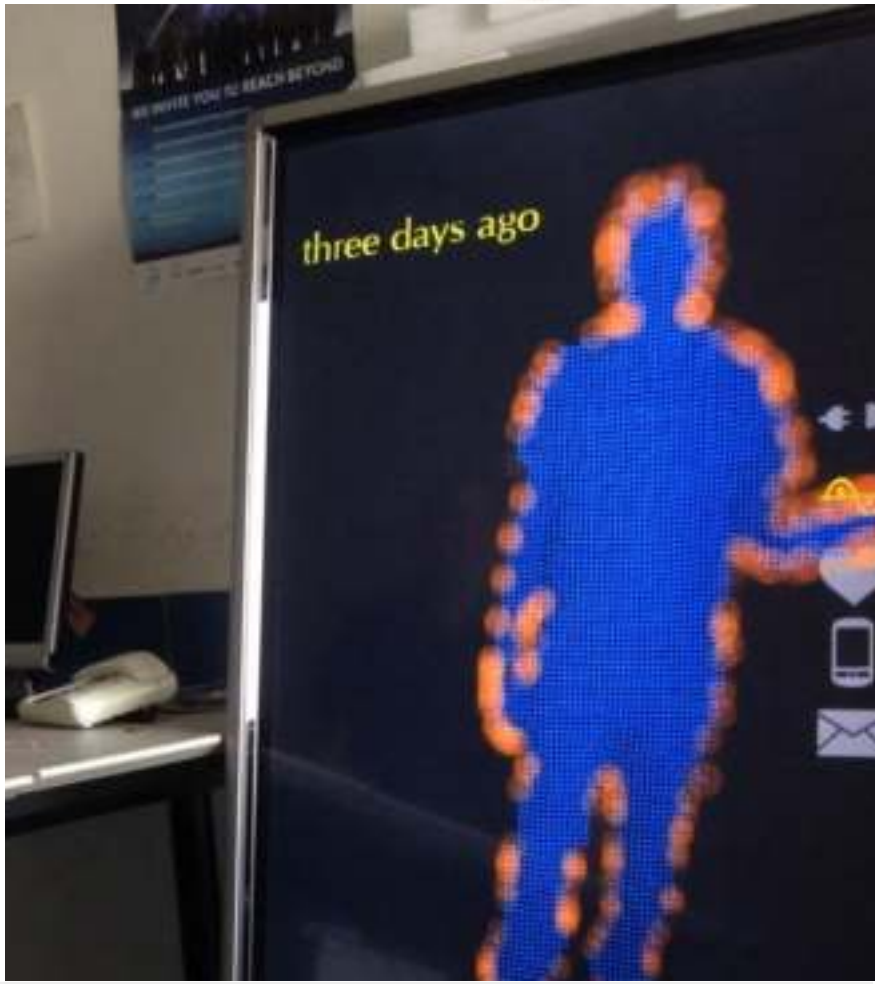
- Diffusione del **self-tracking** di dati personali grazie a strumenti non invasivi (wearable e app) finalizzati al costante monitoraggio di parametri fisiologici.

- **Obiettivo**: studio di nuove tecnologie per tracciamento / analisi / gestione / visualizzazione dati, in modo da migliorare la continuità, la motivazione, efficacia al tracciamento

Tema di ricerca: Self-Tracking for health

- **Progetti:** TIM (2014-ad ora); tracciamento per il wellness in ottica preventiva (num di passi, calorie, per diminuire i rischi della sedentarietà)

- **Sviluppi**: malattie croniche (ad es. diabete, ipertensione, parametri fisiologici e ai farmaci assunti), malattie di difficile diagnosi (ad es. monitoraggio del dolore/sintomi/fattori contestuali per emicrania)



di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Smart e-Health Systems - Prima parte

15:10	15:30	**Analisi dati di stili di vita alimentari**	*Francesca Cordero, Alessandro Mazzei, Luca Anselma, Rossano Schifanella*
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

GLARE (GuideLine Acquisition, Representation and Execution)

Paolo Terenziani, Alessio Bottrighi, Luca Piovesan
DiSIT, Università del Piemonte Orientale, Alessandria, Italy

Luca Anselma
Dip. di Informatica, Università di Torino

Progetto iniziato nel 1997 in collaborazione con
Gianpaolo Molino, Mauro Torchio
Laboratorio di Informatica Clinica, Azienda Ospedaliero Universitaria S. Giovanni Battista, Torino

Paolo Terenziani, Gianpaolo Molino, Mauro Torchio:
A modular approach for representing and executing clinical guidelines. Artificial Intelligence in Medicine 23(3): 249-276 (2001)

Linee Guida Cliniche

USA Institute of Medicine: “sistematically developed statements to assist practitioner and patient decisions about appropriate health care in specific clinical circumstances”.

Vantaggiose per pazienti, medici, ospedali ed organizzazioni sanitarie

Sviluppate migliaia di linee guida (es. ~6500 nella G-I-N Network http://www.g-i-n.net)

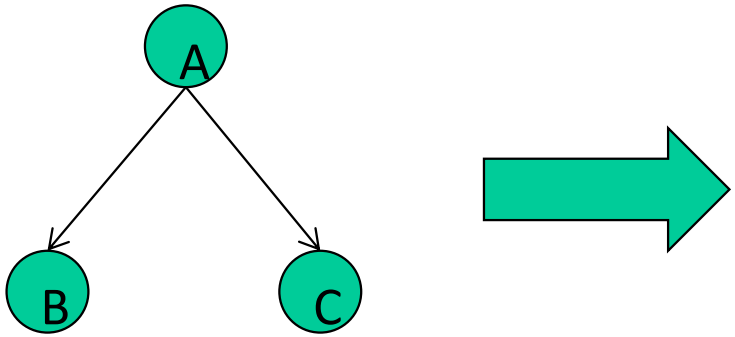
Tuttavia, le linee guida non hanno talvolta ottenuto tutti gli effetti desiderati. Due cause importanti:

Linee Guida Cliniche: problemi

(1) Difficili da interpretare (problemi di formalizzazione)

Il testo (in linguaggio naturale) è non formale, ed ambiguo

Persino i flussi di controllo grafici possono essere ambigui

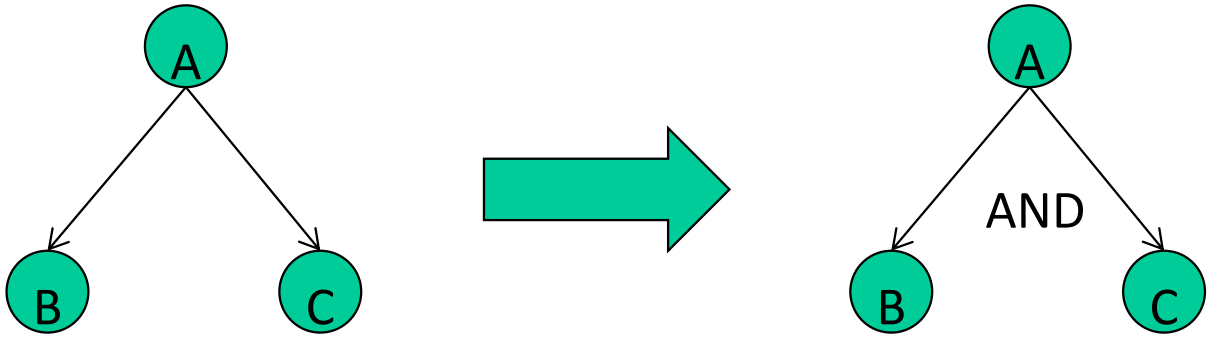


Linee Guida Cliniche: problemi

(1) Difficili da interpretare (problemi di formalizzazione)

Il testo (in linguaggio naturale) è non formale, ed ambiguo

Persino i flussi di controllo grafici possono essere ambigui

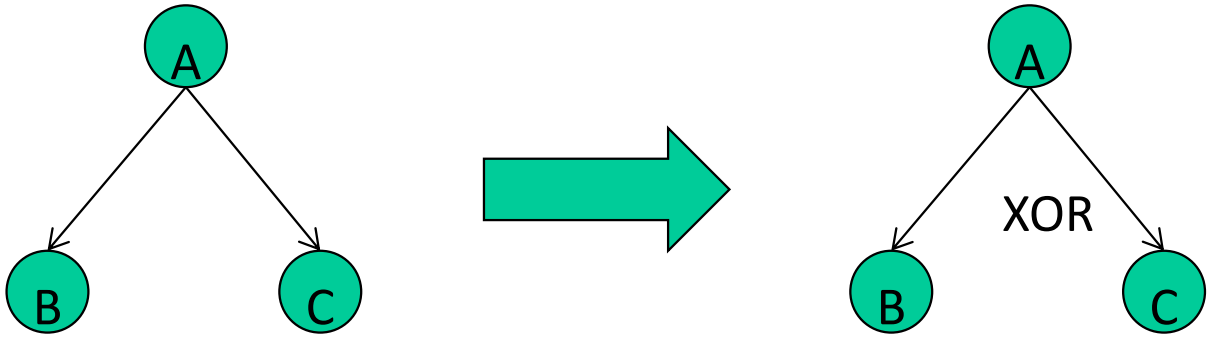


Linee Guida Cliniche: problemi

(1) Difficili da interpretare (problemi di formalizzazione)

Il testo (in linguaggio naturale) è non formale, ed ambiguo

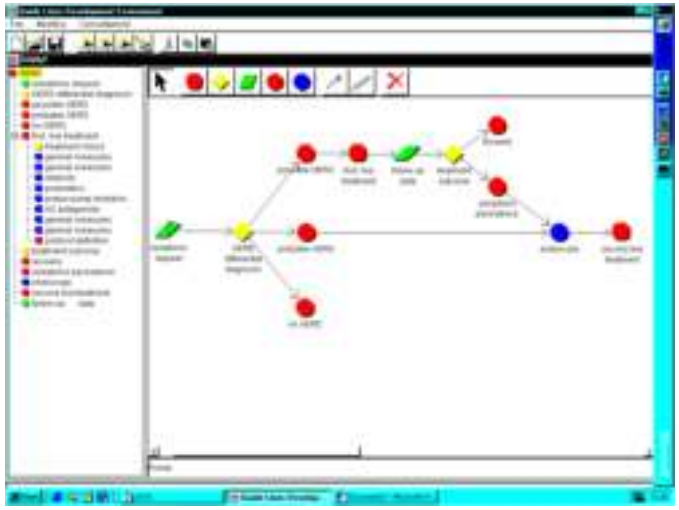
Persino i flussi di controllo grafici possono essere ambigui



GLARE

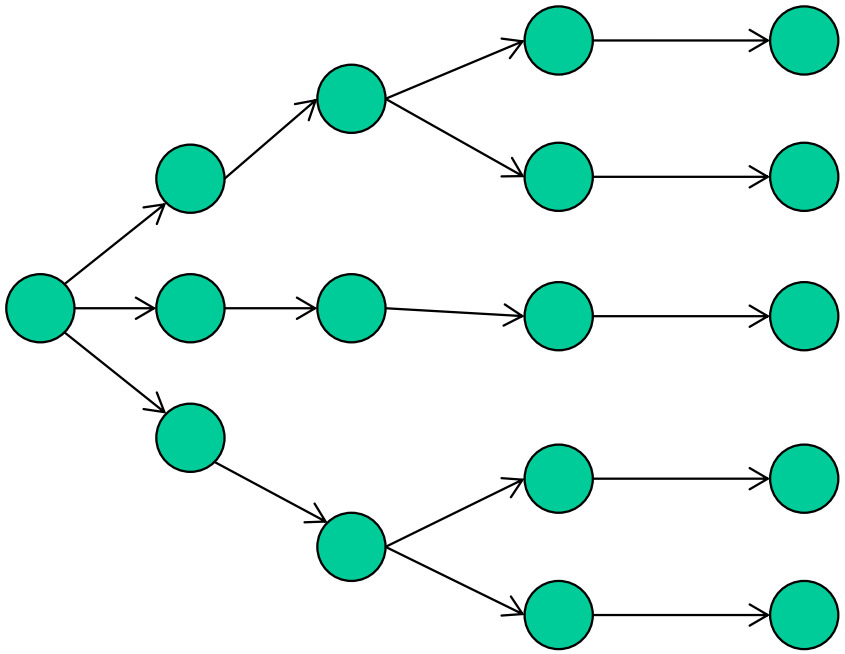
L’adozione di GLARE aiuta a risolvere questo problema

(1) Acquisizione: GLARE fornisce una interfaccia grafica “user-friendly” per acquisire dagli esperti una rappresentazione non ambigua (formale) delle linee guida



Linee Guida Cliniche: problemi

(2) Difficili da applicare allo specifico paziente

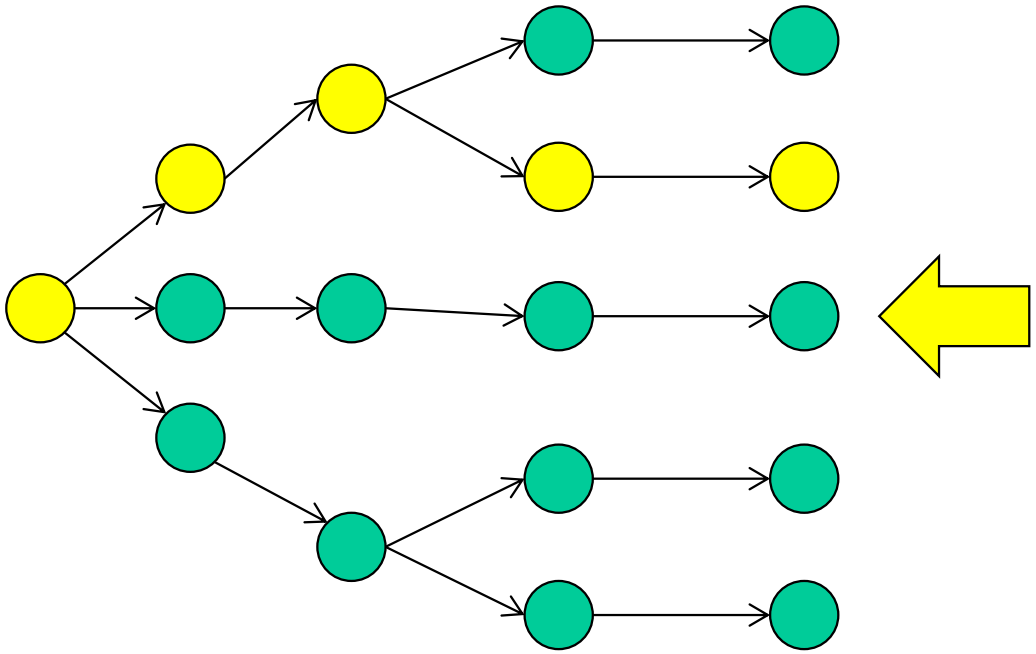


?

Dati Paziente

GLARE

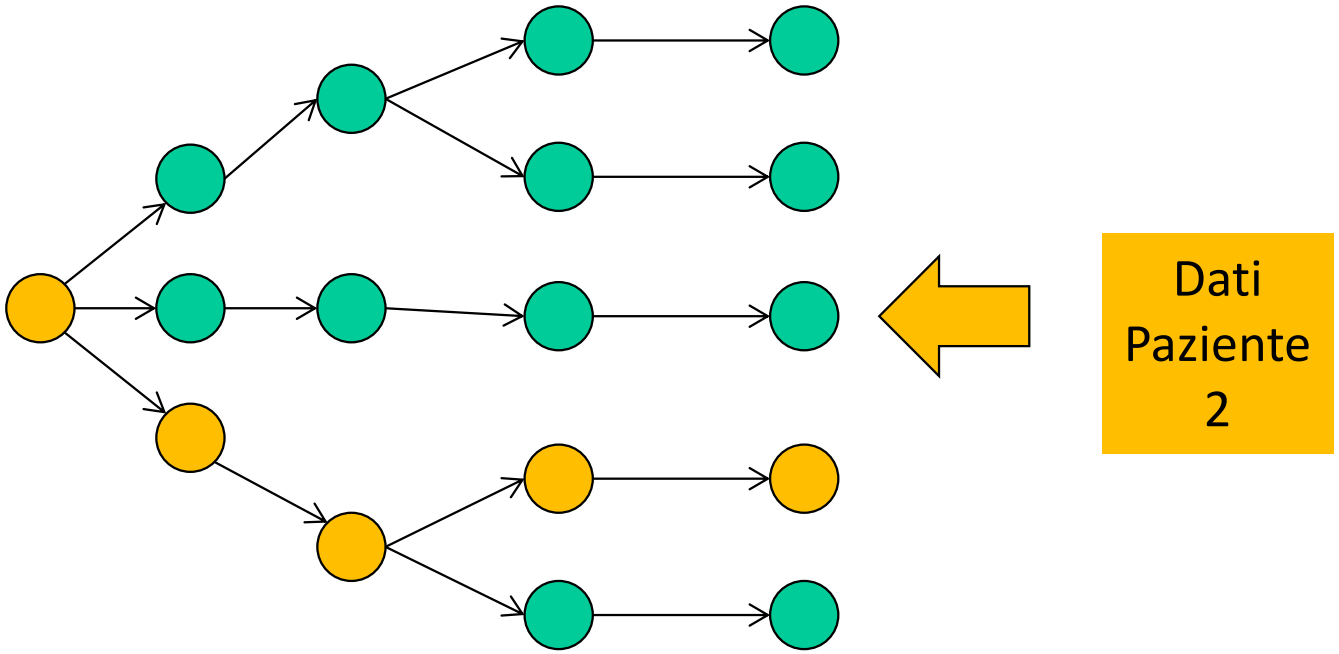
(2) Esecuzione: GLARE fornisce ai medici stumenti per applicare la linea guida allo specifico paziente (sulla base dei dati del paziente), e di supporto alla decisione



Dati Paziente 1

GLARE

(2) Esecuzione: GLARE fornisce ai medici stumenti per applicare la linea guida allo specifico paziente (sulla base dei dati del paziente), e di supporto alla decisione



GLARE
CARATTERISTICHE GENERALI

INDIPENDENTE DAL DOMINIO:

(e.s., cancro alla vescica, reflusso gastro-esofageo, insufficienza cardiaca, ictus ischemico, disturbi legati all'alcol)

VANTAGGIOSO: un unico strumento (ed un unico tipo di rappresentazione) per acquisire ed eseguire linee guida differenti

→ Libreria "omogenea" di linee guida standardizzate

GLARE
CARATTERISTICHE GENERALI

"ORIENTATO AI MEDICI"

Le specifiche di GLARE sono state fornite dai medici del progetto

- Interfaccia grafica
- Meccanismi di supporto alla decisione (es. analisi "what if")
- Formalismo di rappresentazione (7 tipi di entità e 2 tipi di archi)

REPERIMENTO AUTOMATIZZATO DEI DATI DALLA CARTELLA CLINICA DEL PAZIENTE

(progetto DAISY, DBMS Caché)

ANCORA PRESENTE LA "PREDISPOSIZIONE" alla connessione alle cartelle cliniche

GLARE
STATO CORRENTE

ASPETTI SCIENTIFICI

- Parte di OpenClinical (www.openclinical.org)
- Progetti (es. RoPHS –coord. Prof. Renato Balduzzi)
- Oltre 50 pubblicazioni scientifiche
- Relazioni invitate, best paper awards

ASPETTI IMPLEMENTATIVI

- Prototipo in Java del “nucleo” del sistema (studenti)
- In corso una nuova implementazione del software (META-GLARE)

GLARE: estensioni in corso e future

Trattamento di “eccezioni” e\o comorbidita’ (pazienti “atipici”)

Giorgio Leonardi, Alessio Bottrighi, Gabriele Galliani, Paolo Terenziani, Antonio Messina, Francesco Della Corte: Exceptions Handling within GLARE Clinical Guideline Framework. American Medical Informatics Association Symposium, Washington, DC, November 2012. Distinguished Paper Award.

Luca Piovesan, Paolo Terenziani: A Constraint-Based Approach for the Conciliation of Clinical Guidelines. Proc. of the 15th Ibero-American Conference on AI (IBERAMIA’16), San José, Costa Rica, November 2016, 77-88, Best Paper Award.

GLARE: estensioni in corso e future

Analisi di conformità

Matteo Spiotta, Paolo Terenziani, Daniele Theseider Dupré:
Temporal Conformance Analysis and Explanation of Clinical Guidelines Execution: An Answer Set Programming Approach. IEEE Trans. Knowl. Data Eng. 29(11): 2567-2580 (2017)

Supporto al coordinamento di agenti multipli in contesti eterogenei

Alessio Bottrighi, Luca Piovesan, Paolo Terenziani:
Supporting Multiple Agents in the Execution of Clinical Guidelines. Proc. HEALTHINF 2018, Madeira, January 2018, 208-219.

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Smart e-Health Systems - Prima parte

15:10	15:30	**Analisi dati di stili di vita alimentari**	*Francesca Cordero, Alessandro Mazzei, Luca Anselma, Rossano Schifanella*
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

Complex Networks to model epidemics and fake news spreading

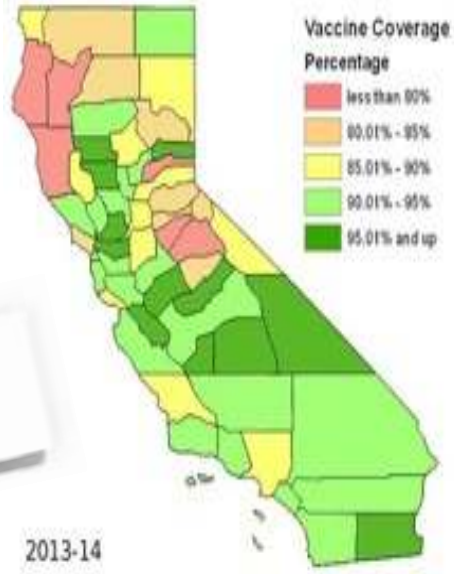
Dipartimento Informatica nel Parco della Salute, della Ricerca e dell'Innovazione - *Ruffo Giancarlo*

The problem of misinformation

di.unito.it
Dipartimento di Informatica
Università degli Studi di Torino

di.unito.it
Dipartimento di Informatica
Università degli Studi di Torino

The Spread of Online Misinformation



source: California Dept. of Public Health

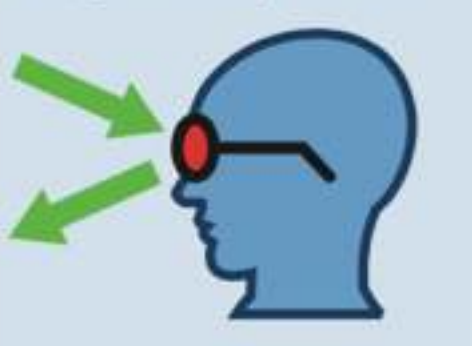
DEBUNKING SERVICES

Fact-checking may be effective …

… or fact-checking may fail

BACKFIRE EFFECT

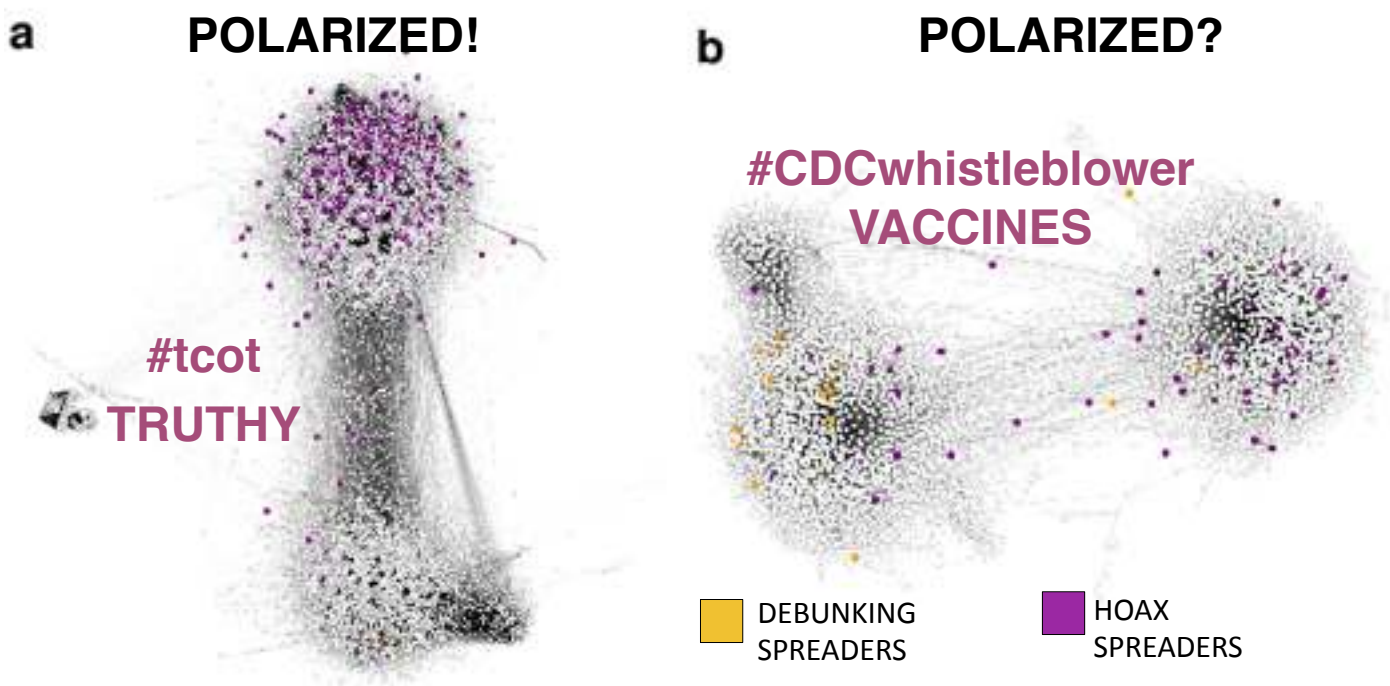
(Nyham - Lewandowsky, 2012)



CONFIRMATION BIAS

cherry-picking data
to support personal
beliefs or expectations

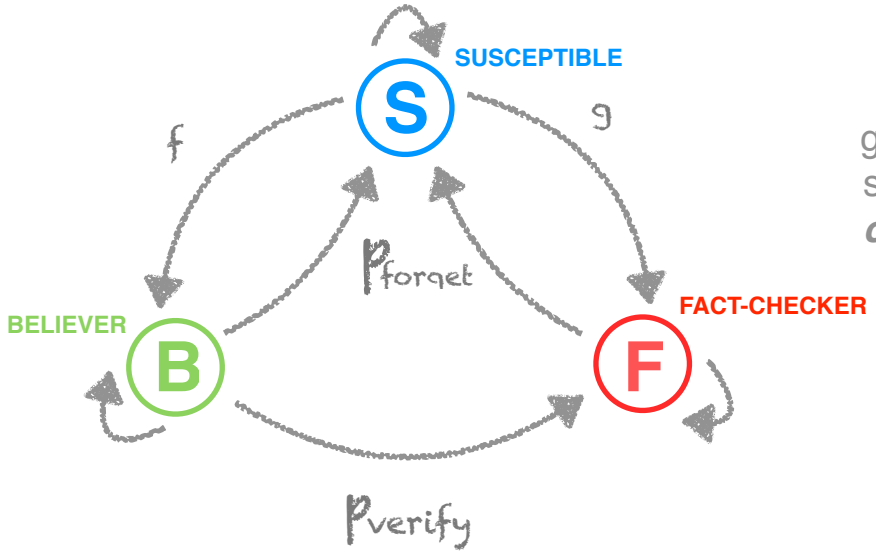
Role of segregation: not clear!



DOES SEGREGATION HELP OR LIMIT THE MISINFORMATION SPREAD?

SBFC MODEL (2015)

Competition **hoax VS debunking**,
forgetting and **fact-checking** mechanisms



CHALLENGING VALIDATION
gauging the exposition is not trivial,
since ***belief does not necessarily correspond to the online activity***

Tambuscio, M., Ruffo, G., Flammini, A., & Menczer, F. (2015, May). Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In Proceedings of the 24th International Conference on World Wide Web (pp. 977-982). ACM.

AGENTS SIMULATIONS (Barabasi-Albert nets)

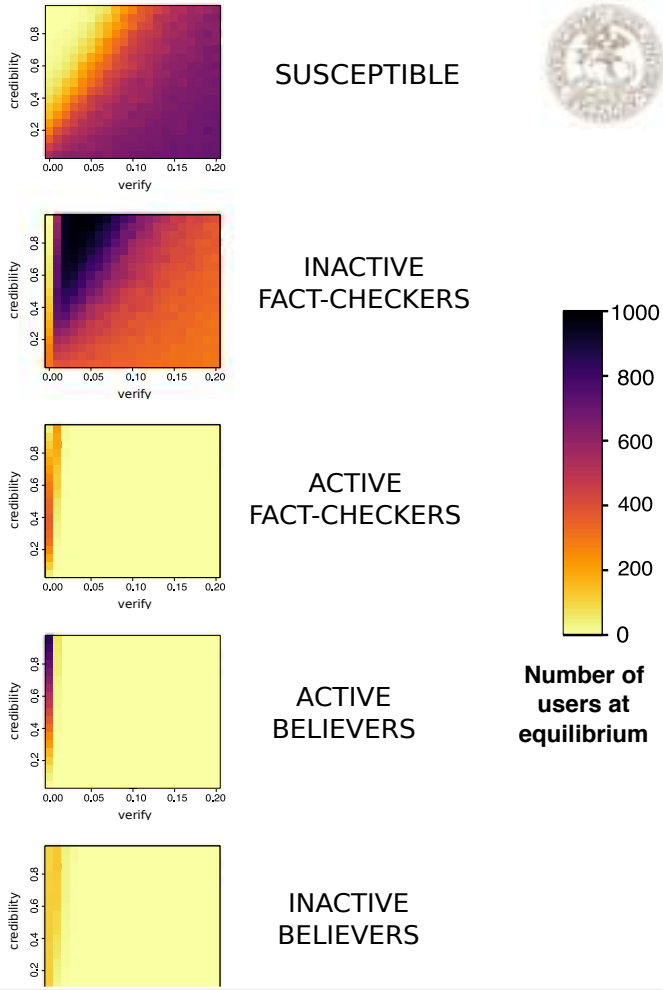
higher credibility

↓

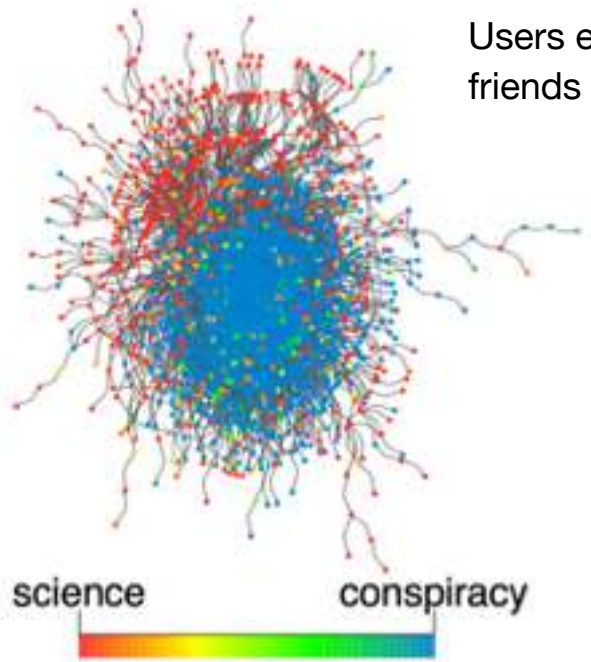
larger diffusion
(less susceptible)

verifying probability is the most important parameter

verify ≃ 0 → hoax survives
verify > 0 → hoax dies out



Misinformation tend to polarize



"Viral Misinformation: The Role of Homophily and Polarization" (Bessi et al. 2015)

Users engagement correlates with the number of friends having similar consumption patterns

homophily!

GULLIBLE USERS
(never verify)

what happens if skeptic interact with gullible?

Gullible Communities

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

SKEPTIC

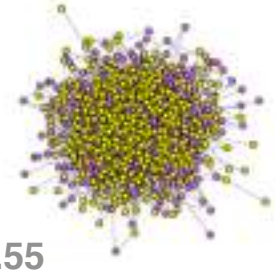
verify > 0

SIZE ($0 < \omega < N$)
#nodes in gullible community

SEGREGATION ($0.5 < s < 1$)
fraction of edges within same community [Gu-Gu, Sk-Sk]

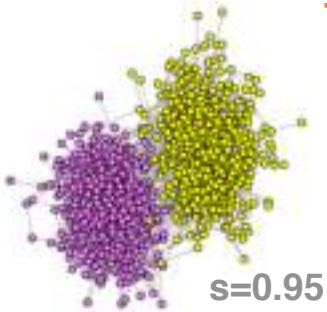
GULLIBLE

verify = 0



s=0.55
ω =500

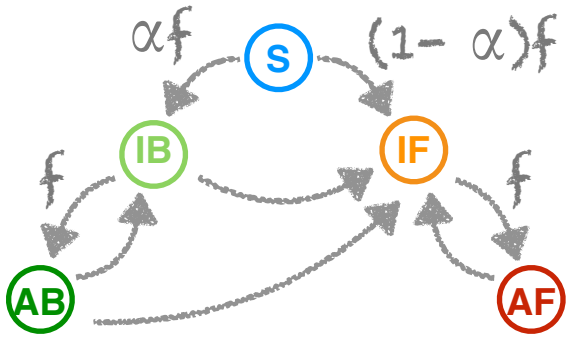
s=0.8
ω =500



s=0.95
ω =500

GULLIBLE

verify = 0



SKEPTIC

verify > 0

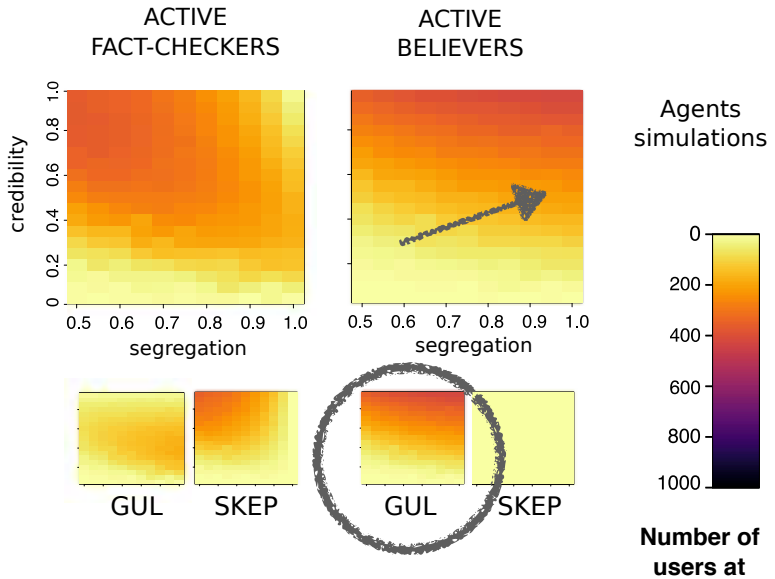
AGENTS SIMULATIONS

AB polarize
in the gullible group

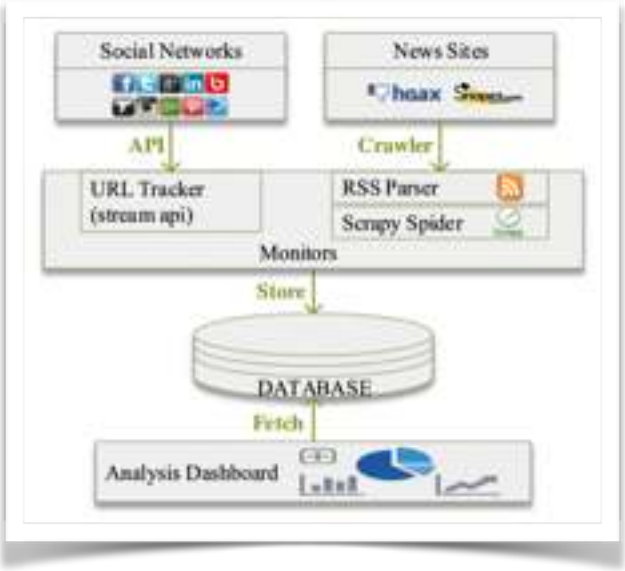
Higher segregation

▼

more Active Believers
(weak effect)



Validation: Hoaxy



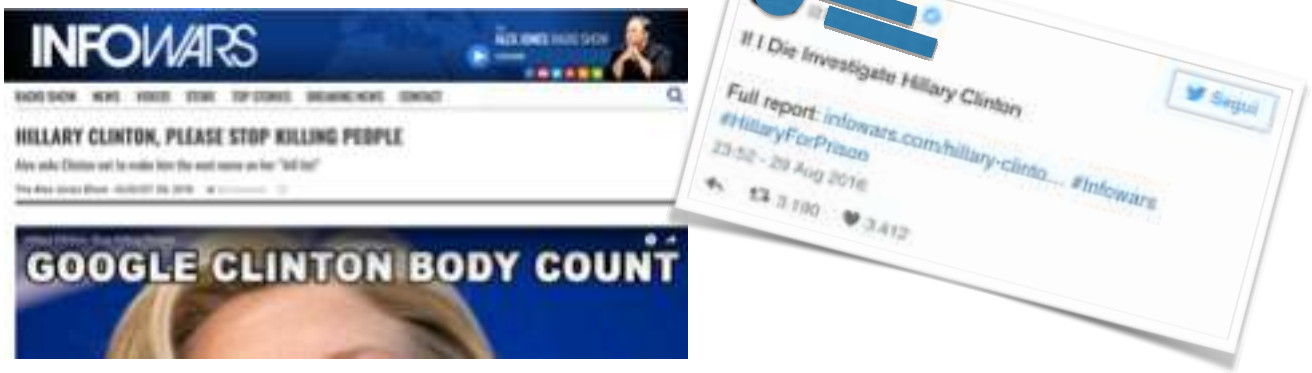
http://hoaxy.iuni.iu.edu/

"Hoaxy: A platform for tracking online misinformation" (Shao et al., 2016)

Validation: data collection

We collected 31 stories
that produced
more than 1000 tweets

FAKE NEWS TWEETS
(containing fake news URLs)
compared to Active Believers



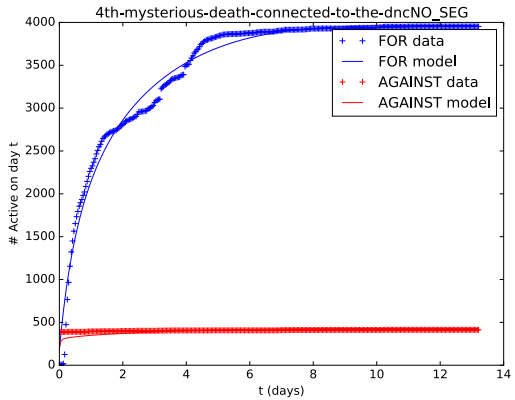
Validation: data collection

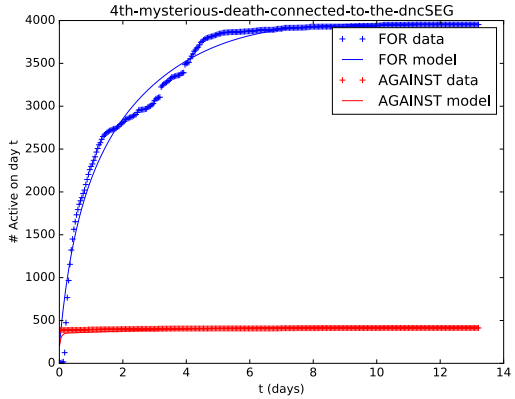
We collected 31 stories
that produced
more than 1000 tweets

FAKE NEWS TWEETS
(containing fake news URLs)
compared to Active Believers

FACT-CHECK TWEETS
(containing Snopes URLs)
compared to Active Fact-checkers

Validation results





Both models fit well data collected with Hoaxy platform!


Marcella Tambuscio **Giancarlo Ruffo**

University of Turin *arcs.di.unito.it*





Chengcheng Shao | Giovanni L Ciampaglia | Diego Fregolente | **Alessandro Flammini** | Filippo Menczer

Indiana University *cnets.indiana.edu*

QUESTIONS?


Daniela Paolotti Ciro Cattuto

ISI Foundation https://www.isi.it

Alessandro Vespignani

Northeastern University https://cos.northeastern.edu

di.unito.it
DIPARTIMENTO DI INFORMATICA
Università degli Studi di Torino

Smart e-Health Systems - Prima parte

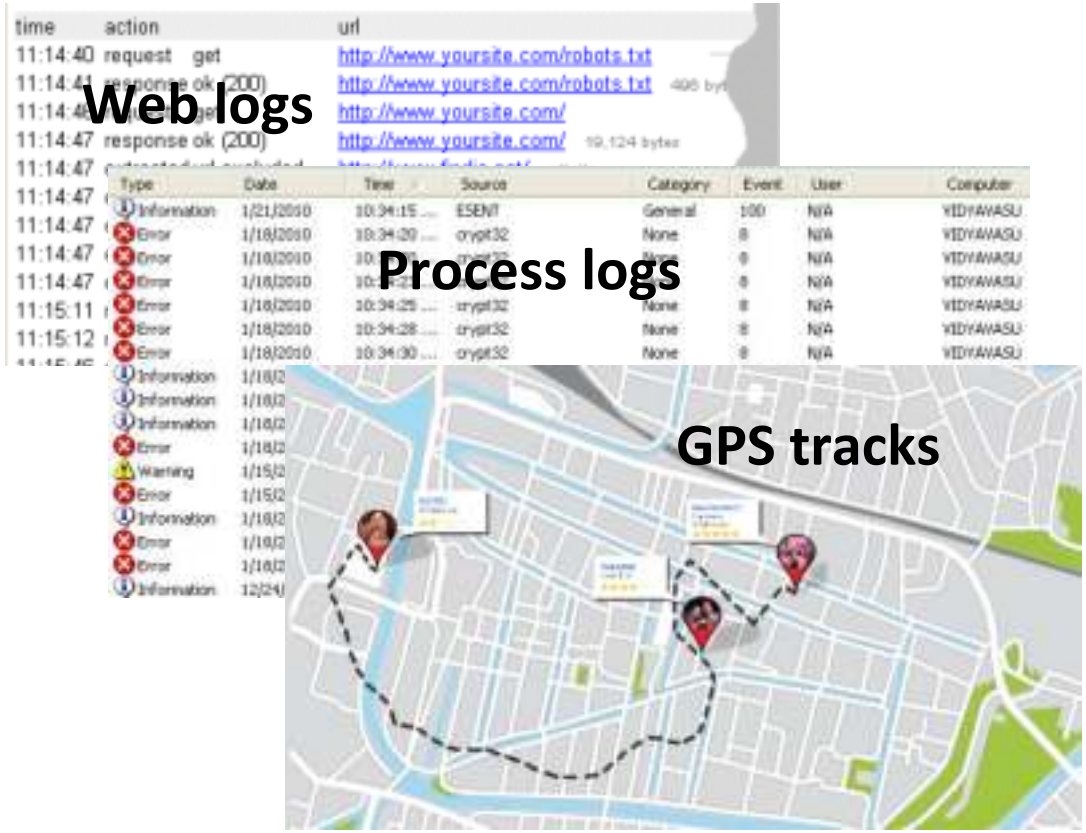
15:10	15:30	**Analisi dati di stili di vita alimentari**	*Francesca Cordero, Alessandro Mazzei, Luca Anselma, Rossano Schifanella*
15:30	15:40	**Personal Informatics & Assistive Technologies**	*Federica Cena / Amon Rapp*
15:40	15:50	**GLARE: AI per le linee guida cliniche**	*Paolo Terenziani/ Luca Anselma*
15:50	16:00	**Social network analysis per la modellazione della diffusione di fake news di contenuto medico**	*Marco Beccuti/ Giancarlo Ruffo*
16:00	16:10	**Privacy-preserving data management and analytics**	*Ruggero Pensa*

Privacy-preserving data management and analytics

Anonymization, differential privacy and privacy metrics for Privacy-by-Design compliant services

Ruggero G. Pensa (pensa@di.unito.it)

Anonymization of sequential data

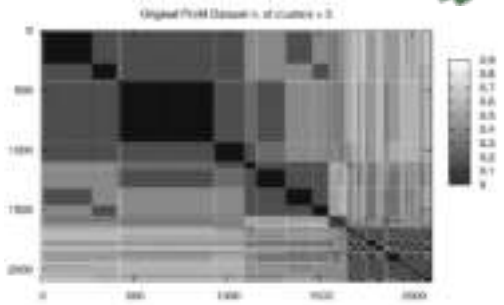


k-anonymous sequence database

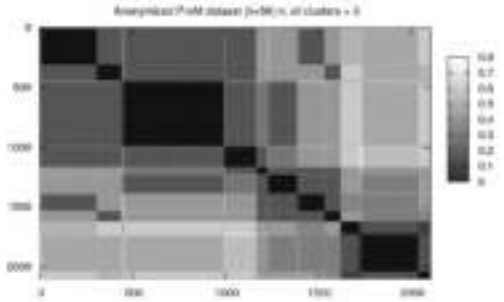
A. Monreale, D. Pedreschi, **R.G. Pensa**, F. Pinelli. Anonymity Preserving Sequential Pattern Mining. Artificial Intelligence and Law. Vol. 22(2) **2014**. pp 141-173. Springer.

Anonymization of sequential data

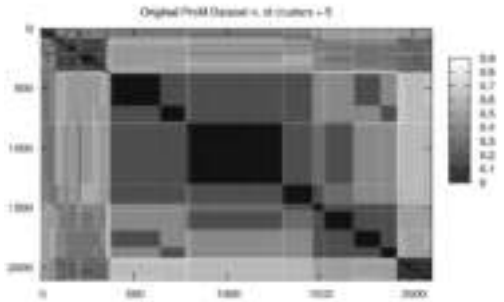
Clustering results on GPS tracks in Milan area



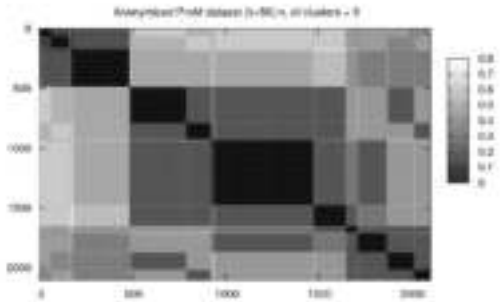
(a) Original data (3 clusters)



(b) k=50 (3 clusters)



(d) Original data (5 clusters)



(e) k=50 (5 clusters)

Metrics for measuring user privacy leakage risk

Fondazione CRT

- Web-service users have a privacy budget (fully-charged battery)
- Each risky action or private data disclosure consumes the budget (discharging battery)

Our privacy score is based on **psychometric** theories, **computational social science** and **machine learning** methods

R.G. Pensa, G. Di Blasi. A privacy self-assessment framework for online social networks. Expert Systems with Applications. Vol. 86 **2017**. pp. 18-31. Elsevier.

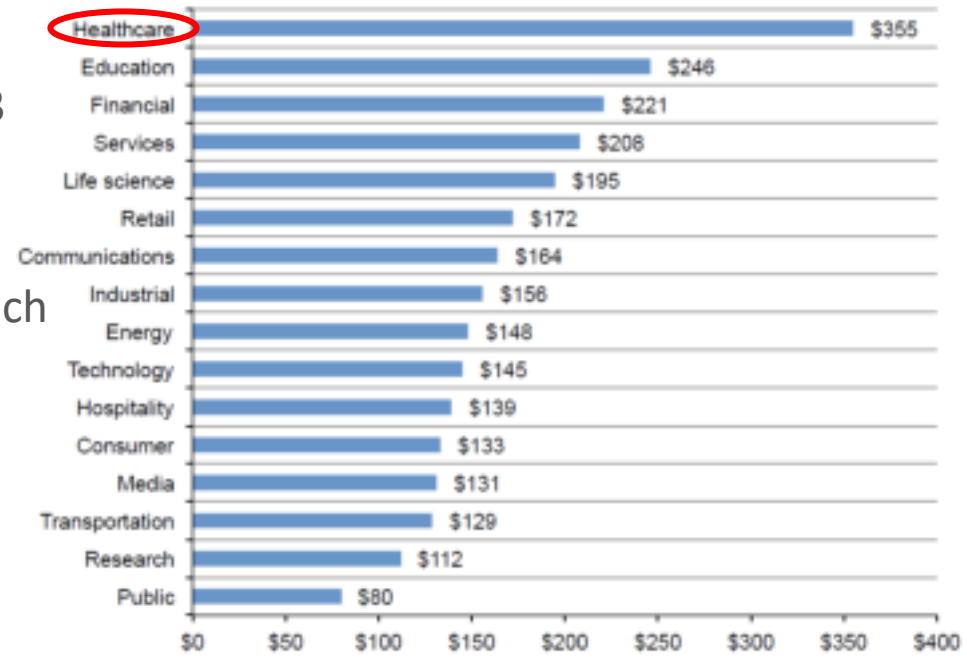
The cost of data breach (2016)

Study involving 383 companies in 12 countries

- $4 million is the average total cost of data breach
- 29% increase in total cost of data breach since 2013
- $158 is the average cost per lost or stolen record
- 15% percent increase in per capita cost since 2013

In Italy (24 companies participated):

- €2.35 million is the average total cost of data breach
- 17.1% increase in total cost of data breach
- €112 is the average cost per lost or stolen record
- 6.3% increase in cost per lost or stolen record



Privacy-by-Design and by Default (GDPR)

All healthcare management and analytical services/processes should be designed and implemented according to privacy-by-design principles

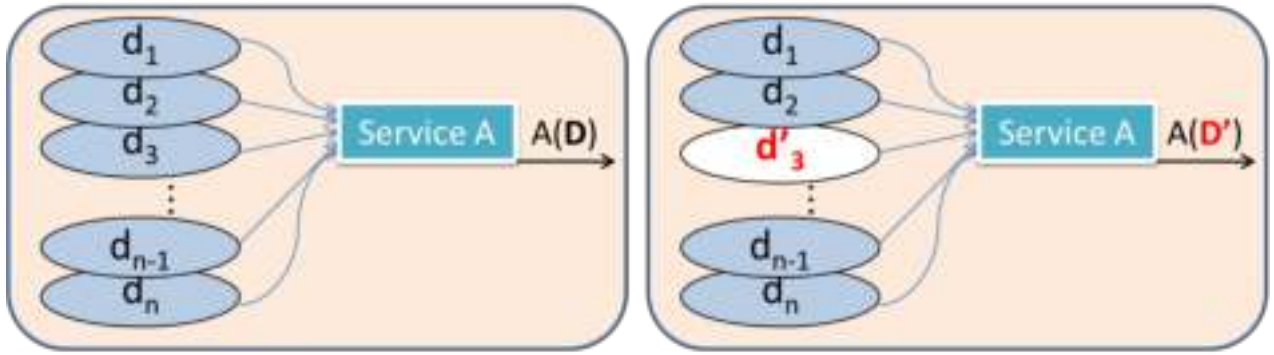
Cybersecurity measures are not sufficient to protect user privacy

- Anonymized research data could be linked to external data and disclose identities (e.g., 1997's Governor of Massachusetts reidentification from medical records)
- Health and clinical analytical services could be misused to perform privacy breach
- Genome data are extremely sensitive

How to manage/store/analyze data without putting patients' privacy at risk?

Future research perspectives (1)

Differentially-private health/clinical data management and analytics



Intuition: Users learn roughly the same thing about me whether or not my data is in the database

A randomized service A is ε-differentially private if

- for all databases **D**, **D'** that differ in one element
- for all sets of answers S

$$Pr[A(\mathbf{D}) \in S] / Pr[A(\mathbf{D'}) \in S] \leq e^{\varepsilon}$$

Examples of services:

- provide common genome properties in specific groups of patients without revealing the full genome profiles
- provide clinical profiles of specific groups of patients without revaling too specific patterns
- alert specific medical staff to emergencies based on geo-localization without disclosing the actual position
- …

Future research perspectives (2)

Provide users/staff/patients with feedbacks on their privacy

- Using machine learning techniques to predict the privacy risk of any process involving personal/sensitive data
- Provide users with clear cost-benefit feedback regarding the disclosure of their personal data

Examples of services:

- provide feedbacks to users involved in clinical studies
- identify potential new participants in studies based on data with privacy guarantees (information about data processing, exchange, outcomes)
- predict emergency/outbreak based on historical data with privacy guarantees
- …

Bibliography

- R.G. Pensa, G. Di Blasi. A privacy self-assessment framework for online social networks. Expert Systems with Applications. Vol. 86 2017. pp. 18-31.
- R.G. Pensa, L. Bioglio. Your Privacy, My Privacy? On Leakage Risk Assessment in Online Social Networks. In Proceedings of the 1st International Workshop on Personal Analytics and Privacy PAP 2017 (in conjunction with ECML PKDD 2017). September 18, 2017, Skopje, Macedonia. LNCS 10708, pp. 3-9.
- L. Bioglio, R.G. Pensa. Impact of Neighbors on the Privacy of Individuals in Online Social Networks. Procedia Computer Science, Vol. 108 2017 (Proceedings of the 2017 International Conference on Computational Science ICCS 2017, June 12-14, 2017, Zurich, Switzerland), pp. 28-37.
- L. Bioglio, R.G. Pensa. Modeling the Impact of Privacy on Information Diffusion in Social Networks. In Proceedings of the 8th Conference on Complex Networks CompleNet 2017, March 21-24, 2017, Dubrovnik, Croatia. pp. 95-107. Springer Proceedings in Complexity.
- R.G. Pensa, G. Di Blasi. A Centrality-based Measure of User Privacy in Online Social Networks. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM 2016, August 18-21, 2016, San Francisco, CA, USA. pp. 1438-1439.
- R.G. Pensa, G. Di Blasi. A Semi-supervised Approach to Measuring User Privacy in Online Social Networks. In Proceedings of the 19th International Conference on Discovery Science DS 2016, October 19-21, 2016, Bari, Italy. LNCS 9956, pp. 392-407.
- Monreale, D. Pedreschi, R.G. Pensa, F. Pinelli. Anonymity Preserving Sequential Pattern Mining. Artificial Intelligence and Law. Vol. 22(2) 2014. pp 141-173.
- Monreale, D. Pedreschi, R.G. Pensa. Anonymity Technologies for Privacy-Preserving Data Publishing and Mining. Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques, F. Bonchi, and E. Ferrari (Eds), pp. 3-33. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. 2010.
- R.G. Pensa, A. Monreale, F. Pinelli, D. Pedreschi. Anonymous Sequences from Trajectory Data. Proceedings of the 17th Italian Symposium on Advanced Database Systems, SEBD 2009, Camogli (GE), Italy, June 21-24, 2009.
- R.G. Pensa, A. Monreale, F. Pinelli, D. Pedreschi. Pattern-Preserving k-Anonymization of Sequences and its Application to Mobility Data Mining. Proceedings of the International Workshop on Privacy in Location-Based Applications PiLBA'08 in conjunction with ESORICS 2008. October 9, 2008, Malaga, Spain.