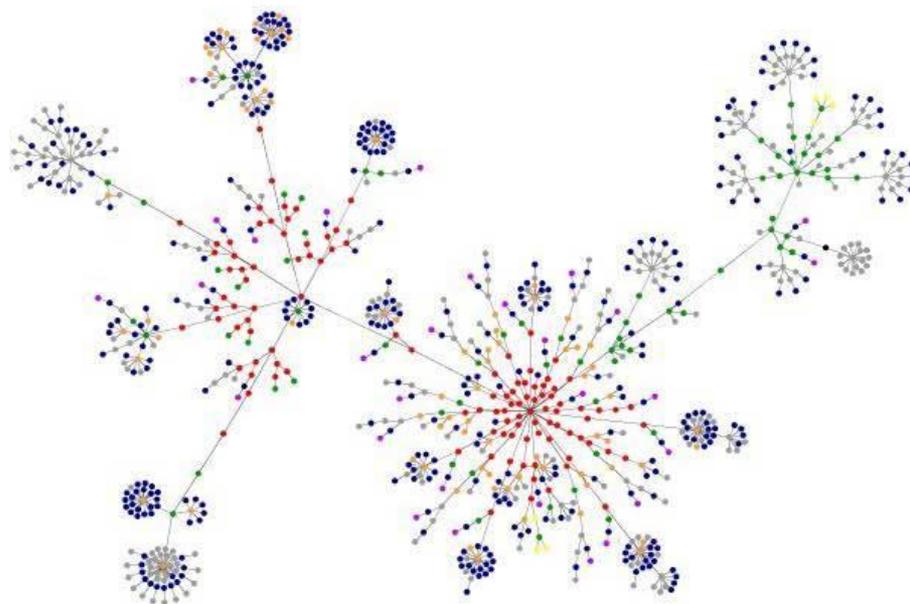


La costruzione dell'universo delle istituzioni e del personale nel settore dell'istruzione pubblica

Vincenzo Spinelli

17 Febbraio 2015



Fonte: wikipedia.it

Indice

1. Introduzione
2. Le scuole nell'indagine RCL 2012
3. Stima dei dipendenti
4. Stima delle istituzioni scolastiche
5. Conclusioni

Introduzione

- L'istruzione pubblica è stata introdotta nell'indagine SES 2006
- La stima dell'universo dei dipendenti e delle scuole (unità locali del MIUR) è basata su archivi amministrativi.
- L'universo considerato comprende (solo) le scuole statali: dalle scuole primarie alle scuole secondarie di II grado (università escluse).
- Gli archivi MIUR forniscono le posizioni lavorative dei dipendenti scolastici e l'anagrafe delle sedi.
- La ricostruzione delle istituzioni scolastiche (cluster di sedi) si basa su strutture formali (grafi).

Le scuole nell'indagine RCL 2012

Per la stima dei dipendenti ed unità locali vengono acquisite ed incrociate le seguenti fonti:

- Fonte **MIUR**
 1. Anagrafica dipendenti 2011-12 e 2012-13
 2. Anagrafica sedi scolastiche 2011-12 e 2012-13
- Fonte **MEF**
 1. Cedolini paga Gennaio – Dicembre 2012
 2. Anagrafica dipendenti Gennaio – Dicembre 2012
 3. Modello di dichiarazione 770 Anno di riferimento 2012 (**)
- Non vengono considerati i dipendenti e le scuole delle province autonome di **Trento**, **Bolzano** e regione autonoma della **Valle d'Aosta**.

Stima dei dipendenti

L'approccio considerato per stimare l'universo dei dipendenti nelle istituzioni scolastiche è descritto in:

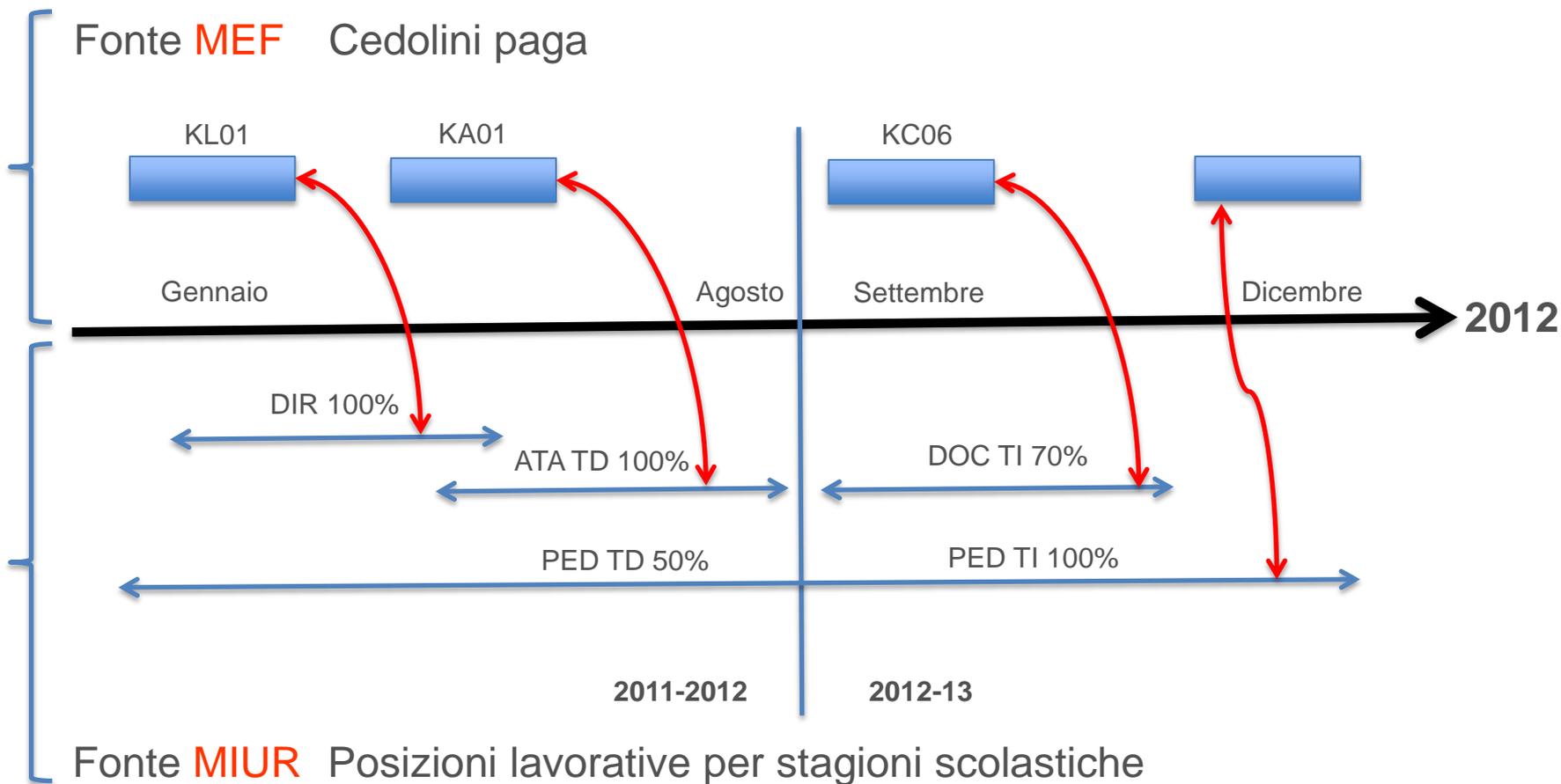
Il Registro dei dipendenti e delle unità locali dell'Istruzione Pubblica in Italia – anno 2010: quadro definitorio e metodologia di calcolo.

Vincenzo Spinelli – Istat Working Paper N. 9 - 2013

Fonti: *archivi amministrativi che descrivono posizioni lavorative e stipendi.*

Problema: *definire la lista dei dipendenti armonizzando le fonti (es. durata e codifiche).*

Stima dei dipendenti



Stima dei dipendenti

Passo 1: ETL degli archivi acquisiti da MEF e MIUR.

Passo 2: Adozione della stessa unità temporale negli archivi (mese vs. stagione scolastica).

Passo 3: Costruzione di alcuni indicatori binari (flag) per la valutazione del grado di similitudine degli archivi amministrativi.

Passo 4: Costruzione delle tabelle mensili del grado di omogeneità (frequenza dei casi riscontrati rispetto agli indicatori).

Passo 5: Scelta del filtro opportuno e cancellazione dei casi non corretti (funzione logica in formato DNF i cui atomi sono i flag)

Passo 6: Raffinamento degli archivi e stima dell'universo dei dipendenti.

Stima dei dipendenti

Aprile 2012 – tabella di raccordo MEF/MIUR (9/16 casi possibili)

Filtro 1	Filtro 2	Filtro 3	Filtro 4	# record	# dipendenti	
	y		y	1	1	
y	y		y	1	1	
	y	y		22	22	
		y		125	125	
	y			154	154	
			y	406	406	
y	y			3823	3822	f(x)
y	y	y		4005	4000	g(x)
				968568	968560	

Funzione filtro: $f(x) = (\text{filtro1}(x) \text{ AND } \text{filtro2}(x))$

$g(x) = (\text{filtro1}(x) \text{ AND } \text{filtro2}(x) \text{ AND } \text{filtro3}(x))$

Stima delle istituzioni scolastiche

L'approccio considerato per stimare l'universo dell'istituzioni scolastiche, come cluster di sedi, è descritto in:

Archivi amministrativi per le statistiche eurostat.

S. Cardinaleschi, V. Spinelli – Convegno SISVSP, 2012.

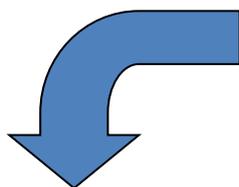
Fonti: lista delle sedi estratta dall'universo dei dipendenti stimato precedentemente.

Problema: definire le istituzioni scolastiche (unità locali) ottenute raggruppando le sedi (clustering).

Stima delle istituzioni scolastiche

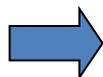
Input: lista delle sedi scolastiche attive (sintesi)

Sede	Root	Indirizzo	Comune	Provincia
BAAA18405V	BAEE18400V	Valenzano	Bari
BGSL010507	BGSL01000T	Bergamo	Bergamo
PGEE827044	PGIC82700V	Deruta	Perugia
CNAA81900Q	CNIC81900X	La Morra	Cuneo



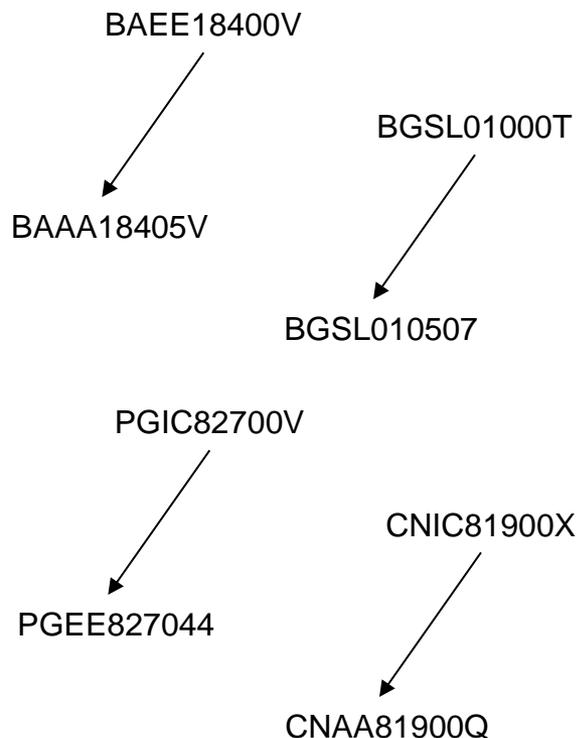
Output: lista delle istituzioni scolastiche (sintesi)

Sede	UL
BAAA18405V	UL007287
BGSL010507	UL001482



UL	N. sedi	N. Dip	FTE
UL007287	3	35	24.5
UL001482	5	9	7.6

Stima delle istituzioni scolastiche



Problema

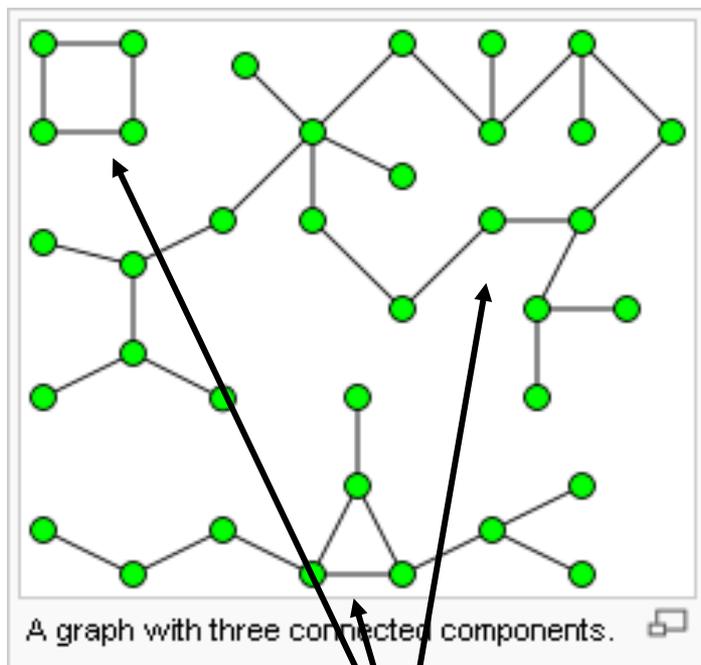
Sulla base delle relazioni gerarchiche definite nel registro delle sedi scolastiche, raggruppare le sedi in cluster che rappresentano le scuole.

In altre parole:

È possibile riconoscere/classificare le unità locali del MIUR, qualsiasi articolazione abbiano?

È possibile definire tecniche di controllo e correzione di queste unità?

Stima delle istituzioni scolastiche



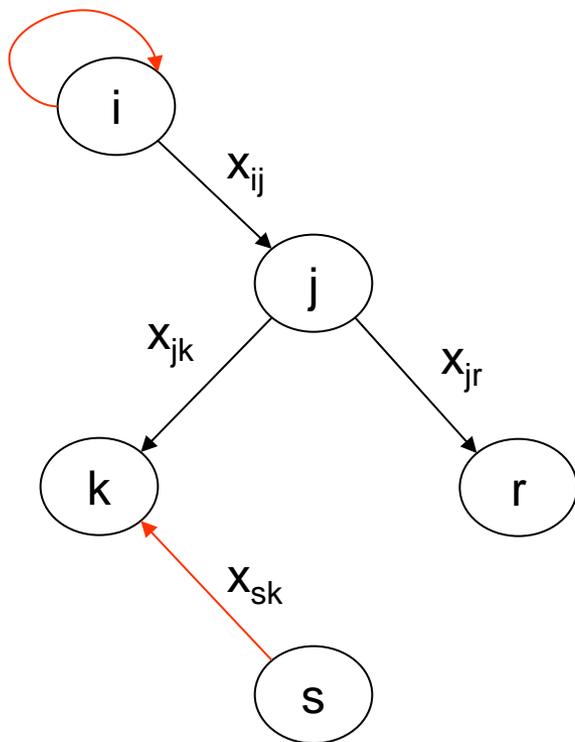
Unità locali?

Riconoscimento

Sia G_S il grafo delle sedi scolastiche (**privo di errori**), allora:

- Le istituzioni scolastiche sono tutte e sole le componenti connesse di G_S
- Ogni componente connessa coincide con il suo “spanning tree”
- La sede principale è unica nelle singole componenti (**grafo orientato**)

Stima delle istituzioni scolastiche



Controllo

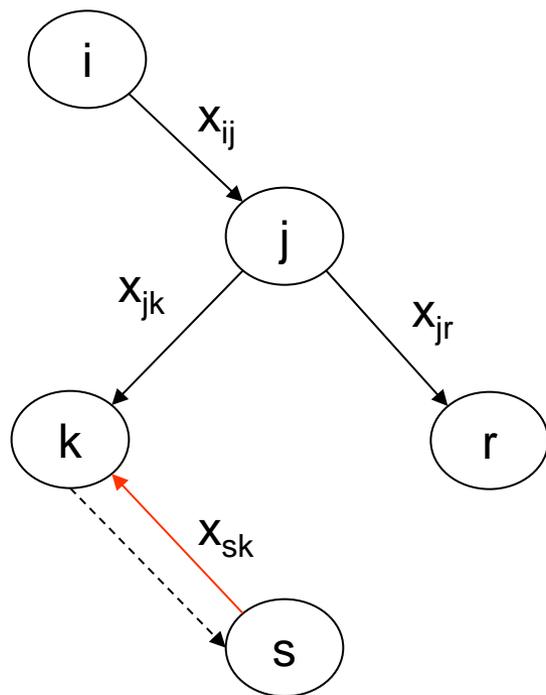
Definizione: $x_{ab}=1$ se $a \rightarrow b$, 0 else

Vincoli: $\forall a,b=1,nv$

- $x_{ab} \in \{0,1\}$ (Integer Problem)
- $x_{aa} = 0$ (no cappi)
- $x_{ab} + x_{ba} \leq 1$ (è una gerarchia)
- $\sum_{ab} x_{ab} = nv-1$ (regola NO CYCLE)
- $\sum_a x_{ab} \leq 1$ (regola FAN IN)

Stima delle istituzioni scolastiche

Correzione



Gerarchia di Problemi

Problema 1

$\min_{G_S} \{\# \text{ orientazioni modificate}\}$

Problema 2a

$\min_{G_S} \{\# \text{ spigoli cancellati}\}$

Problema 2b

$\min_{G_S} \{\# \text{ spigoli inseriti}\}$

Problema 3

$\min_{G_S} \{\# \text{ spigoli inseriti/cancellati}\}$

Stima delle istituzioni scolastiche

• Stagione 2011-12

- 33.825 sedi
- 22.825 spigoli

• di cui

- 11.000 cluster - 32,52% (scuole totali)
- 2.681 scuole con una sola sede (singleton)
- 14 – massimo numero di sedi per scuola



Effetti della
riforma
Gelmini....

• Stagione 2012-13

- 33.109 sedi
- 23.131 spigoli

• di cui

- 9.978 cluster - 30,14% (scuole totali)
- 2.212 scuole con una sola sede (singleton)
- 14 – massimo numero di sedi per scuola

Conclusioni

- La stima del costo del lavoro sull'universo dei dipendenti è una fase successiva al lavoro descritto.
- Il sistema di elaborazione si basa su una parte ingegnerizzata ed una variabile legata alle possibili variazioni nella fornitura dei dati.
- La definizione delle funzioni logiche per la individuazione dei casi anomali si basa su tecniche adattative (legate alla fornitura).
- Le fonti amministrative mostrano generalmente alcuni «effetti di memoria» a lungo termine, quindi le correzioni risultano più semplici con il progredire delle edizioni.
- La calcolabilità della soluzione proposta si basa su alcuni semplici risultati teorici non approfonditi nella presentazione.

