

Esempio: pendolarismo in Abruzzo

L'espressione mercato del lavoro viene utilizzata per indicare l'insieme dei meccanismi che regolano l'incontro tra i posti di lavoro vacanti e le persone in cerca di occupazione e che sottostanno alla formazione dei salari pagati dalle imprese ai lavoratori.

È possibile analizzare quali sono i fattori sociali e culturali che provocano un discostamento della domanda e dell'offerta di lavoro dalle teorie economiche pure.

In particolare l'offerta di lavoro è condizionata:

- Dalle aspirazioni professionali dei lavoratori. Esse a loro volta sono condizionate dai livelli di istruzione e dalla cultura del lavoro.
- Dalla capacità dei lavoratori di mobilitare relazioni (forti e deboli) e risorse sociali.
- Dal ruolo della famiglia; in particolare le differenze di genere nel processo di formazione dell'offerta sono fondamentali.
- Dai vincoli dovuti alle responsabilità extra-lavorative (in termini di orari, mobilità, etc).

Allo stesso modo la domanda di lavoro è condizionata:

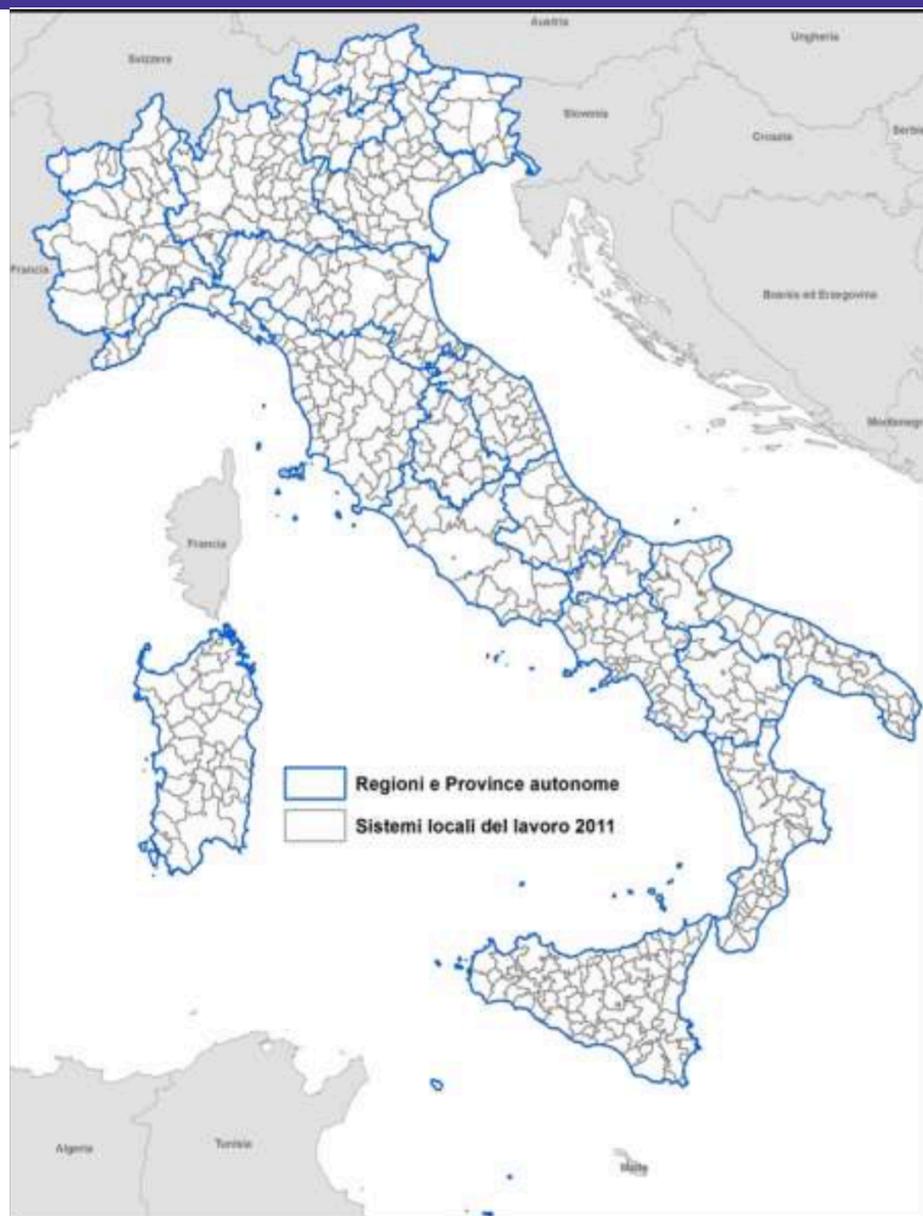
- Dal tipo di posizionamento competitivo delle imprese.
- Dalle strategie di reclutamento e gestione del personale scelte dalle imprese.
- Dalle rappresentazioni condivise in ordine al ruolo di specifiche categorie di lavoratori (per es. la segregazione di genere e la discriminazione razziale).

Esempio: pendolarismo in Abruzzo

L'incontro domanda/offerta di lavoro non avviene su un mercato impersonale e in corrispondenza di un salario d'equilibrio, ma è condizionato:

- Dall'esistenza di una pluralità di mercati della domanda e dell'offerta, in relazione alle caratteristiche delle società locali.
- Da fenomeni di mismatch occupazionale dovuti all'inadeguatezza delle qualifiche, alla selettività dell'offerta, all'incompatibilità delle culture del lavoro, ecc.
- Dall'esistenza di **mercati locali del lavoro** o “zone specifiche” dell'economia.
- Dal ritorno in auge del cosiddetto mercato della vita in cui si scambia non solo capacità lavorativa anonima, qualificata da capacità acquisite di ordine professionale, ma l'intera personalità del lavoratore con tutte le sue caratteristiche ascritte.
- Dalla regolazione giuridica del mercato del lavoro e dalla possibilità di eluderla.
- Dalle politiche del lavoro e dell'occupazione (nazionali e/o locali).
- Dai sistemi di relazioni industriali.
- Dall'intervento di istituzioni di sostegno all'incontro domanda/offerta (centri per l'impiego, agenzie private d'intermediazione, uffici di collocamento).
- Dalla persistenza di segmenti dell'economia regolati dalla tradizione e da vincoli familistici e comunitari.
- Dall'operare di organizzazioni illegali quali ad es. il caporalato.

Esempio: pendolarismo in Abruzzo



Input data:

- **Commuting matrix** from 2011
Population census

LMAs released date:

- 17 December 2014

Tool:

- Script at the base of the **R**
package **LabourMarketAreas**

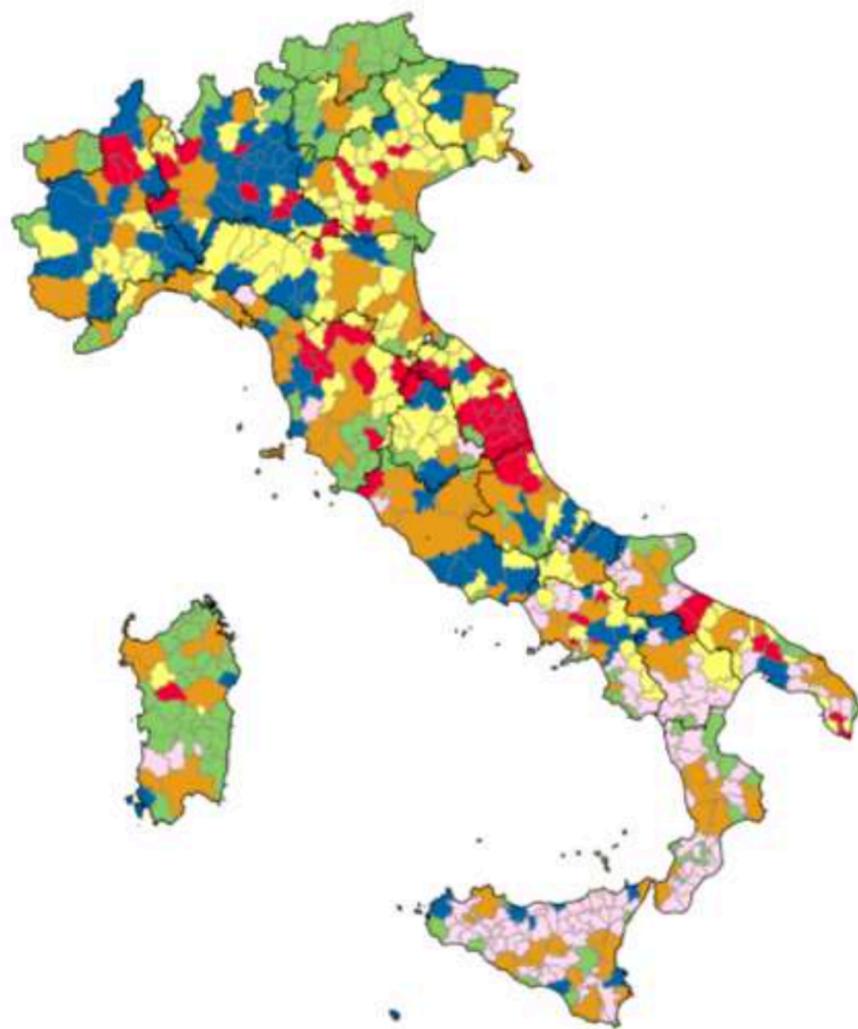
Total number of areas:

- **611 LMAs** in Italy in 2011
(there were 683 in 2001).

Esempio: pendolarismo in Abruzzo

Prevailing production specialisation

Six classes have been identified to classify the Italian productive system



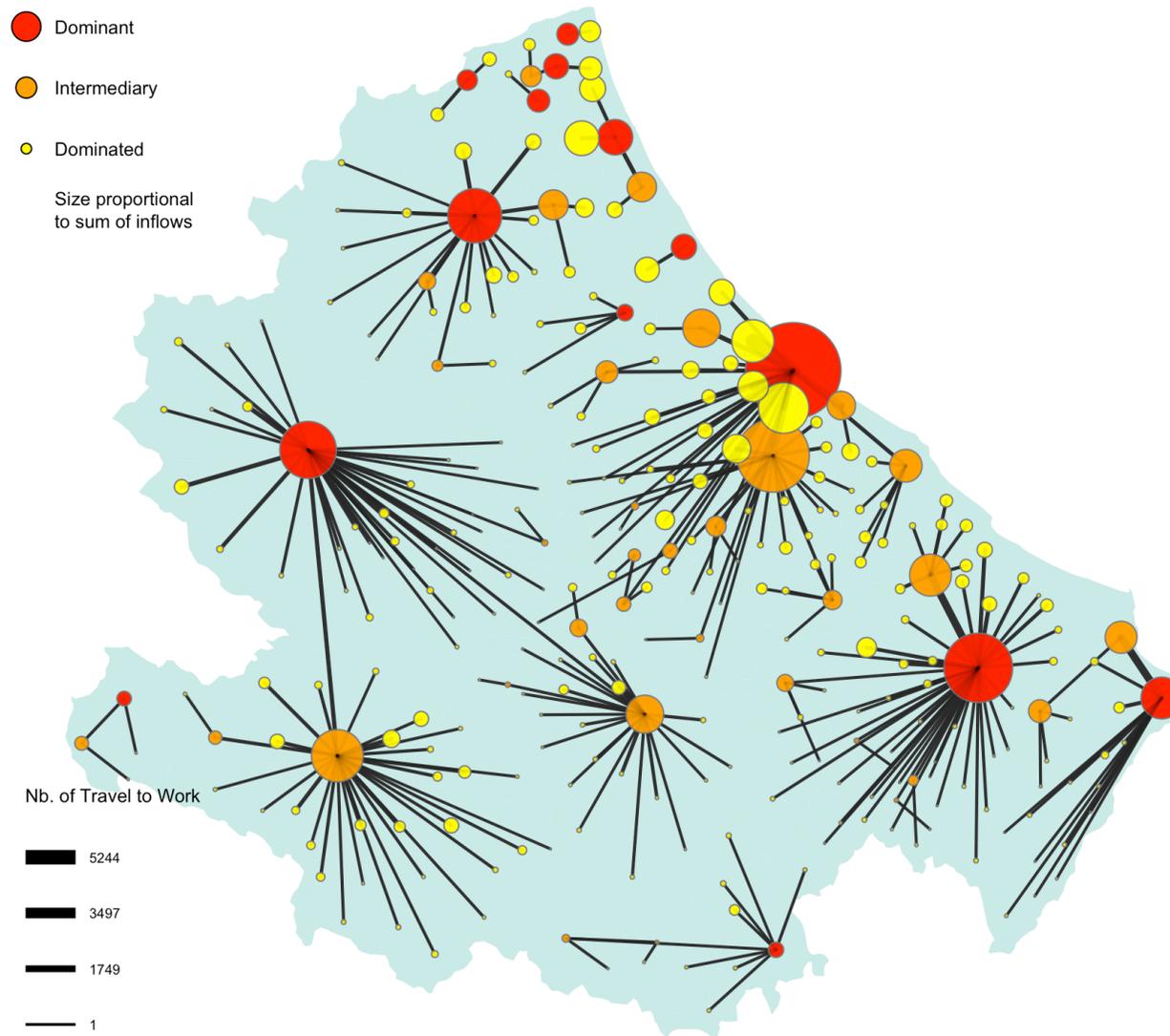
- Texile, clothing and leather (60)
- Other firms of the «made in Italy» (129)
- Heavy manufacturing (85)
- Urban LMA (91)
- Non manufacturing (133)
- Unspecialised (113)

Esempio: pendolarismo in Abruzzo

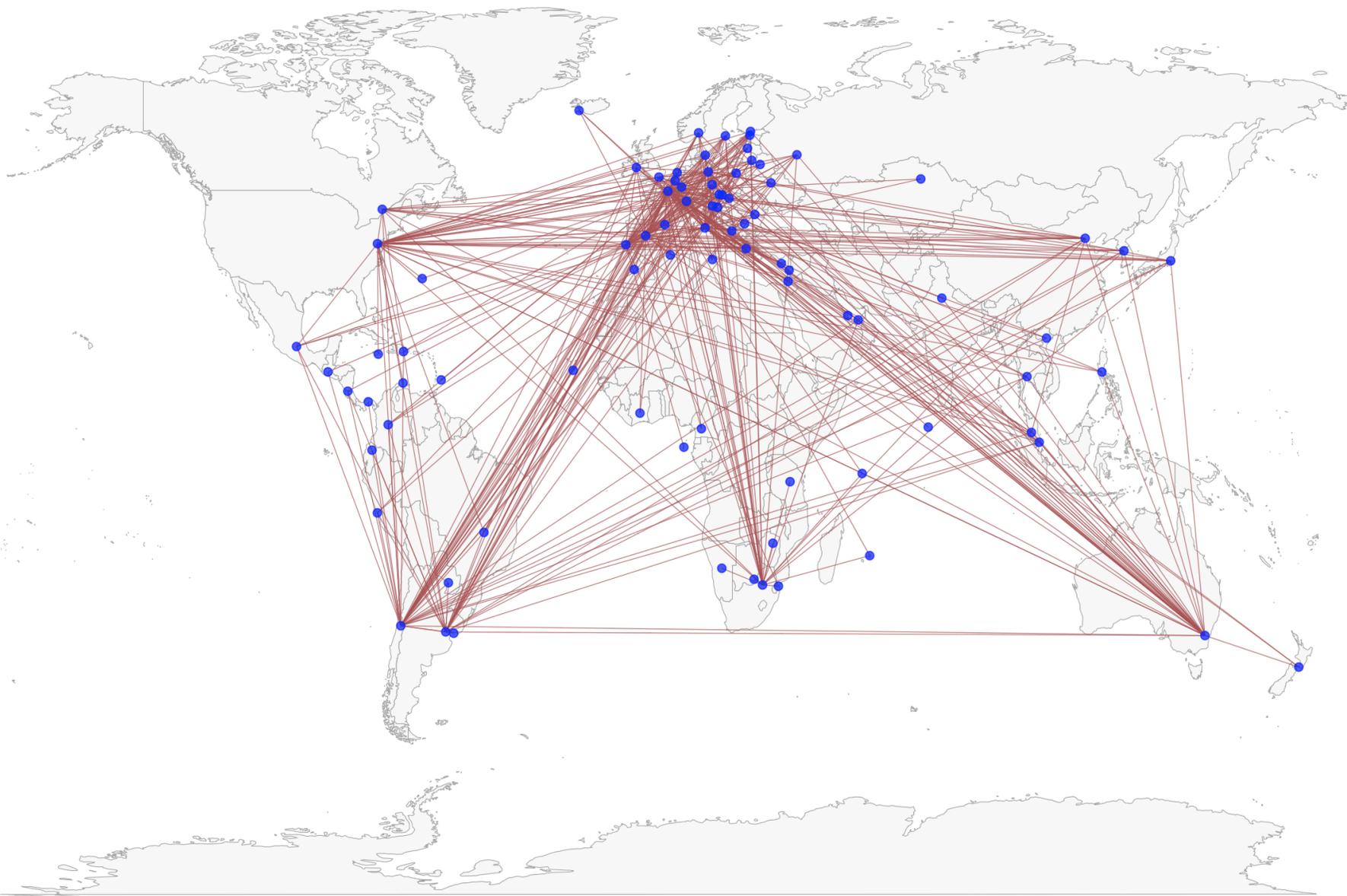
```
setwd("~/Documents/prova r pkg/MATRICE PENDOLARISMO 2011")
datipend <- read.table("matrix_pendo2011_10112014.txt",dec=".",header = T,na.strings = "ND")
datipend <- datipend[datipend$tipo == "S",]
datipend <- datipend[datipend$provres <= 69 & datipend$provres >= 66 & datipend$provdes <= 69
& datipend$provdes >= 66,]
datipend <- datipend[datipend$motivo == 2,]
datipend$codcomres <- datipend$provres * 1000 + datipend$comres
datipend$codcomdes <- datipend$provdes * 1000 + datipend$comdes
sumpend <- aggregate(datipend$stiflusso,by=list(codcomres = datipend$codcomres,codcomdes =
datipend$codcomdes),sum,na.rm = T)
library(maptools)
comabr <- readShapePoly("/Users/utente/Desktop/Dati Archivio/Limiti
Amministrativi/comuni2008/com2008_s.shp")
comabr <- subset(comabr,COD_REG == 13)
library(flows)
myflows <- prepflows(mat = sumpend, i = "codcomres", j = "codcomdes", fij = "x")
mystats <- statmat(mat = myflows, output = "all", verbose = TRUE)
diag(myflows) <- 0
flowSel1 <- domflows(mat = myflows, w = colSums(myflows), k = 1)
flowSel2 <- firstflows(mat = myflows, method = "nfirst", ties.method = "first",k = 1)
flowSel <- myflows * flowSel1 * flowSel2
inflows <- data.frame(id = colnames(myflows), w = colSums(myflows))
opar <- par(mar = c(0,0,1,0))
sp::plot(comabr, col = "#ccea7", border = NA)
plotMapDomFlows(mat = flowSel, spdf = comabr, spdfid = "PRO_COM", w = inflows, wid = "id",
wvar = "w", wcex = 0.05, add = TRUE, legend.flows.pos = "bottomleft",
legend.flows.title = "Nb. of Travel to Work")
title("Dominant Travel to Work Flows - ISTAT, 2011")
par(opar)
```

Esempio: pendolarismo in Abruzzo

Flussi Dominanti - ISTAT, 2011

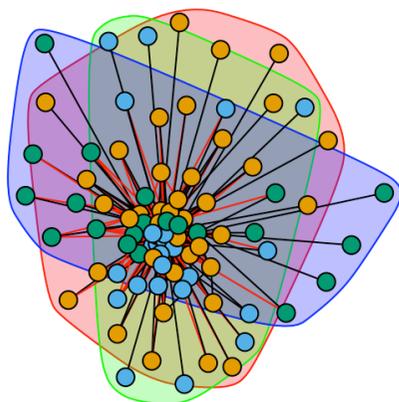


Esempio: commercio estero del vino

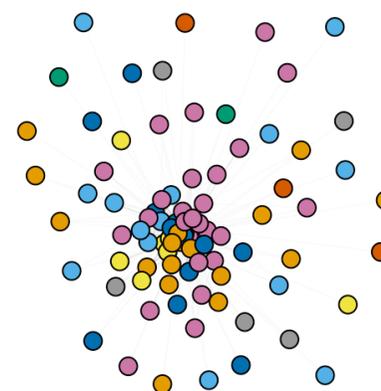


Esempio: commercio estero del vino

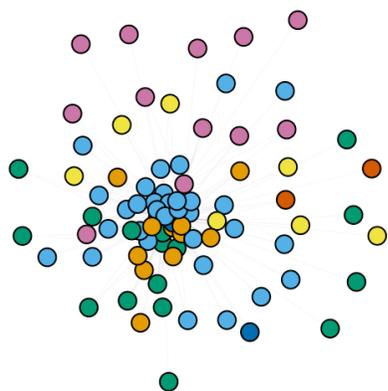
Cluster Fast Greedy ITN Plot



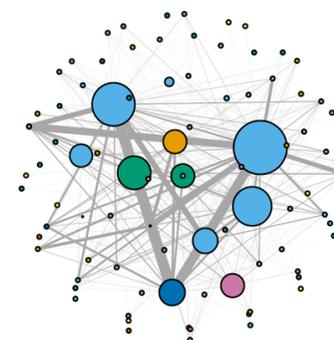
Spinglass ITN Plot



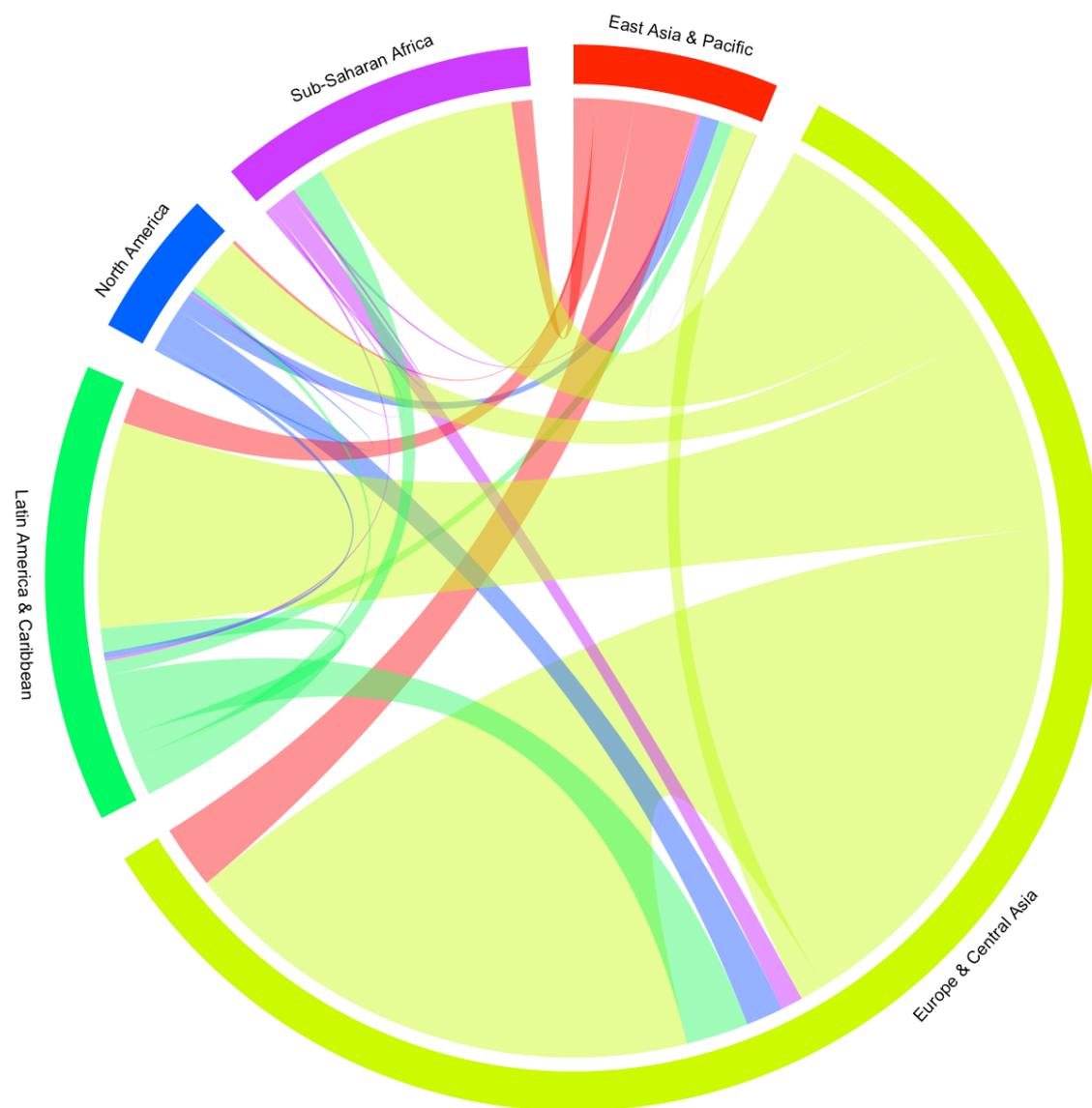
Region ITN Plot



Region ITN Plot - Node Size Outdegree



Esempio: commercio estero del vino



Esempio: commercio estero del vino

	CLUS	Weighted.Degree.All	Binary.Degree.All	Betweenness	Closeness	Eigenvector	Hub	Authority
ARG	1	3.0156	29	1	0.1517	0.1085	0.6010	0.1011
AUS	2	7.0511	29	22	0.1638	0.2751	0.6336	0.7873
CHL	3	6.8541	41	1	0.1165	0.1717	0.7840	0.0775
CHN	2	1.8598	2	6	0.0708	0.3564	0.0370	0.9822
DEU	1	3.2024	31	231	0.0984	0.5079	0.6714	0.9062
ESP	1	10.0353	48	103	0.0557	0.3282	0.8789	0.5124
FRA	2	33.6984	66	657	0.0872	1.0000	1.0000	0.8308
GRC	1	0.1999	6	0	0.0839	0.0143	0.1382	0.3617
HUN	1	0.2911	11	0	0.1349	0.0075	0.2559	0.2059
ITA	1	22.7949	53	135	0.0703	0.8452	0.9250	0.5711
PRT	1	2.4803	28	33	0.1103	0.0921	0.5922	0.3617
ROU	1	0.0272	2	0	0.1078	0.0065	0.0558	0.4813
RUS	1	0.0000	0	0	0.0595	0.0857	0.0000	0.8580
USA	1	4.6274	32	511	0.0949	0.9528	0.6509	0.9088
ZAF	3	2.3469	29	44	0.1288	0.0628	0.5654	0.1290

```
ITNplotset(wineITN)
ITN_make_plot(wineITN,TRUE,TRUE)
ITNdegdist(wineITN)
ITNhistdegdist(wineITN)
par(mfrow=c(1,1),mar=c(1,1,1,1))
region_circle_plot(wineITN)
ITN_map_plot(wineITN)
CLUwine<-ITNcluster(wineITN)
CENTwine<-ITNcentrality(wineITN)
cbind(CLUwine$cluster.fast.greedy.mem,CENTwine[,-c(1,2:5)])
```

Le tavole Intersettoriali

- Le tavole input-output sono strumenti di analisi che traggono la loro origine dai modelli matematici dello statistico-economico russo Leontiev (1929) e poi da Stone e da Sraffa e Pasinetti.
- Input-output è un modello disaggregato (non Keynes = modello aggregato)
- Lo strumento delle tavole input-output consente di contabilizzare ed esprimere sotto forma di modello matematico i rapporti che legano i vari settori di un sistema economico con l'obiettivo di spiegare e prevedere l'andamento delle grandezze macroeconomiche.

Le tavole Intersettoriali

destinazione	1	2	3	Totale domanda intermedia	Flussi finali	Produzione
origine						
1	x_{11}	x_{12}	x_{13}	x_{10}	F_1	X_1
2	x_{21}	x_{22}	x_{23}	x_{20}	F_2	X_2
3	x_{31}	x_{32}	x_{33}	x_{30}	F_3	X_3
Totale costi intermedi	x_{01}	x_{02}	x_{03}			
Flussi primari	Y_1	Y_2	Y_3			
Produzione	X_1	X_2	X_3			

La tavola input-output è composta di tre parti: quella relativa ai *flussi intermedi* (riportati nella parte centrale della tavola), quella relativa ai *flussi primari* e quella relativa ai *flussi di domanda finale*.

Le tavole Intersectoriali

Flussi primari :

- Valore Aggiunto
- Trasferimenti
- Imposte
- Importazioni

Flussi finali:

- Consumi delle famiglie
- Consumi pubblici
- Investimenti
- Esportazioni
- Variazioni delle scorte

Le tavole Intersettoriali

Le k equazioni di bilancio sono definite dall'uguaglianza: $\mathbf{x}\mathbf{i} + \mathbf{F} = \mathbf{X}$

dove:

$$\mathbf{x}_{j0} = \begin{bmatrix} x_{10} \\ x_{20} \\ x_{30} \end{bmatrix} = \mathbf{x}\mathbf{i}$$

le equazioni di bilancio possono essere riscritte come: $\mathbf{x}_{j0} + \mathbf{F} = \mathbf{X}$

Le k equazioni dei costi sono, invece, date dalla seguente uguaglianza: $\mathbf{x}^t\mathbf{i} + \mathbf{Y} = \mathbf{X}$

dove:

$$\mathbf{x}_{0i} = \begin{bmatrix} x_{01} \\ x_{02} \\ x_{03} \end{bmatrix} = \mathbf{x}^t\mathbf{i}$$

Le tavole Intersettoriali

Coefficienti di spesa e coefficienti tecnici

Consideriamo i flussi intermedi x_{ji} di una generica colonna i e la produzione vendibile della branca i -esima X_i ; allora è possibile ottenere la matrice dei coefficienti di spesa come:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \frac{x_{11}}{X_1} & \frac{x_{12}}{X_2} & \frac{x_{13}}{X_3} \\ \frac{x_{21}}{X_1} & \frac{x_{22}}{X_2} & \frac{x_{23}}{X_3} \\ \frac{x_{31}}{X_1} & \frac{x_{32}}{X_2} & \frac{x_{33}}{X_3} \end{bmatrix}$$

I **coefficienti di spesa** esprimono l'ammontare dei beni e servizi prodotti dalla branca j -esima che vengono impiegati per la produzione di una unità monetaria di produzione della branca i -esima e costituiscono, pertanto, una misura del grado di dipendenza della branca i -esima dalla branca j -esima.

Le tavole Intersettoriali

Il modello domanda finale-produzione

Un primo modello utile nella programmazione economica è costruito a partire dalla equazione di bilancio che considerata la *j-esima* branca risulta pari a:

$$X_j = \sum_{i=1}^k x_{ji} + F_j$$

Questa espressione può essere riscritta in termini matriciali come:

$$\mathbf{AX} + \mathbf{F} = \mathbf{X}$$

Esplicitando rispetto a \mathbf{X} si ottiene:

$$\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{F}$$

Matrice di Leontief



Le tavole Intersettoriali

Da quest'ultima espressione emerge come una variazione della domanda finale di una branca determini effetti anche sulle altre branche tranne nel caso banale in cui la matrice dei coefficienti di spesa sia una matrice diagonale ovvero tutti i settori siano autosufficienti.

Da un punto di vista analitico il processo economico sottostante può essere rappresentato dal seguente sviluppo in serie:

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \dots$$

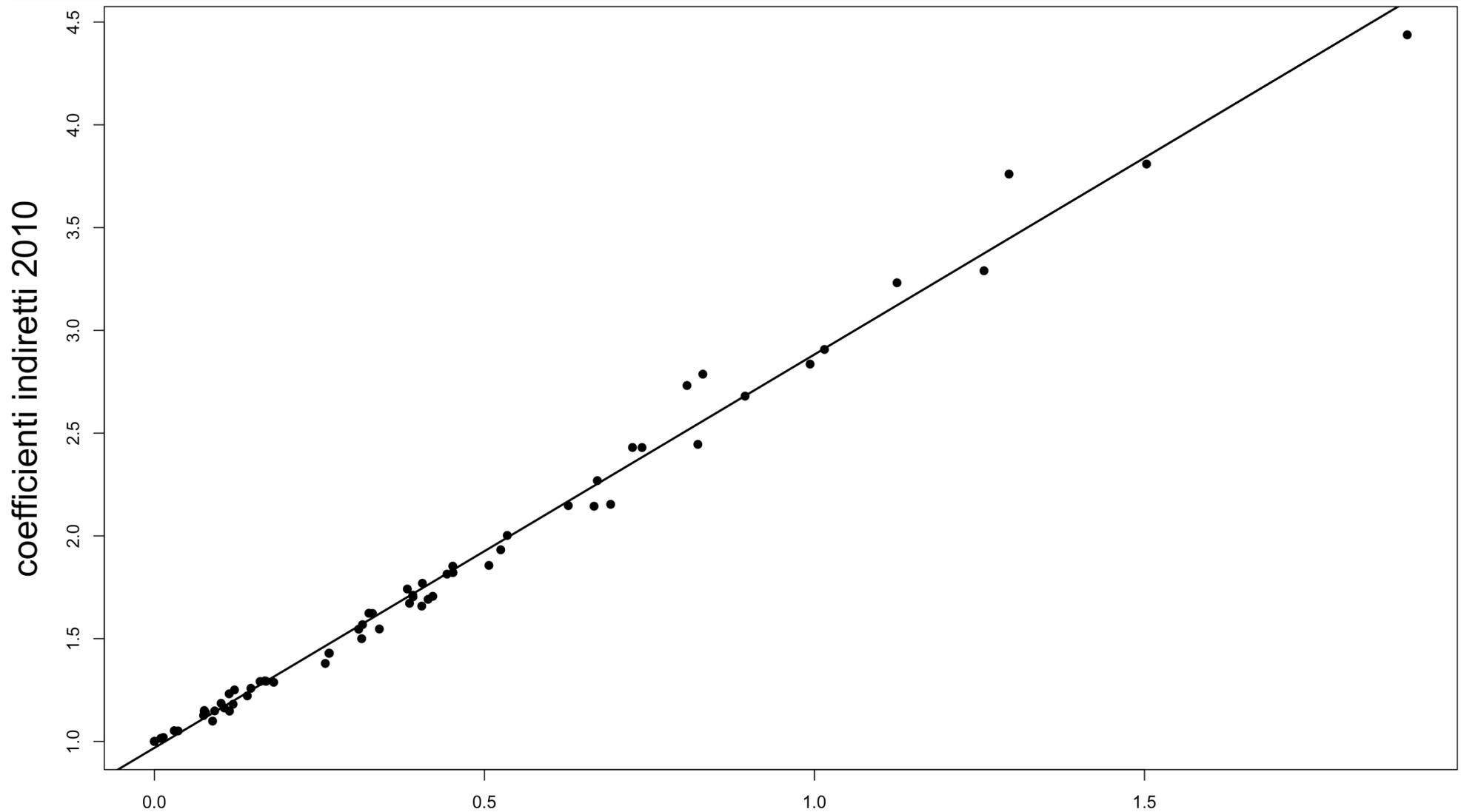
A fronte di un aumento della domanda in un settore, inizialmente si ha un aumento della produzione nello stesso settore per un pari importo, in seguito, al fine di realizzare la nuova produzione richiesta, è necessario un livello superiore della domanda intermedia, il che provoca un ulteriore aumento della produzione negli altri settori. Coefficienti indiretti sono i totali della matrice di Leontief:

$$\Delta \mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \Delta \mathbf{F}$$

Esempio: tavola IO Italia 2010

```
library(iotables)
io_tables <- iotables_download ( source = "naio_10_cp1750" )
na <- io_tables %>% dplyr::filter ( geo %in% "IT" ) %>% dplyr::filter ( year %in%
2010 ) %>% dplyr::filter ( unit %in% "MIO_EUR" ) %>% dplyr::filter ( stk_flow %in%
"DOM" )
an <- as.data.frame(na$data)
ITtable <- iotable_get( source = "naio_10_cp1750", geo = 'IT', year = 2010, unit =
"MIO_EUR", stk_flow = "DOM", labelling = "iotables" )
df <- as.data.frame(ITtable)
cc <- t(t(as.matrix(ITtable[1:65,2:66]))/as.vector(as.matrix(ITtable[75,2:66])))
inp <- input_coefficient_matrix_create(data_table=ITtable)
leo <- leontieff_inverse_create(inp)
com <- input_indicator_create(ITtable, 'compensation_employees')
ind <- rowSums(as.data.frame(leo[,2:66]))
dir <- rowSums(as.data.frame(inp[,2:66]))
plot(dir, ind, cex=1, pch=19, xlab="coefficienti diretti 2010", ylab="coefficienti
indiretti 2010", cex.lab=2)
names(ind) <- colnames(leo[,2:66])
names(dir) <- colnames(leo[,2:66])
rcoe <- lm(ind ~ dir)
abline(rcoe, lwd=2)
cbind(dir, ind)[order(-ind), ]
com <- input_indicator_create( data_table = ITtable, input_vector = c("gva"), digits
= 4, indicator_names = c("GVA indicator") )
ind2 <- indirect_effects_create(input_requirements = com, inverse = leo)
cbind(dir, ind, t(ind2[2:length(ind2)]))[order(-t(ind2[2:length(ind2)])), ]
```

Esempio: tavola IO Italia 2010



(Intercept) 0.96853 slope 1.91445 R-squared 0.99030

Esempio: tavola IO Italia 2010

	Coef. Dir.	Coef. Ind
wholesale_trade	1.898078839	4.437794
legal_accounting_consulting	1.503148358	3.809388
electricity_gas_steam	1.294731051	3.760269
financial_services	1.256864348	3.289937
land_transport	1.125096896	3.231397
security_investigation	1.015146598	2.906861
real_estate_services_b	0.993352176	2.835719
construction	0.830914506	2.786790
warehousing	0.806879578	2.731605
fabricated_metal	0.894887098	2.679964
coke_refined_petrol	0.823354823	2.445244
computer_programing_consulting	0.724302565	2.430125
sewage	0.738772975	2.429954
architectural_engineering	0.671040137	2.268627
auxiliary_financial_services	0.691254726	2.153458
basic_metals	0.627034747	2.147221
chemical	0.666072333	2.144218
telecommunications	0.534591963	2.002190

Esempio: tavola IO Italia 2010

	C. Dir	C. Ind	GVA indicator ind
coke_refined_petrol	0.823354823	2.445244	5.8304207
travel_agency_services	0.180759565	1.287602	4.3583854
advertising_marketing	0.324967673	1.623569	2.7274882
food_beverages_tobacco	0.524686296	1.932237	2.7121005
basic_metals	0.627034747	2.147221	2.6585027
air_transport	0.121332493	1.250535	2.3816309
motor_vechicles	0.391786724	1.703342	2.2719452
insurance	0.105888006	1.162847	2.1904387
water_transport	0.075553311	1.150395	2.1798149
paper_products	0.452022627	1.852711	2.0810613
sewage	0.738772975	2.429954	2.0035548

Esempio: Ranking sport

Un sistema di valutazione sportiva è un sistema che analizza i risultati delle competizioni sportive per fornire punteggi per ogni squadra o giocatore. I sistemi comuni includono sondaggi di elettori esperti, mercato delle scommesse e sistemi informatici. Le classificazioni sono rappresentazioni numeriche della forza competitiva, spesso direttamente confrontabili in modo da poter prevedere il risultato del gioco tra due squadre qualsiasi.

I sistemi automatici possono tendere verso l'obiettività, senza preconcetti specifici per giocatore, squadra, regione o stile. Un loro vantaggio è che sono verificabili, ripetibili, complete e richiedono la valutazione di tutti i criteri selezionati. In confronto, i sistemi di classificazione che fanno affidamento sui sondaggi umani includono la soggettività umana intrinseca; questo può o non può essere una proprietà attraente a seconda delle esigenze del sistema.

Esempio: Ranking sport

I sistemi di valutazione dello sport utilizzano una varietà di metodi per classificare i contendenti, ma il metodo più diffuso è denominato power rating. La potenza nominale di una squadra o di un giocatore è un calcolo della forza della squadra rispetto ad altre squadre della stessa lega o divisione. L'idea di base è di massimizzare la quantità di relazioni transitive in un dato insieme di dati a causa dei risultati di gioco. Ad esempio, se A sconfigge B e B sconfigge C, allora si può tranquillamente affermare che $A > B > C$.

Ci sono ovvi problemi nel basare un sistema esclusivamente su vittorie e perdite. Ad esempio, se C sconfigge A, viene stabilita una relazione intransitiva ($A > B > C > A$) e si verificherà una violazione della classifica se questo è l'unico dato disponibile. Scenari come questo avvengono abbastanza regolarmente negli sport.

Inoltre la matrice dei risultati è spesso non piena (dati mancanti)

Esempio: Serie A 1934-2016

	0	1	2	3	4	5	6	7	8	Tot
0	11.297	6.109	3.401	1.169	0.398	0.102	0.031	0.016	0.008	22.532
1	12.443	13.175	4.468	1.952	0.543	0.201	0.024	0.016	0.000	32.822
2	8.089	8.955	4.940	1.193	0.417	0.102	0.051	0.004	0.004	23.756
3	4.287	4.629	2.181	0.807	0.173	0.071	0.020	0.008	0.000	12.175
4	1.941	2.043	0.996	0.390	0.087	0.028	0.000	0.004	0.000	5.487
5	0.744	0.807	0.362	0.142	0.031	0.004	0.000	0.000	0.000	2.090
6	0.228	0.268	0.150	0.075	0.016	0.008	0.000	0.000	0.000	0.744
7	0.083	0.106	0.047	0.016	0.000	0.000	0.000	0.000	0.000	0.252
8	0.047	0.016	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.079
9	0.016	0.031	0.008	0.004	0.000	0.000	0.000	0.000	0.000	0.059
10	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
Tot	39.179	36.140	16.568	5.747	1.665	0.516	0.126	0.047	0.012	100.000

Esempio: Serie A 1934-2016

$$X_{i,j} \sim \text{Poisson}(\alpha_i \beta_j \gamma_h),$$

$$Y_{i,j} \sim \text{Poisson}(\alpha_j \beta_i),$$

dove $X_{i,j}$ e $Y_{i,j}$ sono indipendenti e $\alpha_i, \beta_i > 0 \forall i$. L'abilità di attacco è misurata da α_i , mentre β_i e $\gamma_h > 0$ misurano rispettivamente l'abilità di attacco, difesa ed il vantaggio di giocare in casa.

Segue che con n squadre, $(\alpha_1, \dots, \alpha_n)$ i parametri di attacco, $(\beta_1, \dots, \beta_n)$ i parametri di difesa e il parametro dell'effetto home γ_h , devono essere stimati. Poiché il modello avrà troppi parametri, è necessario imporre il seguente vincolo:

$$n^{-1} \sum_{i=1}^n \alpha_i = 1,$$

Con le corrispondenze indicizzate $k = 1, \dots, N$ e i punteggi corrispondenti (x_k, y_k) la verosimiglianza prende la forma:

$$L(\alpha_i, \beta_i, \gamma_i; i = 1, \dots, n) = \prod_{k=1}^N \exp(-\lambda_k) \lambda_k^{x_k} \exp(-\mu_k) \mu_k^{y_k},$$

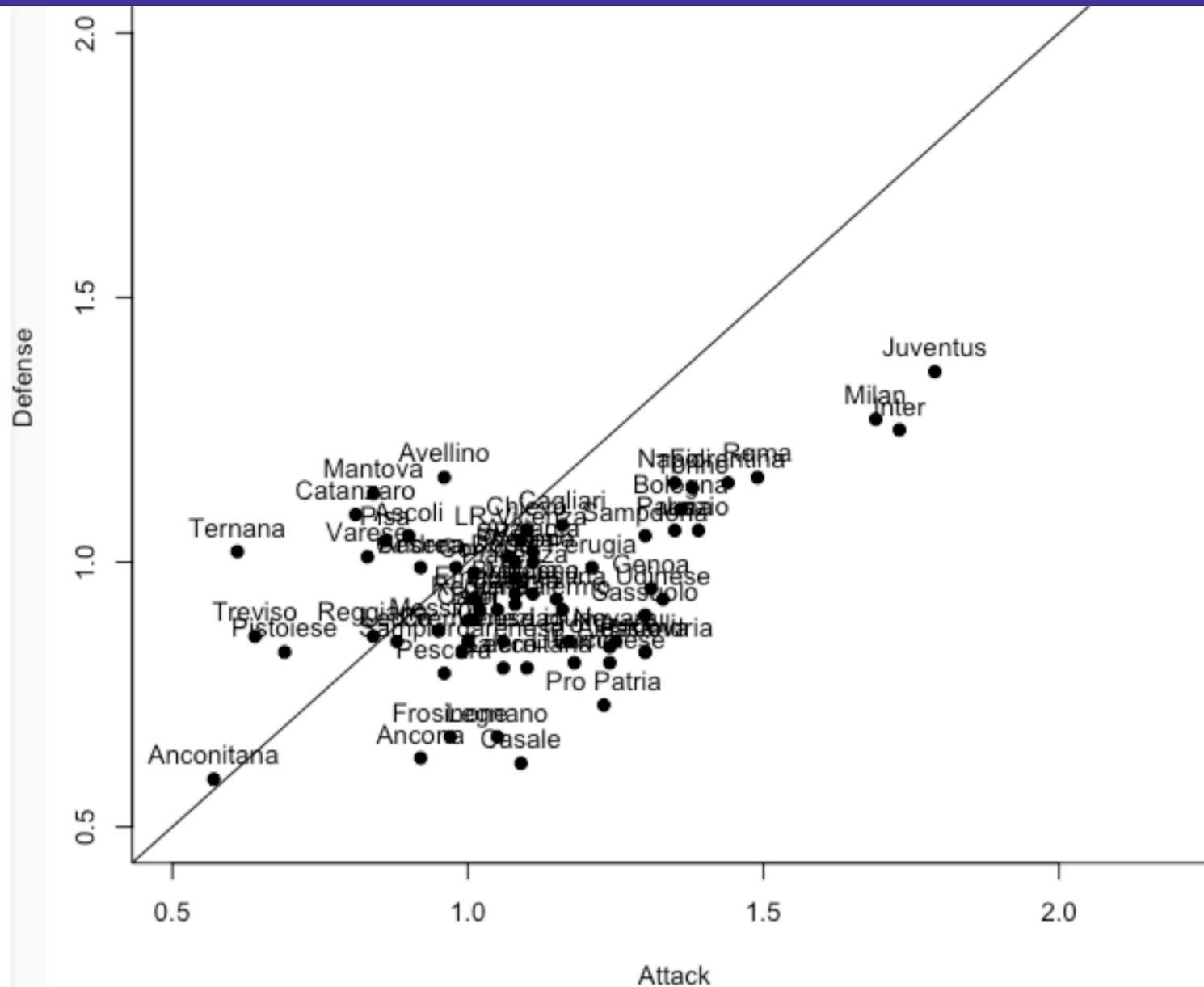
dove:

$$\lambda_k = \alpha_{i(k)} \beta_{j(k)} \gamma_h,$$

$$\mu_k = \alpha_{j(k)} \beta_{i(k)},$$

mentre $i(k)$ e $j(k)$ misurano gli indici della squadra di casa e di trasferta che giocano la partita k .

Esempio: Serie A 1934-2016



Esempio: Serie A 1934-2016

Rank	Team Name	Data 1934-2016				Data 2011-2016				Data 2015-2016			
		score	att.	def.	games	score	att.	def.	games	score	att.	def.	games
1	Juventus	1.28	1.79	1.36	2808	1.96	1.82	2.14	190	1.91	1.74	2.16	38
2	Inter	1.11	1.73	1.25	2847	0.56	1.45	1.02	190	0.46	1.18	1.17	38
3	Milan	1.10	1.69	1.27	2786	0.79	1.54	1.12	190	0.26	1.16	1.03	38
4	Roma	0.78	1.49	1.16	2802	0.99	1.73	1.15	190	1.04	1.97	1.04	38
5	Fiorentina	0.72	1.44	1.15	2626	0.72	1.50	1.10	190	0.57	1.42	1.04	38
6	Torino	0.65	1.38	1.14	2419	0.33	1.31	0.96	152	0.01	1.25	0.80	38
7	Napoli	0.63	1.35	1.15	2348	1.09	1.86	1.15	190	1.33	1.88	1.34	38
8	Bologna	0.58	1.36	1.10	2330	-0.09	0.94	0.99	152	-0.35	0.78	1.00	38
9	Lazio	0.57	1.39	1.06	2489	0.59	1.45	1.04	190	0.08	1.25	0.85	38
10	Parma	0.53	1.35	1.06	856	0.11	1.22	0.89	152	-	-	-	-
11	Sampdoria	0.45	1.30	1.05	2008	0.13	1.20	0.91	152	-0.24	1.16	0.73	38
12	Cagliari	0.32	1.16	1.07	1216	-0.12	1.04	0.88	152	-	-	-	-
13	Genoa	0.32	1.31	0.95	1690	0.15	1.21	0.92	190	-0.01	1.07	0.93	38
14	Udinese	0.31	1.33	0.93	1486	0.21	1.20	0.96	190	-0.66	0.85	0.75	38

Modelli Grafici

Un modello grafico consiste in una rappresentazione di una distribuzione di probabilità congiunta in termini di Grafi e di corrispondenti funzioni definite su tali grafi.

Ogni grafo sottende un set di relazioni condizionatamente dipendenti appartenenti alle variabili casuali presenti, che permettono alla distribuzione congiunta di tali variabili di essere scomposta (fattorizzata).

Vi sono due classi principali di modelli grafici, i *modelli grafici diretti* ed i *modelli grafici indiretti*, basati su grafi aciclici diretti e su grafi indiretti, rispettivamente.

Modelli grafici diretti

Sia X_1, \dots, X_N un set di N variabili casuali. Un modello grafico diretto permette la decomposizione della distribuzione congiunta $P(X_1, \dots, X_N)$, rispetto a delle relazioni di indipendenza indicizzate da relazioni *genitore-figlio* (*parent-child*) in un grafo aciclico diretto (DAG) $G = (V, E)$ dove $V = \{1, \dots, N\}$.

Formalmente, sia G un *DAG* e sia $\{f(x_i | x_{pa(i)})\}$ per $i \in V$ un insieme di densità di probabilità condizionate per ogni variabile casuale X_i , dato il suo genitore $X_{pa(i)}$. Un modello grafico diretto definisce la probabilità congiunta P come:

$$f(x_1, \dots, x_N) = \prod_{i \in V} f(x_i | x_{pa(i)})$$

Esempio: Movies

L'industria cinematografica sta diventando sempre più competitiva. Secondo IMDB, il numero medio di film prodotti ogni anno è 2577. Nel 2015 sono stati prodotti 9387 film nel mondo. È impossibile per tutti questi film sopravvivere in questo feroce mercato competitivo. Tutte le compagnie cinematografiche stanno esplorando il modo di aumentare il loro incasso, scegliendo accuratamente il cast, cambiando regista, aumentando il budget e così via. È difficile conoscere il gradimento del pubblico prima di vedere effettivamente le recensioni. Molti fattori possono influenzare il pubblico e gli incassi del film, come il regista, il cast, il genere e il budget. E' importante quindi cercare di usare tutti i possibili fattori e dati per costruire il modello per prevedere gli incassi di un film al botteghino con una maggior precisione. I modelli statistici aiutano le compagnie cinematografiche ed i registi in questo compito. Grazie ad accurate previsioni sarebbe infatti possibile ottimizzare il film, aggiustandone ad esempio il piano finanziario. E' inoltre importante per gli investitori cinematografici per aumentare i loro profitti e ridurre i rischi.

Esempio: Movies

```
load("/Users/utente/Documents/prova r pkg/lucidi del corso/movies.RData")
regfilm <- lm(gross ~ budget + score + year + votes,data=revfilm)
summary(regfilm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.328e+08	1.502e+08	4.213	2.57e-05	***
budget	8.250e-01	1.778e-02	46.401	< 2e-16	***
score	-4.172e+05	6.965e+05	-0.599	0.549	
year	-3.157e+05	7.508e+04	-4.205	2.66e-05	***
votes	1.903e+02	5.280e+00	36.039	< 2e-16	***

Multiple R-squared: 0.6057, Adjusted R-squared: 0.6053

I dati riguardano 6820 film prodotti tra il 1986 ed il 2016

1	budget	2	company	9	released
		3	country	10	runtime
		4	director	11	score
		5	genre	12	star
		6	gross	13	votes
		7	name	14	writer
		8	rating	15	year

Esempio: Movies

```
load("/Users/utente/Documents/prova r pkg/lucidi del corso/movies.RData")
library(cluster)
revfilm <- revfilm[!is.na(revfilm$budget),]
revfilm <- revfilm[revfilm$budget > 50000000,]
regfilm <- lm(gross ~ budget + score + year + votes + runtime,data=revfilm)
disf <- as.matrix(1-daisy(revfilm[,c(3,4,5,12,14)],metric = "gower"))
disf[disf < 0.2] <- 0
rownames(disf) <- colnames(disf) <- revfilm$name
library(igraph)
grfilm <-graph_from_adjacency_matrix(disf, mode = "max", weighted = T,diag = F, add.rownames = T)
centr <- data.frame(degree = degree(grfilm),betweenness = betweenness(grfilm),
  closeness = closeness(grfilm),eigenvector = eigen_centrality(grfilm)$vector)
datifilm <- cbind(revfilm,centr)
kk <- datifilm[order(datifilm$eigenvector,decreasing = T),]
regfilm <- lm(gross ~ budget + score + year + votes + runtime+degree+betweenness+
closeness+eigenvector,data=datifilm)
g1 <- as.matrix(as_adjacency_matrix(grfilm))
del <- which(rowSums(g1)==0)
g1 <- g1[-del,-del];datifilm <- datifilm[-del,]
library(spdep)
nb <- mat2listw(g1,style="W")
mor <- moran(datifilm$gross, nb,length(nb),Szero(nb))
sar <- lagsarlm(gross ~ budget + score + year + votes + runtime+degree+betweenness+closeness+
eigenvector, Durbin=T,data=datifilm, nb,tol.solve = 1e-100)
impacts(sar, listw=nb)
sem <- errorsarlm(gross ~ budget + score + year + votes + runtime+degree+betweenness+closeness+
eigenvector, data=datifilm, nb,tol.solve = 1e-100)
```

Esempio: Movies

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.057e+08	7.293e+08	0.145	0.88479	
budget	8.826e-01	5.617e-02	15.712	< 2e-16	***
score	6.804e+06	2.978e+06	2.285	0.02253	*
year	-7.723e+04	3.627e+05	-0.213	0.83144	
votes	2.363e+02	1.390e+01	17.002	< 2e-16	***
runtime	-3.910e+05	1.226e+05	-3.190	0.00147	**
degree	2.772e+05	6.316e+04	4.389	1.26e-05	***
betweenness	-2.690e+03	2.838e+03	-0.948	0.34348	
closeness	1.001e+11	2.006e+11	0.499	0.61772	
eigenvector	-1.039e+08	2.138e+07	-4.860	1.37e-06	***

Multiple R-squared: 0.5636, Adjusted R-squared: 0.5596

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.0118e+09	4.6532e+09	-0.2174	0.8279
budget	8.0604e-01	5.9031e-02	13.6546	<2e-16
score	3.4579e+06	3.1810e+06	1.0870	0.2770
year	-1.3309e+05	3.6136e+05	-0.3683	0.7127
votes	2.4232e+02	1.4108e+01	17.1767	<2e-16
runtime	-1.0523e+05	1.4035e+05	-0.7498	0.4534
degree	-1.6621e+05	2.1691e+05	-0.7663	0.4435
betweenness	-1.8679e+03	3.2004e+03	-0.5837	0.5595
closeness	1.4305e+12	1.9859e+12	0.7203	0.4713
eigenvector	7.8659e+07	8.7310e+07	0.9009	0.3676

Rho: 0.22035, LR test value: 3.4431, p-value: 0.063518

Log likelihood: -19406.02 for mixed model

Number of observations: 999

LM test for residual autocorrelation

test value: 3.003, p-value: 0.083113

Esempio: Movies

Impact measures (mixed, exact):

	Direct	Indirect	Total
budget	8.072265e-01	3.066013e-01	1.113828e+00
score	3.512832e+06	1.416941e+07	1.768224e+07
year	-1.301804e+05	7.498124e+05	6.196320e+05
votes	2.421251e+02	-5.105185e+01	1.910733e+02
runtime	-1.081286e+05	-7.482501e+05	-8.563787e+05
degree	-1.646819e+05	3.952678e+05	2.305859e+05
betweenness	-1.952475e+03	-2.181791e+04	-2.377039e+04
closeness	1.425177e+12	-1.366882e+12	5.829482e+10
eigenvector	7.801581e+07	-1.659830e+08	-8.796724e+07

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.2505e+08	7.5356e+08	0.5641	0.572715
budget	8.5202e-01	5.7233e-02	14.8868	< 2.2e-16
score	5.6790e+06	3.0752e+06	1.8467	0.064784
year	-1.5171e+05	3.6187e+05	-0.4192	0.675042
votes	2.3828e+02	1.4025e+01	16.9891	< 2.2e-16
runtime	-2.7087e+05	1.2977e+05	-2.0873	0.036863
degree	3.3506e+05	9.5027e+04	3.5259	0.000422
betweenness	-1.7587e+03	2.9555e+03	-0.5951	0.551795
closeness	-1.1593e+12	1.2378e+12	-0.9365	0.348995
eigenvector	-1.1609e+08	2.9405e+07	-3.9479	7.882e-05

Lambda: 0.30626, LR test value: 7.4396, p-value: 0.0063802

Log likelihood: -19413.49 for error model

ML residual variance (sigma squared): 4.4243e+15, (sigma: 66515000)

AIC: 38851, (AIC for lm: 38856)