

L'analisi dei cluster e l'analisi delle concatenazioni probabilistiche di nodi semantici

Maria Paola Piccini

mariapaola.piccini@uniroma1.it

mariapaolapiccini@virgilio.it

Università Pontificia Salesiana –

Facoltà di Scienze della Comunicazione Sociale

9. La cluster analysis

Spad.T (Lebart, Morineau, Bécue, 1989) prevedeva anche la possibilità applicare procedure di *clustering* sulla base dei risultati dell'analisi fattoriale.

Con la denominazione *cluster analysis* si indicano le procedure con le quali è possibile costruire, coinvolgendo più variabili contemporaneamente, **classi** o gruppi (*clusters*) di unità di un determinato insieme, in modo che le unità di una stessa classe siano **il più possibile omogenee tra loro** rispetto alle modalità delle variabili considerate e che ogni unità sia contenuta in una classe soltanto.

Tale procedura consente di aggregare un insieme di unità di analisi in un numero ridotto di gruppi, in maniera tale che siano il più possibile omogenei al loro interno e ben distinti tra loro.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Il risultato di questo tipo di analisi, applicata al corpus dei metatesti relativi ai protocolli delle interviste condotte in occasione del Giubileo del 2000, consiste in **una tipologia sintetica di interviste**, differenziate in funzione delle **categorie tematiche riscontrate**.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Rientrano nel **primo cluster** poco meno della metà delle interviste, che si sviluppano toccando temi relativi alla **famiglia**, riportando episodi relativi sia a quella d'origine che a quella acquisita, ma anche vicende connesse con l'istruzione e l'educazione ricevute, con la propria occupazione, con problematiche esistenziali e di salute, per proporre una **presentazione di sé** che va ben **oltre l'esperienza giubilare**.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Cluster 1, n = 42 interviste

Item	V.T.	% interna	% globale	Freq. interna	Freq. globale	Prob.
Famiglia	10.7	18.2	10.9	274	492	0.000
Scuola	6.4	6.9	4.1	104	185	0.000
Lavoro	5.3	5.9	3.7	89	167	0.000
Soldi	2.9	1.7	1.0	25	45	0.002
Male	2.3	6.9	5.8	105	261	0.011

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Nel **secondo cluster** sono incluse interviste che legano l'esperienza di fede alla sfera **emozionale**, alla religiosità intesa come pratica religiosa vissuta attivamente in una **dimensione comunitaria**.

Il Giubileo è un evento considerato anche per come i mass media lo rappresentano, senza tralasciare le implicazioni **turistiche** e rivelando un esplicito interesse per la città di Roma.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Il cluster, n = 18 interviste

Item	V.T.	% interna	% globale	Freq. interna	Freq. globale	Prob.
Tempo	7.1	7.1	4.3	121	193	0.000
Fede	5.4	15.6	12.2	266	548	0.000
Emozione	4.2	3.7	2.4	63	109	0.000
Religiosità	4.1	9.9	7.7	168	347	0.000
Altri	3.3	5.2	3.9	89	178	0.000
Mass media	3.1	2.9	2.0	50	92	0.001
Roma	2.8	4.6	3.5	78	159	0.002
Chiese	2.7	2.8	2.0	47	90	0.004

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Rientrano invece nel **terzo cluster** le interviste di quanti considerano l'esperienza giubilare essenzialmente come **esperienza collettiva**, da condividere con la comunità di appartenenza, i gruppi di amici, i gruppi parrocchiali o di volontariato, come occasione da vivere con gli **altri** e che avvicina agli **altri**.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

III cluster, n = 17 interviste

Item	V.T.	% interna	% globale	Freq. interna	Freq. globale	Prob.
Giubileo	11.7	34.3	15.8	194	713	0.000
Altri	3.1	6.5	3.9	37	178	0.001

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Le interviste classificate nel **quarto cluster** si caratterizzano per i forti richiami ai temi della **politica**, del **potere**, del **lavoro**. Abbastanza forte in queste interviste è anche il riferimento a **valori** come l'amicizia e l'amore, le tradizioni e la patria, la solidarietà e l'impegno nel volontariato.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

IV cluster, n = 11 interviste

Item	V.T.	% interna	% globale	Freq. interna	Freq. globale	Prob.
Politica	7.5	7.2	1.1	23	50	0.000
Lavoro	3.2	7.5	3.7	24	167	0.001
Valori	1.9	11.0	7.9	35	357	0.028

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Infine, **nell'ultimo cluster** sono incluse le interviste di persone per le quali l'intervista stessa è un'occasione per esprimere le proprie **opinioni** in merito a temi diversi, quali i rapporti con le **altre religioni**, le **regole** proprie della pratica religiosa, apprese attraverso l'educazione ricevuta in famiglia e durante il **percorso scolastico**. Per finire, nelle interviste di questo ultimo cluster si fa riferimento alle **questioni economiche** e a tematiche relative ai **mezzi di comunicazione di massa**, in considerazione del ruolo da essi svolto durante il Giubileo del 2000.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

V cluster, n = 8 interviste

Item	V.T.	% interna	% globale	Freq. interna	Freq. globale	Prob.
Altre religioni	7.5	4.9	0.7	20	31	0.000
Opinione	4.7	3.4	0.8	14	36	0.000
Scuola	3.8	8.1	4.1	33	185	0.000
Regole	2.1	1.9	0.8	8	38	0.018
Mass media	1.8	3.4	2.0	14	92	0.035
Soldi	1.7	2.0	1.0	8	45	0.045

10. Analisi delle concatenazioni probabilistiche di nodi semantici

Si effettua con il programma **DiscAn** (**Discourse Analysis** Maranda, 1995).

Si applica su metatesti ciascuno costituito dalla successione delle etichette verbali associate alle **categorie** in cui sono state classificate le unità testuali dei testi originari.

DiscAn calcola le **probabilità di transizione**, nel corpus di testi in esame, da ciascuna categoria a quelle che immediatamente la precedono e la seguono, consentendo di ricostruire la rete delle **concatenazioni probabilistiche** tra le categorie stesse.

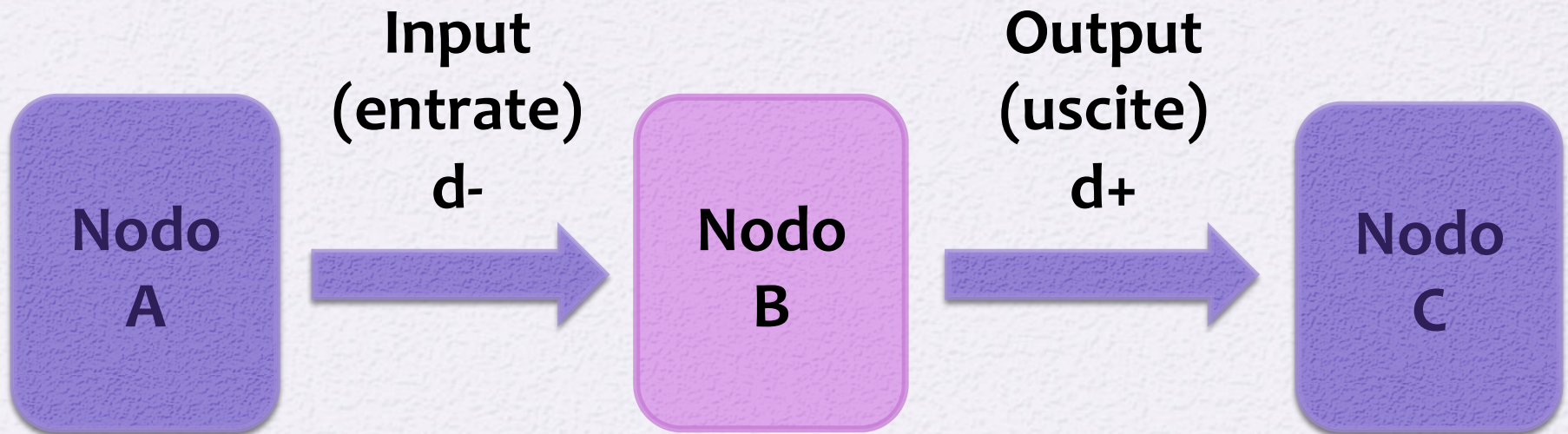
10. Analisi delle concatenazioni probabilistiche di nodi semantici

Le **categorie** rappresentano **nodi** semantici per ciascuno dei quali è possibile calcolare:

- il grado di **ricezione**, denominato *semigrado interno* e indicato con d_- , che corrisponde all'**input** del nodo;
- il grado di **emissione**, denominato *semigrado esterno* e indicato con d_+ , che corrisponde agli **output** del nodo.

Considerando il rapporto d_+/d_- per ogni nodo, è possibile definirne il ruolo nella dinamica discorsiva del corpus in esame.

10. Analisi delle concatenazioni probabilistiche di nodi semantici



10. Analisi delle concatenazioni probabilistiche di nodi semantici

Se gli output sono più numerosi degli input,
si ha

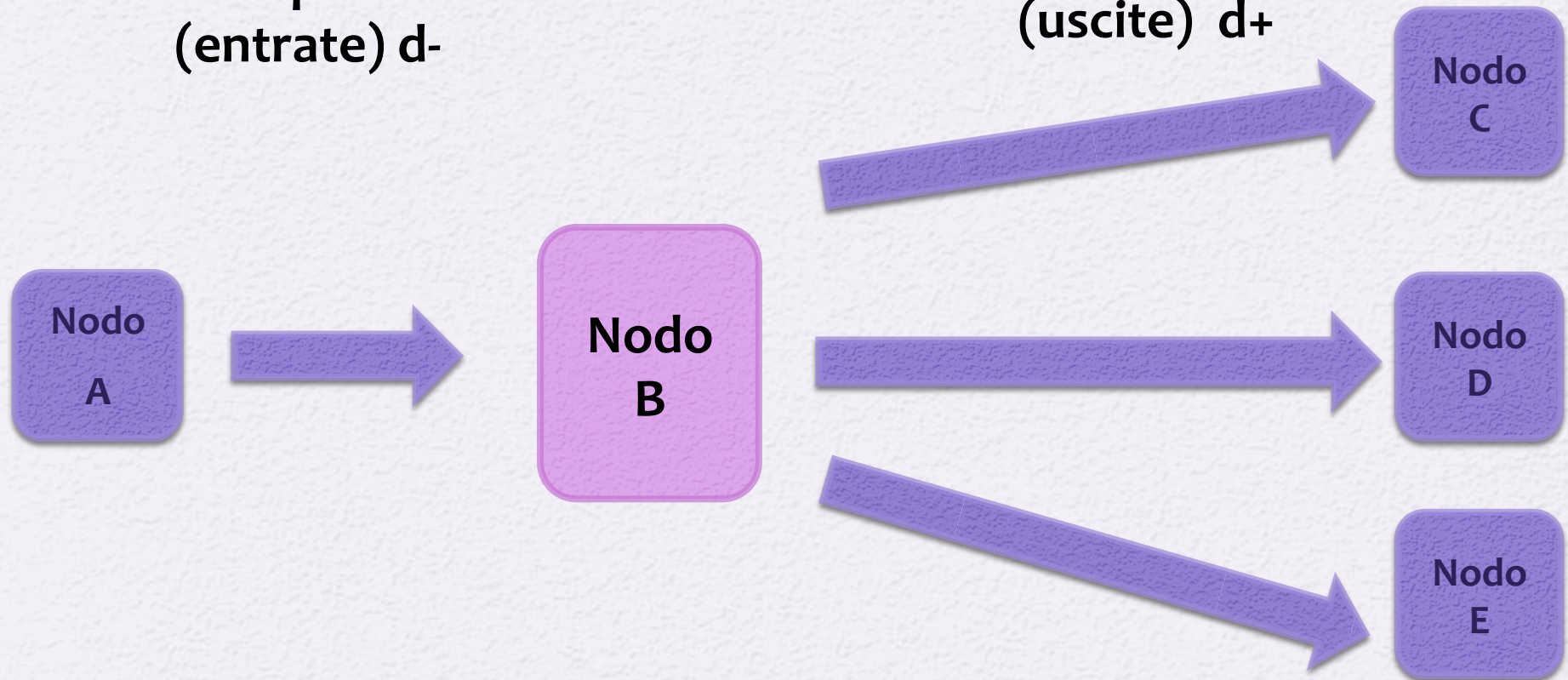
$$d+/d- > 1$$

e la categoria è un nodo diffrattore o *fonte*

10. Analisi delle concatenazioni probabilistiche di nodi semantici

Input
(entrate) d-

Output
(uscite) d+



$d+/d- = 3/1 = 3 > 1$ perciò **Nodo B** è una **FONTE**

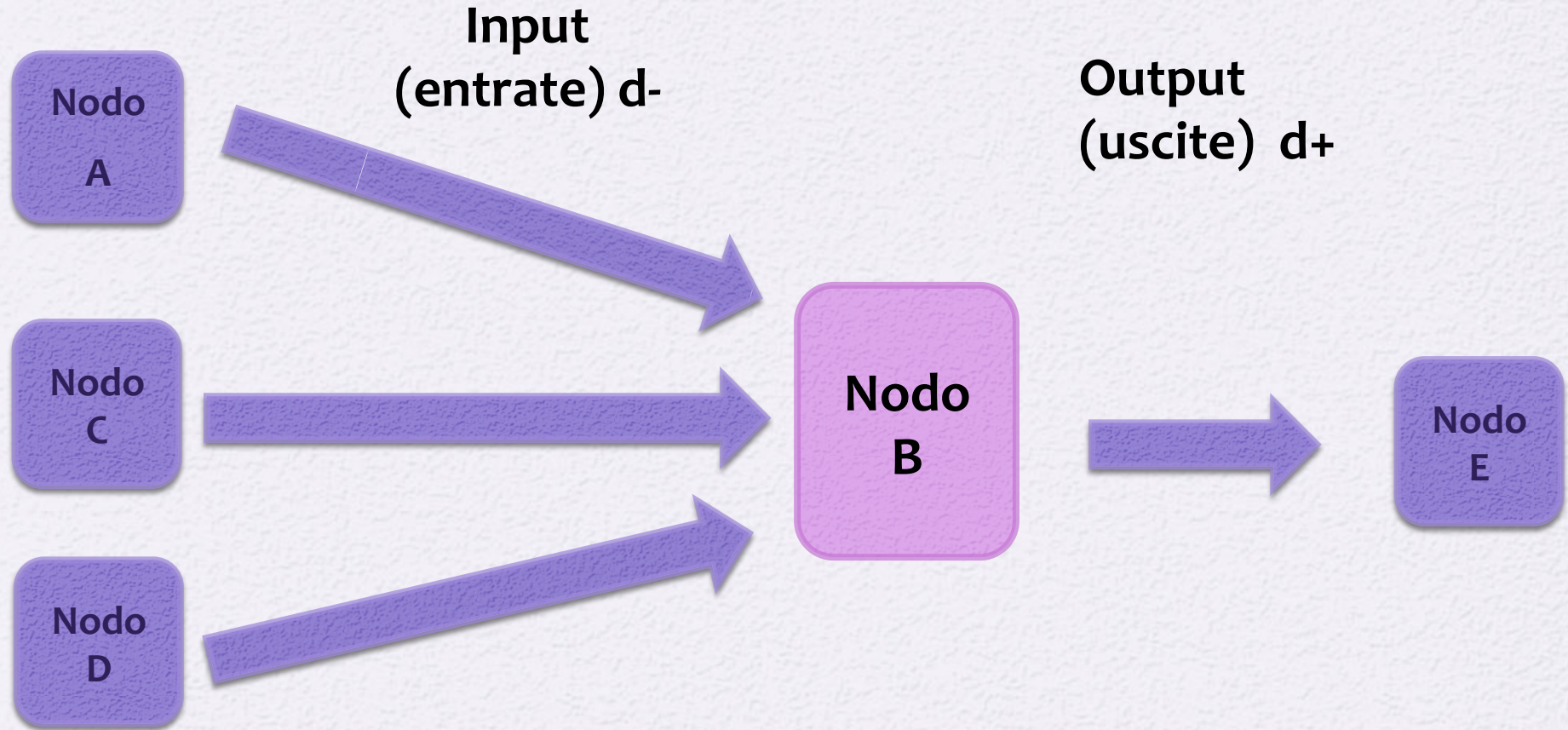
10. Analisi delle concatenazioni probabilistiche di nodi semantici

Se gli input sono più numerosi degli output,
si ha

$$d+/d- < 1$$

e la categoria è un nodo attrattore o
condensatore

10. Analisi delle concatenazioni probabilistiche di nodi semantici



$d+/d- = 1/3 = 0,33 < 1$ perciò Nodo B è un **CONDENSATORE**

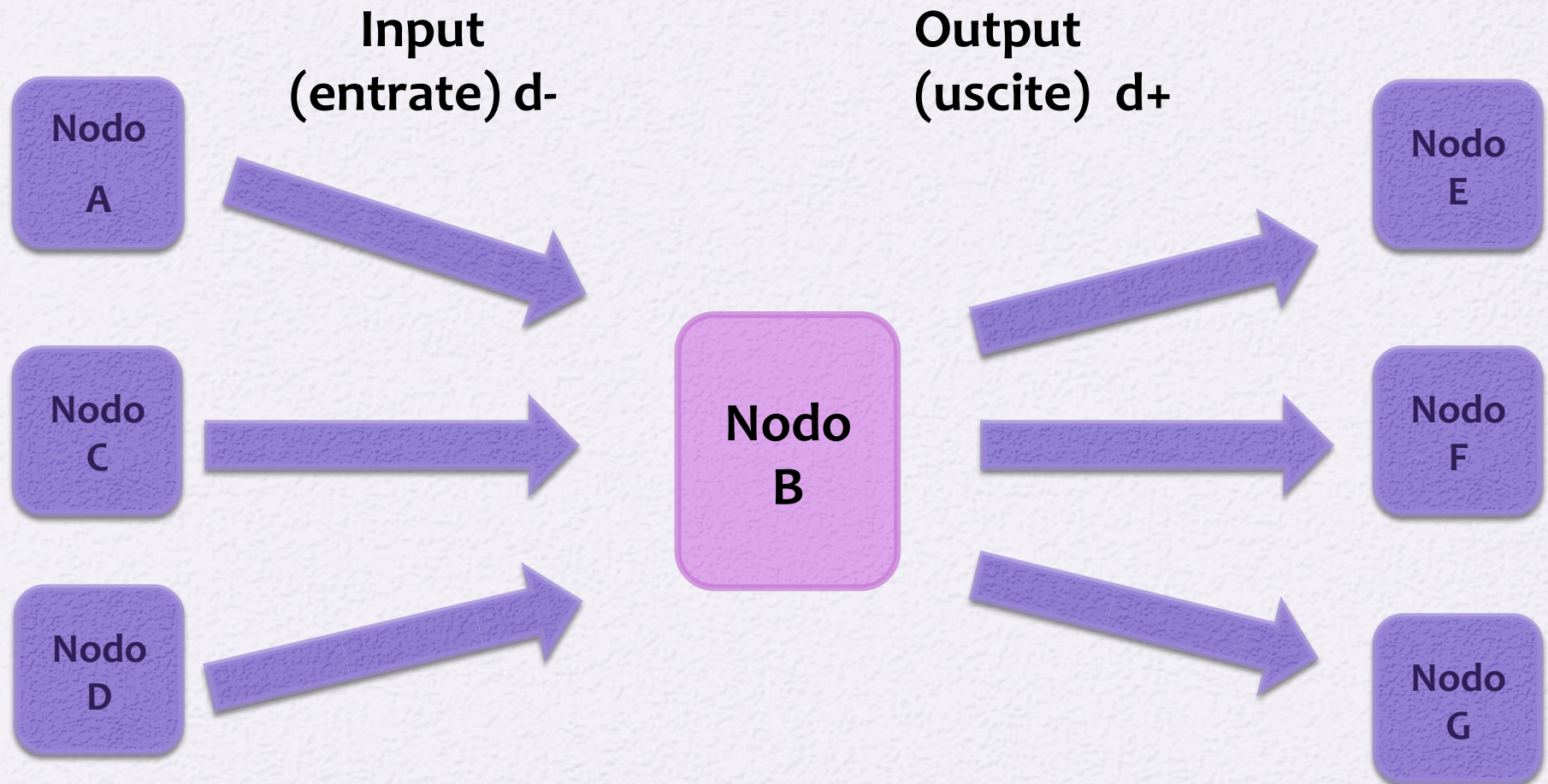
10. Analisi delle concatenazioni probabilistiche di nodi semantici

Se input e output si equivalgono, si ha

$$d+/d- = 1$$

e la categoria è un nodo trasmettitore o
relay

10. Analisi delle concatenazioni probabilistiche di nodi semantici



$d+/d- = 3/3 = 1$ perciò **Nodo B** è un **RELAY**

10. Analisi delle concatenazioni probabilistiche di nodi semantici

Per ciascun nodo (mediante la somma dei prodotti del semigrado interno e del semigrado esterno per la frequenza con cui il nodo stesso si presenta) DiscAn calcola l'**attività semantica**, ovvero il **contributo** di esso alla **dinamica discorsiva** del corpus nel suo complesso:

$$(d_+ \times f) + (d_- \times f)$$

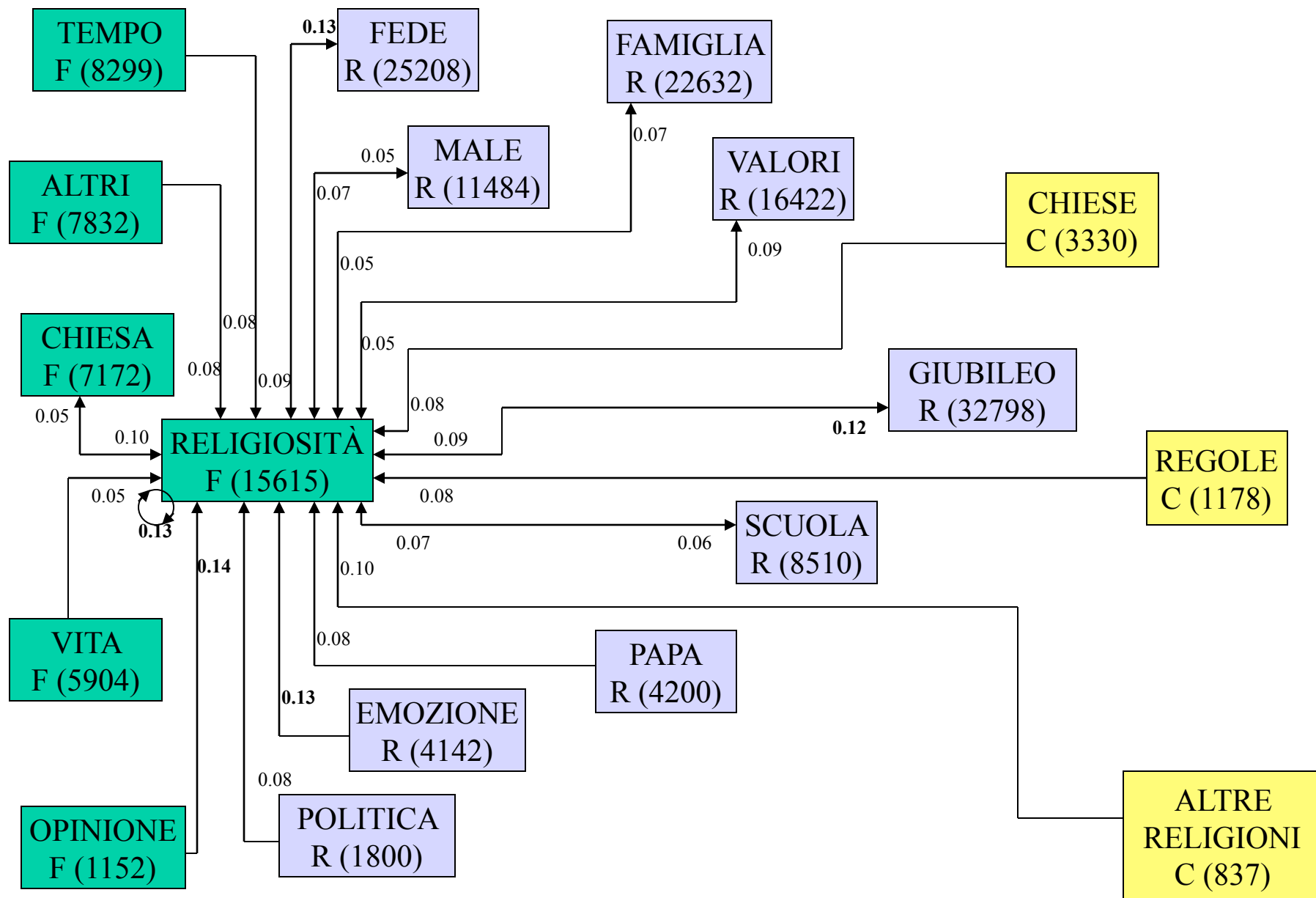
L'attività semantica sostanzialmente esprime l'**intensità con cui ogni nodo svolge una delle tre funzioni**, di attrazione, diffrazione o trasmissione.

10. Analisi delle concatenazioni probabilistiche di nodi semantici

Sulla base delle informazioni fornite da DiscAn, quindi tenendo conto delle probabilità di transizione tra i nodi e del ruolo di ciascuno di essi, è possibile disegnare una **mappa semantica** che rappresenta graficamente le concatenazioni dei nodi. Quando i nodi semantici e le relazioni di transizione da rappresentare sono numerosi, come nel presente caso, la costruzione della mappa complessiva appare piuttosto complessa.

Dunque, si procede in genere alla costruzione di più **'micro-mappe'** che focalizzano le relazioni di transizione dei **nodi più attivi** - rispettivamente come fonte, come relay e come condensatore - con gli altri nodi.

Oppure si possono visualizzare solo i nodi con le probabilità di transizione più elevate (**relazioni più forti** fra i nodi semantici).



Un esempio di applicazione: le interviste ai Giubilanti del 2000:

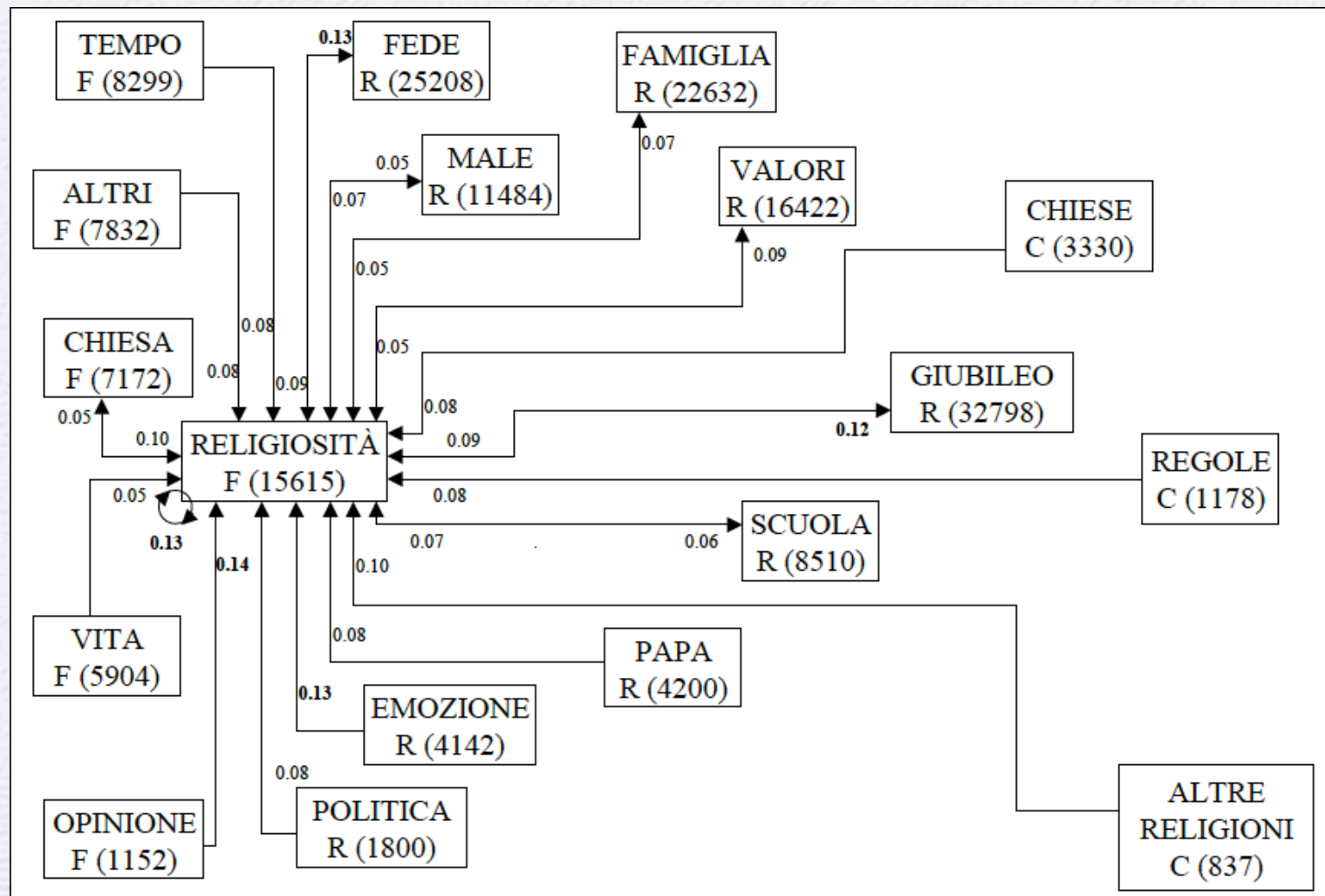
Da questa categoria si dipartono differenti percorsi discorsivi che si caratterizzano, da un lato, per l'attenzione posta su temi attinenti all'esperienza religiosa riferita alla **fede** e, dall'altro, per il ricondurre tale esperienza anche alle concrete dimensioni della vita terrena e ai problemi ad esse sottesi (**il male, la famiglia, i valori**).

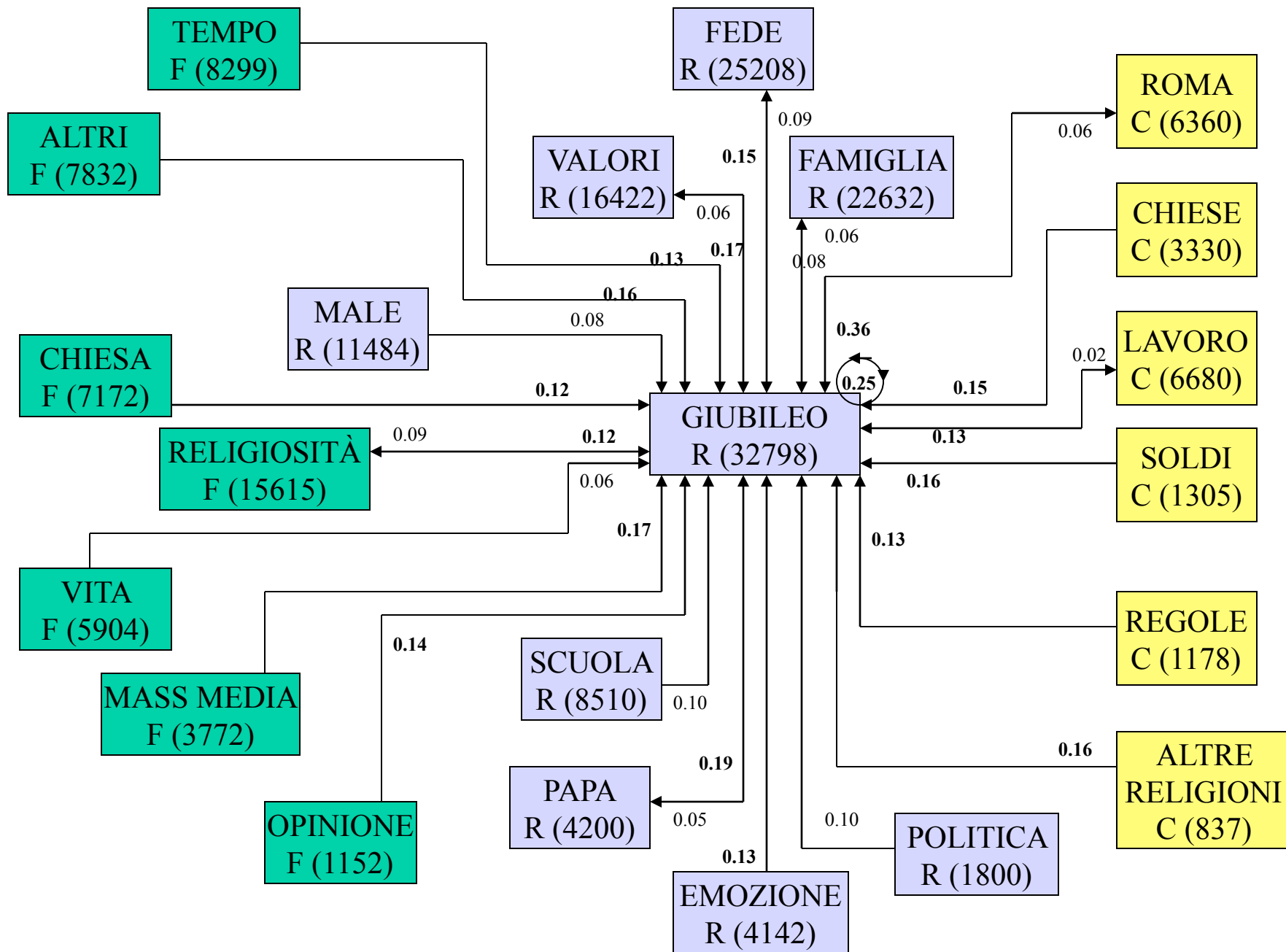
D'altra parte, come del resto era da attendersi, si tende ad associare alla *religiosità* anche il riferimento alla personale condizione esistenziale (**vita**), per come si declina nella particolare stagione (**tempo**) che si sta percorrendo, ed anche e soprattutto alla sfera emotiva (**emozione**). La *religiosità* assume dunque, in parte delle interviste dei giubilanti, il profilo di un'esperienza complessa e coinvolgente, nella quale si integrano aspetti propri della sfera interiore, della sfera individuale-privata, della sfera relazionale, un'esperienza che nella partecipazione all'evento **Giubileo** sembra trovare una compiuta espressione oltre che sul piano della devozione, anche sul piano simbolico.

Da notare, infine, come la categoria **religiosità** torni più volte su se stessa (ovvero sia preceduta e seguita da se stessa, con eguale probabilità pari a $p=0,13$) a sottolineare la propensione degli intervistati ad approfondire i temi da essa implicati e, dunque, la particolare rilevanza che essa assume nella dinamica associativa del corpus oggetto d'indagine.

Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Mappa semantica focalizzata sulla FONTE con attività più elevata: “religiosità”





Un esempio di applicazione: le interviste ai Giubilanti del 2000:

La categoria più attiva con funzione di nodo di collegamento, o relay, è la categoria *Giubileo* che è anche la più attiva in assoluto indipendentemente dalla funzione svolta, come del resto era da attendersi in rapporto la sua centralità nelle interviste che appunto su questo evento erano mirate.

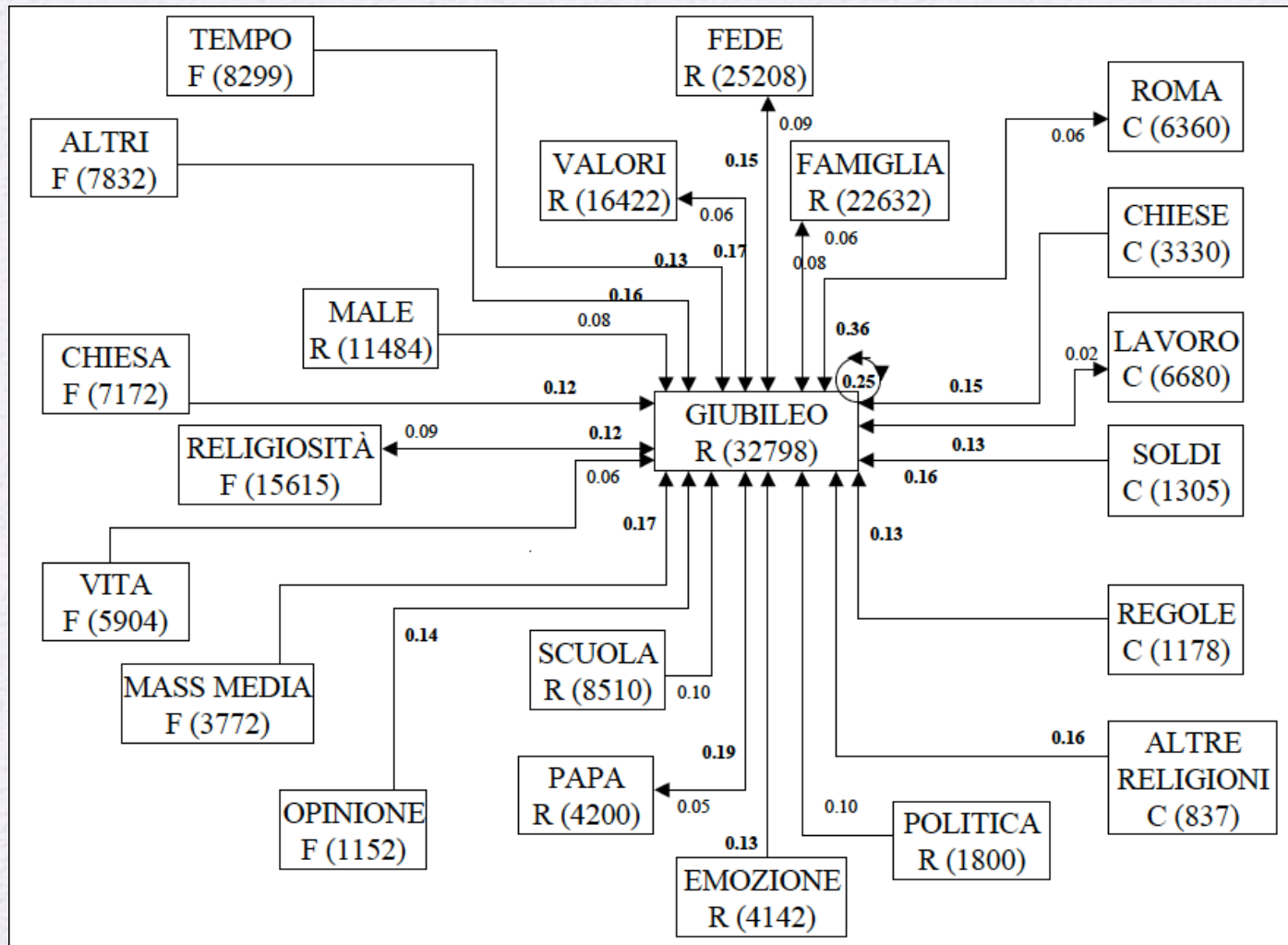
Già si è detto della relazione di **Giubileo** con **religiosità** e, quindi, della forte valenza religiosa e di fede attribuita all'esperienza giubilare. Stante questo legame e coerentemente con esso, il **Giubileo** coinvolge gli intervistati non solo in relazione alla fede (**fede**), ma anche al vissuto interiore (**emozione**) e a quello relazionale e sociale (**famiglia, lavoro, regole**).

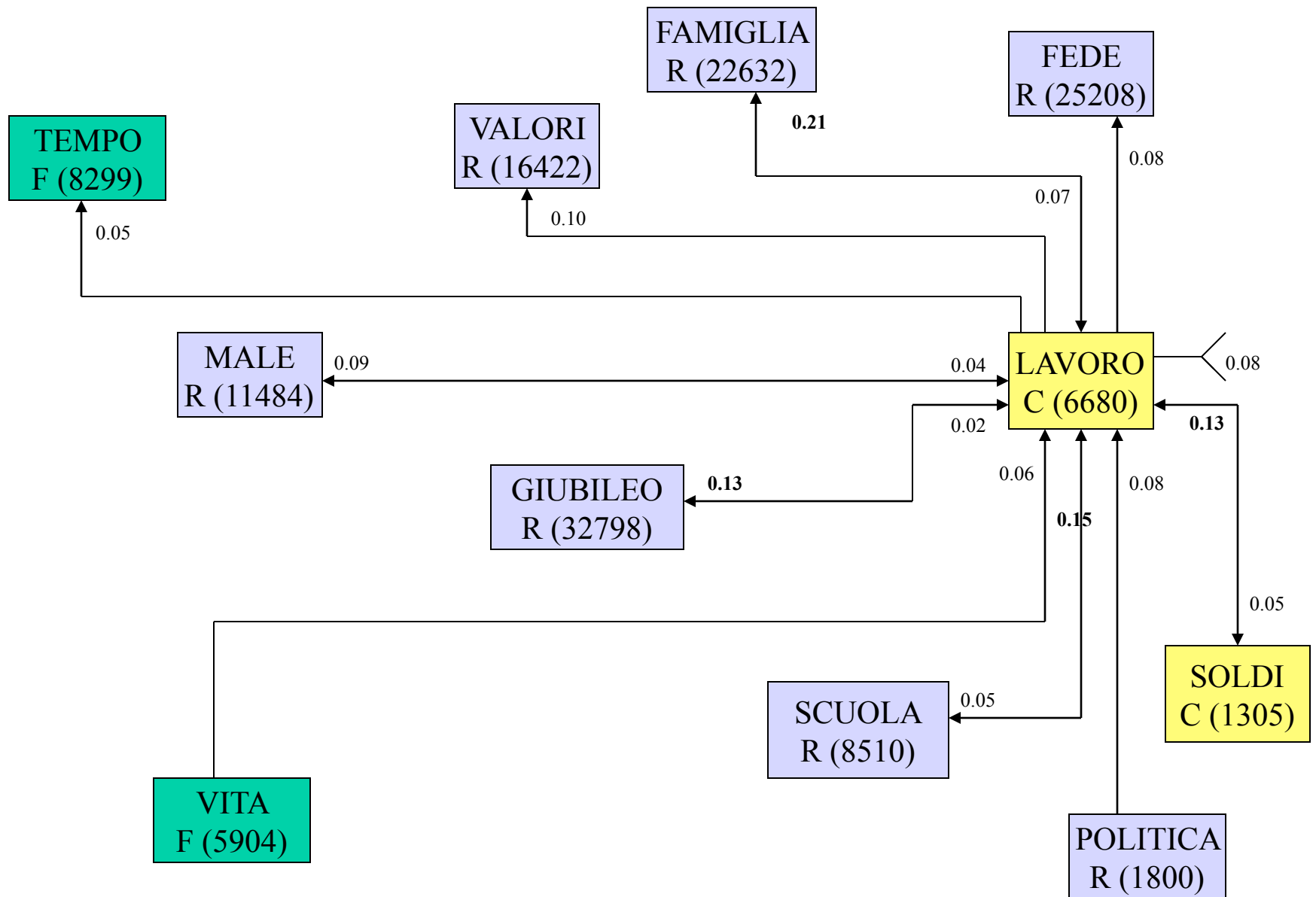
Oltre a ciò si evidenzia, in parte delle interviste, la consapevolezza di talune implicazioni «profane» di questo grande evento: l'attenzione ad esso riservata dai mezzi d'informazione, spesso in relazione alle sole manifestazioni esteriori (**mass media**), la valenza anche turistica del pellegrinaggio (**Roma, chiese**), gli aspetti economici legati al viaggio e al soggiorno nella Capitale (**soldi**).

Così, il racconto degli intervistati segue percorsi discorsivi che giungono al Giubileo partendo da richiami all'esperienza quotidiana, con i problemi e le difficoltà che la caratterizzano, talvolta coniugandoli con espliciti riferimenti all'esperienza religiosa, per poi proseguire con la narrazione del viaggio e del rapporto con Roma.

Inoltre, anche per la categoria *Giubileo*, come per *religiosità*, si riscontrano frequenti ritorni su se stessa, a testimoniare il ricorrere in sequenza, nelle interviste, di temi ad essa riconducibili.

Mappa semantica focalizzata sul RELAY con attività più elevata: “Giubileo”



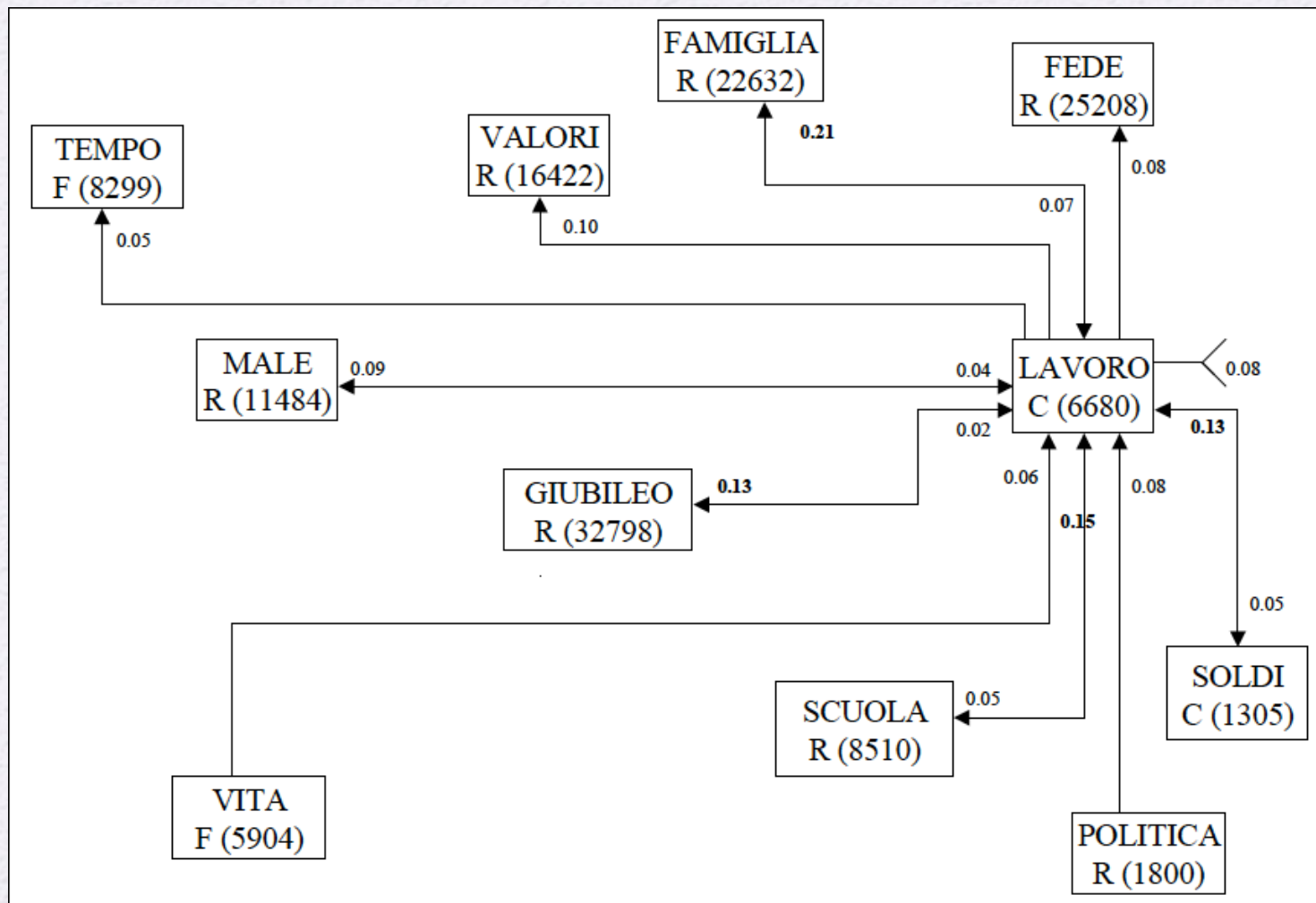


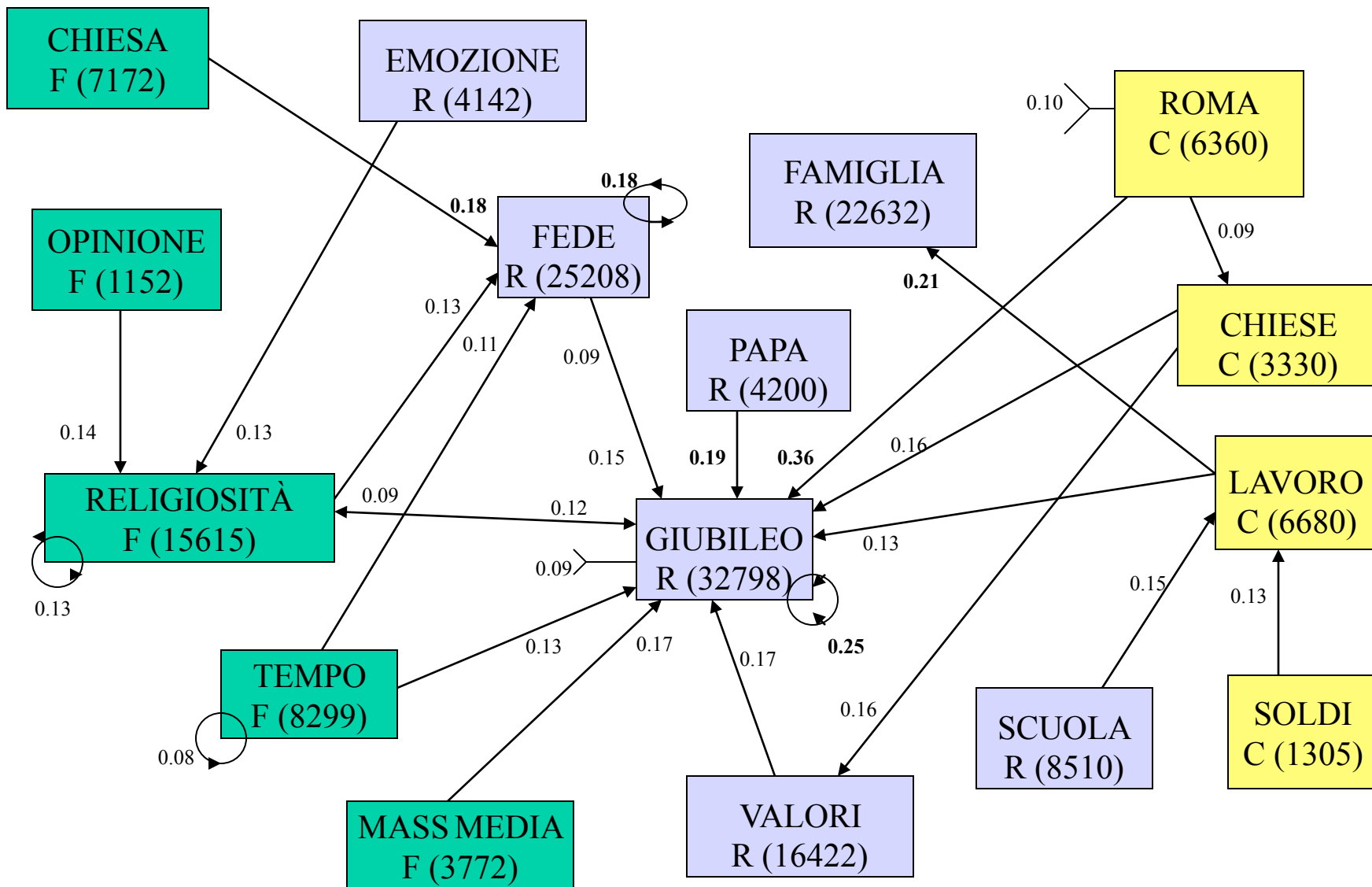
Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Quanto ai nodi attrattori o condensatori, la categoria più attiva risulta essere **lavoro**, anch'essa in relazione con nodi riferibili alla dimensione religiosa ma anche, e soprattutto, alla dimensione relazionale e sociale. In parte delle interviste, il discorso tende a convergere verso temi relativi al lavoro degli intervistati o dei loro congiunti, soprattutto dei figli, lavoro che sembra imporsi come problema prevalente, talvolta irrisolto e per questo centrale.

Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Mappa semantica focalizzata sul nodo attrattore con attività più elevata: “lavoro”





Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Un primo percorso è quello che testimonia di un'esperienza giubilare vissuta all'insegna di una religiosità più tradizionale, in qualche modo legata agli aspetti "istituzionali" della fede. È la dinamica discorsiva centrata sulle successioni

chiesa → fede ↔ giubileo

Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Un secondo tipo di percorso è invece caratterizzato dall'enfasi posta su un'esperienza religiosa, associata alla partecipazione all'evento giubilare, più "intima", più esplicitamente riferita al vissuto individuale, alla sfera interiore, alla dimensione emotiva. In questo caso la dinamica discorsiva è prevalentemente espressa dalla successione

emozione → religiosità ↔ giubileo ↔ religiosità

Per tale successione si registra una variante che contempla, anche, richiami ai temi della fede:

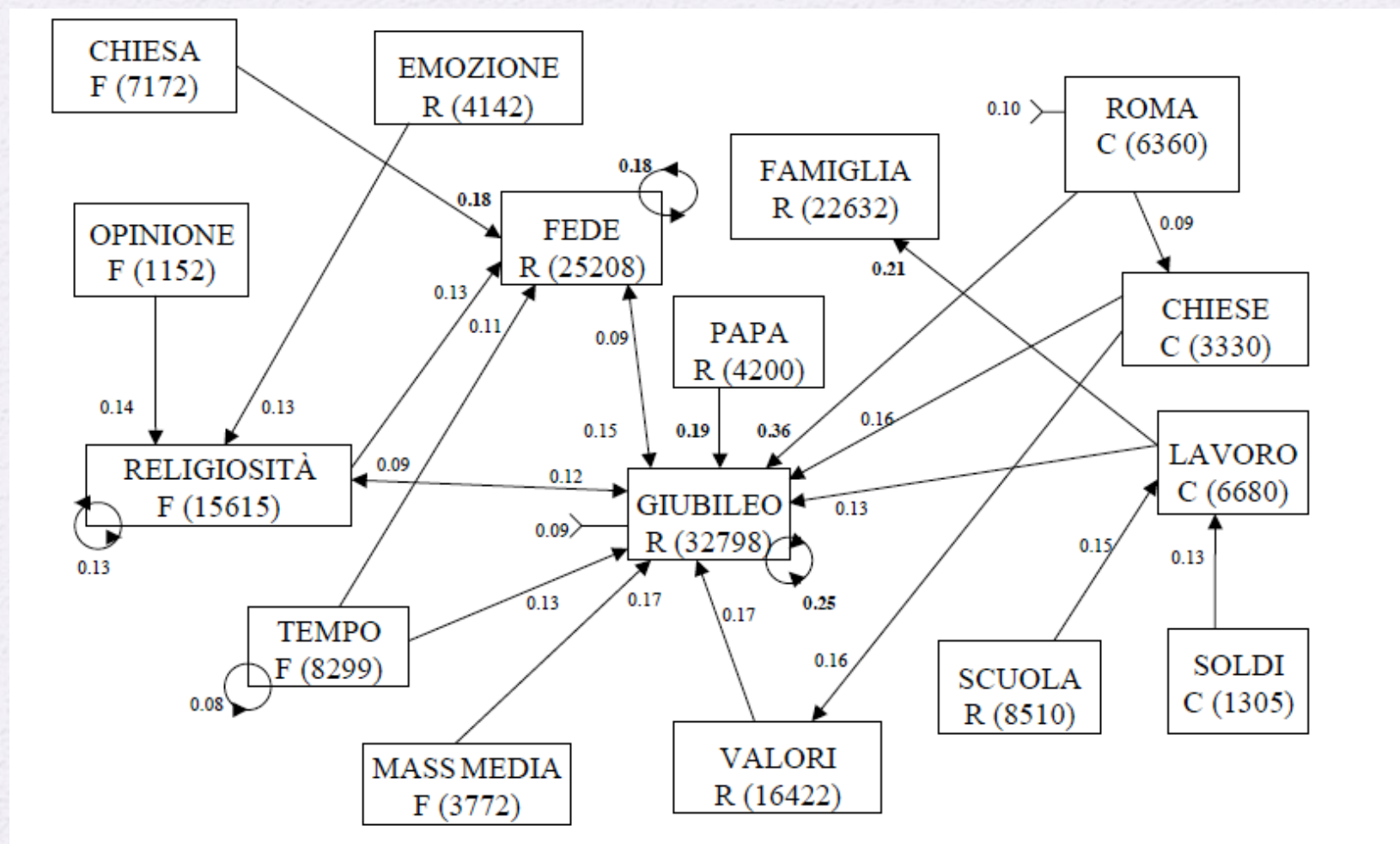
emozione → religiosità → fede ↔ giubileo ↔ religiosità

Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Un terzo tipo, infine, è quello che estende i due precedenti percorsi associando ai temi che li caratterizzano riferimenti, sovente anche dettagliati, alle concrete circostanze, ai concreti problemi, alle concrete vicissitudini della vita relazionale, soprattutto per come si presentano nei due "luoghi" chiave dell'esperienza sociale di ciascuno: la **famiglia** e il **lavoro**.

Un esempio di applicazione: le interviste ai Giubilanti del 2000:

Mappa semantica delle relazioni più forti



Bibliografia essenziale

- Aureli Cutillo E., Bolasco S. (a c. di), 2004, *Applicazioni di analisi statistica dei dati testuali*, Roma, Casa Editrice Università degli studi di Roma La Sapienza.
- Bardin L., 1977, *L'analyse de contenu*, Paris, Presses Universitaires de France.
- Bellelli G. (a c. di), 1989, *Il metodo del discorso. L'analisi delle produzioni discorsive in psicologia e in psicologia sociale*, Napoli, Liguori.
- Benzécri J.P., 1973, *L'analyse des données*, Paris. Dunod.
- Berelson B., 1952, *Content Analysis in Communication Research*, New York, The Free Press, (nuova ed. Hafner Publishing Co., 1971).
- Berelson B., Lazarsfeld P.F., 1948, *The Analysis of Communication Content*, Chicago-New York, University of Chicago and Columbia University.
- Bolasco S., 1999, *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*, Roma, Carocci.
- Cannavò L., Frudà L. (a c. di), 2007, *Ricerca sociale*, 3 voll., Roma, Carocci.
- Cecconi L. (a c. di), 2002, *La ricerca qualitativa in educazione*, Milano, Angeli.
- Cipriani R. (a c. di), 2006, *L'approccio qualitativo. Dai dati alla teoria nell'analisi sociologica*, Milano, Guerini Scientifica.
- Cipriani R., Bolasco S. (a c. di), 1995, *Ricerca qualitativa e computer. Teorie, metodi e applicazioni*, Milano, Franco Angeli.
- Cipriani R., Losito G. (a c. di), 2008, *Dai dati alla teoria sociale. Analisi di un evento collettivo*, Roma, Anicia.
- della Ratta-Rinaldi F., 2002, “L'analisi testuale, uno strumento per la ricerca qualitativa”, in Cecconi L. (a c. di), *La ricerca qualitativa in educazione*, Milano, Angeli, 2002.
- della Ratta-Rinaldi F., 2007, “L'analisi testuale computerizzata”, in Cannavò L., Frudà L. (a c. di), *Ricerca sociale*, vol. II, Roma, Carocci, 2007, pp. 133-152.
- della Ratta-Rinaldi F., 2007, “L'analisi multidimensionale dei testi”, in Cannavò L., Frudà L. (a c. di), *Ricerca sociale*, vol. III, Roma, Carocci, 2007, pp. 133-150.
- De Lillo A. (a c. di), 1971, *L'analisi del contenuto*, Bologna, Il Mulino.
- De Sola Pool I. (ed.), 1959, *Trends in Content Analysis*, Urbana, University of Illinois Press.
- D'Unrug M.-C., 1974, *Analyse de contenu et acte de parole*, Paris, Delarge.
- Ercolani A.P., Areni A., Mannetti L., 2004⁹, *La ricerca in psicologia*, Roma, Carocci (I ed. 1990).
- Ericsson K.A., Simon H.A., 1984, *Protocol Analysis. Verbal Reports as Data*, Cambridge, MIT Press.
- Fielding N.G., Lee R.M. (eds.), 1991, *Using Computers in Qualitative Research*, London, Sage.

Bibliografia essenziale

- Gerbner G., Holsti O.R., Krippendorff K., Paisley W.J., Stone Ph. J., 1969, *The Analysis of Communication Content. Developments in Scientific Theory and Computer Techniques*, New York, Wiley.
- Ghiglione R., 1995, “Teorie, metodi e uso del computer nell’analisi del contenuto: alcuni problemi”, in Cipriani R., Bolasco S. (a c. di), *Ricerca qualitativa e computer. Teorie, metodi e applicazioni*, Milano, Franco Angeli, 1995, pp. 71-86.
- Ghiglione R., Blanchet A., 1991, *Analyse de contenu et contenu d’analyse*, Paris, Dunod.
- Giuliano L., 2004, *L’analisi automatica dei dati testuali. Software e istruzioni per l’uso*, <http://www.ledonline.it>, LED Edizioni Universitarie di Lettere Economia Diritto.
- Giuliano L., La Rocca G., 2008, *L’analisi automatica e semi-automatica dei dati testuali*, Milano, LED Edizioni Universitarie di Lettere Economia Diritto.
- Giuliano L., La Rocca G., 2010, *L’analisi automatica e semi-automatica dei dati testuali II strategie di ricerca e applicazioni*, Milano, LED Edizioni Universitarie di Lettere Economia Diritto.
- Krippendorff K., 1980, *Content Analysis: An Introduction to its Methodology*, London, Sage (trad. it. *Analisi del contenuto. Introduzione metodologica*, Torino, Eri, 1983).
- Lebart L., Morineau A., 1982, *SPAD Système Portable Pour l’Analyse des Données*. Paris, Cisia.
- Lebart L., Morineau A., Bécue M., 1989, *Spad.T Système Portable Pour l’Analyse des Données Textuelles*, Paris, Cisia.
- Lebart L., Salem A., 1988, *Analyse statistique des données textuelles. Question ouvert et lexicométrie*, Paris, Dunod.
- Lebart L., Salem A., 1994, *Statistique textuelle*. Paris, Dunod.
- Losito G., 2002⁴, *L’analisi del contenuto nella ricerca sociale*, Milano, Franco Angeli (1 ed. 1993).
- Losito G., 2004, *L’intervista nella ricerca sociale*, Roma-Bari, Laterza.
- Losito G., Piccini M.P., 2006, “L’analisi della dinamica discorsiva per la ricerca qualitativa”, in Cipriani R. (a c. di), *L’approccio qualitativo. Dai dati alla teoria nell’analisi sociologica*, Milano, Guerini Scientifica, 2006, pp.62-83.
- Losito G., 2007, “L’analisi del contenuto nella ricerca sociale”, in Cannavò L., Frudà L. (a c. di), *Ricerca sociale*, Vol II, Roma: Carocci, 2007, pp. 117-132.

Bibliografia essenziale

- Maranda, P. (1990), *DiscAn: Users Manual*, Nadeau Caron Informatique, Québec.
- Maranda, P., “DiscAn: un programma di analisi reticolare per la costruzione di mappe semantiche”, in Cipriani, R., Bolasco, S. (a cura di) (1995), pp. 171-183.
- Ogilvie D.M., Stone P.J., Kelly E.F., 1980, “Computer-Aided Content Analysis”, in Smith R.B., Manning P.K. (eds.), *Handbook of Social Science Research Methods*, New York, Irvington, 1980.
- Piccini M.P., 2008, “Discan, Spad.T e l’analisi qualitativa”, in Cipriani R., Losito G. (a c. di), *Dai dati alla teoria sociale. Analisi di un evento collettivo*, Roma, Anicia, 2008, pp. 169-197.
- Rosengreen K.E. (ed.), 1981, *Advances in Content Analysis*, London, Sage.
- Rositi F., 1970, *L’analisi del contenuto come interpretazione*, Torino, Eri.
- Rositi F., 1989, “L’amore folle fra l’analisi del contenuto e il computer”, in Bellelli G. (a c. di), *Il metodo del discorso. L’analisi delle produzioni discorsive in psicologia e in psicologia sociale*, Napoli, Liguori, 1989, pp. 107-114.
- Smith R.B., Manning P.K. (eds.), 1980, *Handbook of Social Science Research Methods*, New York, Irvington.