

Science network in Italian population research: an application of Social Network Analysis

*Reti Scientifiche Nella Ricerca Sulla Popolazione In Italia:
Un'applicazione Della Social Network Analysis*

Giulia Rivellini,

Università Cattolica del Sacro Cuore,
Largo Gemelli 1, Milano
giulia.rivellini@mi.unicatt.it

Ester Rizzi

Università degli Studi di Padova,
via Cesare Battisti 241/243, Padova
ester@hal.stat.unipd.it

Riassunto: Il presente lavoro intende essere una prima sintetica analisi dell'attività scientifica degli studiosi di popolazione in Italia. In particolare, si vanno a considerare le collaborazioni tra demografi, in forma di contributo scritto, pubblicato o non pubblicato, nel biennio 1998-1999. Queste costituiscono le "relational ties" o diadi di una rete di relazioni, i cui nodi sono i demografi stessi, e rappresentano le unità statistiche dell'analisi intesa ad indagare le caratteristiche strutturali della rete di relazioni, quali l'appartenenza degli autori alla medesima università, area geografica, ruolo accademico, sesso.

Keywords: references review, social networks, logit models, population studies, dyads

1. Introduction

Since 1992 and every two years, during the Workshop on Italian Population Studies (*Giornate di Studio sulla Popolazione*), the Coordination Group for Demography (*Gruppo di Coordinamento per la Demografia*) delivers a References Review on Population Studies (*Rassegna Bibliografica*) of Italian scientists for the last two years. For the recent edition, references data have been collected and transformed in an electronic form, i.e. a data set useful for the statistical analysis. Regular and not regular university lecturers and professors of Demography, Statistics, Social Statistics, Geography, Human Geography and Sociology (equal in all to 77% of respondents) have been directly involved. The remaining 23% who filled out the form was mainly represented by people belonging either to national, regional, provincial, municipal statistical offices staff, or to national and international research institutes. In addition to these we have looked for new references among the most renown Italian and international journals, web sites or editorial series. The review assembles 967 papers, among those 200 have been written only by demographers. The authors involved were 610.

For references, it was asked to classify works according to the list of themes proposed (principally inspired to Population Index and Revue de Revues Demographiques classifications). The most frequent topics are Internal and External migration, 17% and 7% respectively, Historical Demography (11%), Fertility Behaviours and Studies on Intermediate Fertility Variables (6%). Family Formation, Ageing, Methods, Demographic Sources, Demographic and Economic Interrelation studies count each 4-

5% of papers. Other mentioned topics are, in order of importance, Studies on Mortality, Local Demography, Population Theories, Population and Market Labour, Nuptiality, Divorcing and Populations Reconstruction. Authors could also specify other topics than those listed and in this case some innovative research subjects appear, such as Gender, Sexuality, Youth, Transition to Adulthood, Education, Metropolitan Area, Environment (Rivellini, Rizzi, 2001).

The review does not include all the production on population studies, nevertheless it has two main advantages: 1) review includes also provisional papers not published yet; 2) it is complete for academics' works. The first feature is connected to the aims of the review: (a) to make a useful job tool for the population analysts, who can examine easily references for their preliminary research; (b) to contribute towards an historical database setting for studying the demographic topics evolution; (c) to give a better enlightenment to the demographic scientific research and to promote a better exchange of ideas and methods among who is involved in population studies. The second point, concerning completeness of review for academic's papers, lets us try some analysis on the scientific networks. The available information on authors (n=86) are presented in table 1. In the next paragraph these attributes will be adapted to ties.

Table 1: Actor attributes (for regular academic demographers)

Position	%	Gender	%	Faculty	%	Territorial Area	%
Professor	35.7	Men	61.9	Economy	41.7	North West	17.9
Associate Professor	28.6	Women	38.1	Political Sciences	16.7	North East	2.4
Lecturer	35.7			Statistical Sciences	41.6	Centre	50.0
						South	16.6
						Islands	13.1

Source: Information on academics' features are from <http://www.murst.it>, October 2000

2. Methods and results

A social network can be represented by a sociomatrix where the (i, j) entry is the value of tie between actor i and j . Since the late 1930's sociometricians used statistical models to study sociomatrix. When estimating nonstochastic models, goodness-of-fits statistics and parametric significance tests of network structural properties are not allowed. A new generation of models appears with Markov random graphs. According to Strauss and Ikeda's theorem, Markov random graphs can be approximated by logistic regressions for independent observations, despite existing dependencies among any ties nested in a node. Such logistic regression can be fitted using iteratively reweighted Gauss-Newton computational techniques, as implemented by any logistic regression model package (Wasserman S., Pattison P., 1996).

Looking at our data (table 2) on scientific works from a social network perspective, the authors are the actors or *nodes* of a network and all potential collaborative dyads are the *relational ties*, defined by y_{ij} , where $i=1, \dots, n$ is the index for the first component of the tie and $j=1, \dots, n$ for the second-one (Carolyn J. A et al., 1999). All potential ties are considered, i.e. $n(n-1) = 7,310$. Only actual ties, in the form of scientific work, take value 1 ($y_{ij}=1$), otherwise tie is 0. Total events are 88 ($\sum y_{ij}$). Explanatory variables are indicated by X_{kij} with $k=1, \dots, h$. In our specific analysis variables are the following:

Y ="scientific cooperation between i and j ", X_1 ="faculty", X_2 ="university", X_3 ="geographic area", X_4 ="position", X_6 ="sex". Variables X_1 , X_2 , X_3 are dummies. In order to avoid multicollinearity, we combine variables X_2 and X_3 , university and geographic area. We label the new variable $X_{2,3}$ ="cluster".

Table 2: Data set for logistic regression in social network analysis

Obs	i	J	y_{ij}	X_{1ij}	...	X_{kij}	...	X_{hij}
1	1	2	y_{12}	X_{112}	...	X_{k12}	...	X_{h12}
2	1	3	y_{13}	X_{113}	...	X_{k13}	...	X_{h13}
...	..	..	...	...	...	...	...	...
$n-1$	1	N	y_{1n}	X_{11n}	...	X_{k1n}	...	X_{h1n}
...	..	..	...	...	...	...	...	...
...	n	1	y_{n1}	X_{1n1}	...	X_{kn1}	...	X_{hn1}
...	..	..	...	...	...	...	...	...
$N(n-1)$	n	$n-1$	$y_{n,n-1}$	$X_{1,n,n-1}$	...	$X_{k,n,n-1}$	...	$X_{h,n,n-1}$

One expected result of logistic regression (table 3) is the higher odds of cooperation for people of the same university than for those of a different university and of a different geographic area (baseline). On the contrary, belonging to the same faculty is not statistical significant. As far as academic position, the odds of cooperation is higher for dyads of professors than for any other dyad. But except for this couple, generally, mixture of positions have higher odds than homogeneity of positions.

Table 3: Results from logistic model and network data

Parameters	Estimation	Pr>Chi	Odds Ratio
Intercept	-5,0104	0,0001	
X_1 – Faculty			
a) same faculty	0,3053	0,1928	1,357
b) different faculty (ref.)	0	-	1
$X_{2,3}$ – Cluster			
a) same university-same geographic area	2,5016	0,0001	12,202
b) different university-same geographic area	0,422	0,2177	1,525
c) different university-different geographic area (ref.)	0	-	1
X_4 – Position			
a) PaCRuC	-1,522	0,0003	0,218
b) PoRuC	-0,7346	0,0226	0,480
c) PoPaC	-0,4383	0,1592	0,645
d) RuCRuC	-1,8169	0,0014	0,162
e) PaCPaC	-1,4909	0,0094	0,225
f) PoPo (ref.)	0	-	1
X_5 – Sex			
M-M	0,8373	0,0015	2,310
F-F	1,9447	0,0001	6,991
M-F (ref.)	0	-	1

Legend: Po= professor, PaC=associate professor, RuC= lecturer

Finally, the most interesting result is in higher probability of cooperation for people of the same sex and more for women than men: the odds is 7 times higher for FF than for MF (for sex composition in authors population see table 1).

Conclusions

These results are the very beginning of the more ambitious aim to study the structural properties of population science network in Italy. A limit of analysis is our restricting to a specific academic network without taking into account other sciences (economy, sociology, psychology, anthropology, biology). In order to study interactions with different disciplines, and also with non-academics, future setting of relational databases on research population system could be conceived as multiple networks and the actors of network should be institutions, not authors, for the data set to have a manageable number of nodes. Furthermore, in our work we implicitly assume that cooperation improves progress in science by ideas and methods circulation. This is a very limited dimension of the dynamic of a science, that we realise having other aspects which worth considering for a reflexive process of population science (Piché and Poirier, 1997). As far as methodology is concerned, the information on “relation as scientific cooperation” was suitable for fitting network model via logistic regression. In literature, it is also shown that multilevel methods are particularly well-suited for the analysis of relations and the changes in these relations (Van Duijn et al., 1999; Snijders, 2001). Our data have a hierarchical structure as ties (first level) are nested within ego-networks (second level). Variables can be referred to the level of the relation (i.e. tie strength, dyad homogeneity in sex, position, geographic area) or to the level of the respondents (i.e. sex, position, geographic area). To study change over time in networks, the whole data set of References Review since 1992 to 1999 should be considered, letting us to compare the science network as in a four wave study. In this case, the dependent variable should be the change in scientific relationship frequency.

References

- Carolyn J. A., Wasserman S., Crouch B. (1999) A p* primer: logit models for social networks, *Social Network*, 21, 37-66.
- Poirier J., Piché V. (1999) Trente ans de recherche explicative en démographie. Réflexions autour des dangers du cloisonnement, in: *Théories, Paradigmes et courants explicatifs en démographie*, Actes de la Chiare Quetelet 1997, Academia-Bruylant L'Harmattan (Eds.), 41-64.
- Snijders T.A.B. (2001), The statistical evaluation of social network dynamics, M.E. Sobel and M.P. Becker (edited by) *Sociological Methodology* 2001, Boston and London, Basil Blackwell.
- Van Duijn M.A.J., van Busschbach J.T., Snijders T.A.B. (1999), Multilevel analysis of personal network as dependent variables, *Social Networks*, 21, 187-209.
- Wasserman S., Pattison P. (1996) Logit models and logistic regression for social networks: I. an introduction to Markov graphs and p*, *Psicometrika*, 61, 3, 301-425.
- Rivellini G., Rizzi E. (2001) *La ricerca sulla popolazione in Italia*, anni 1998-1999. Rassegna Bibliografica, Società Italiana di Statistica, Gruppo di Coordinamento per la Demografia (Cd-Rom).