Exponential Random Graph Model for multivariate networks: an application in knowledge network analysis

Domenico De Stefano and Susanna Zaccarin

Abstract In this contribution, we aim to identify the complex structure of relationships at the base of knowledge and innovation diffusion among actors from different organizations on a given territory. Considering two forms of knowledge-based relations - co-authorship and co-inventions - we fit a multivariate p^* model to capture the variety and the complexity of actor interactions.

Key words: University-industry relations, social network analysis, p^* models

1 Introduction

The last decades have witnessed increasing interest in the network approach as a tool to explore knowledge flows related to innovative processes. Basically, innovative processes take place thanks to complex knowledge interchange between science and industry. Knowledge spillovers from academic research to industry are one of the most important forms of knowledge exchange between these two realms [1]. Individual actors play key roles in determining all these processes, and analysis of their social networks is crucial in understanding how their relations affect the diffusion mechanisms of knowledge. In the analysis of science-technology relationships, only a few proposals have examined multivariate formulations of networks [2]. We argue that, especially at individual level, knowledge production and scientific or technological innovation implies the collective participation of many researchers with differing technical skills and competences, and it is unlikely that these interactions are confined to univariate dyadic exchanges. It is more realistic to assume that, in a scientific community, knowledge flows through various kinds of links which jointly interact in very complex ways (e.g. by informal contacts or by codified knowledge

Dep. of Economics, Business, Mathematics and Statistics, University of Trieste, P.le Europa, 1 - 34127 Trieste, Italy
e-mail: domenico.destefano@econ.units.it, susanna.zaccarin@econ.units.it

transfer and formal collaboration activities), and the complexity of the possible interconnected patterns is proportional to the potential number of knowledge-based relations which may be activated among the actors embedded in those organizations. In this paper we propose a multiplex approach to explain two kinds of these links - co-invention (technology-based relation) and co-authorship (science-based relation) - in a scientific community.

2 Multivariate networks and p^* class of models

The conceptualization of the interactions among multiple knowledge-based relations is represented by a multivariate networks (commonly referred to as multiplex networks). Multiplex networks are multi-relational structures that can be defined as a network $\mathbf{Y}$ composed of a set V of n social actors and a collection R of r social relations that describe how the n actors are connected to each other. The m-th social relations in R represents an univariate network on the actors in V. Each of the r social relationships is intended to express a distinctive relational content (for instance, co-authorship or co-invention ties) not implying their mutual independence. For the m-th social relation is defined a binary variable Y_{ijm}, which is 1 if there is a relational tie of type m between actor i and actor j, and equals 0 otherwise. The m-th univariate network is non-directed if $Y_{ijm} = Y_{jim}$. For directed networks this is not generally true. In this paper we deal with undirected multiple relations.

A fundamental question is related to the identification of the processes guiding the simultaneous functioning of multiple relations within one system. To this end an useful approach consists in identifying the structural (endogenous) and exogenous determinants at the basis of the interactions observed in real networks. The p^* class of models (also known as exponential random graph models or ERGM) allows to analyze interdependencies among network ties. In their most general form, p^* models express the probability of an overall multirelational network structure in terms of parameters associated with particular network substructures. A substructure represents a specific hypothetical configuration of network ties linking a small subset of individuals.

The general form of p^* models is:

$$Pr(\mathbf{Y} = \mathbf{y}) = (1/\kappa) exp(\sum_A \lambda_A \prod_{Y_{ijk \in A}} y_{ijk}) \tag{1}$$

where $\mathbf{Y}$ is a random multivariate network with possible ties Y_{ijm}; $\mathbf{y}$ is a realization of $\mathbf{Y}$ with y_{ijm} denoting an observed tie of type m between actors i and j; A is a subset of possible ties (defining a network substructure of interest); λ_A denotes a parameter associated with the substructure A (that has to be estimated); the terms $\prod_{Y_{ijm \in A}} y_{ijm}$, indicates the network statistic corresponding to the substructure A; and κ is a normalizing constant. Eq.1 states that the probability of observing a network is dependent on the presence of various configurations (substructures) included in the model [4].

For our purpose, the p^* models are used here to analyze interactions among the co-authorship and co-invention relations and in order to verify some specific assumption on the logic that determines the observed multiple network [4].

3 Research hypotheses and model results

This study aims to analyse the interplay between co-inventions and co-authorship ties, observed in a particular scientific community operating in a local environment, i.e. the A-I community in the Trieste area. Interest in this area is due to its very high concentration of (mainly international) research organizations and a lively sector of small R&D firms. In order to retrieve information on the local A-I community, a relational dataset was built by matching lists of actors involved in technology-oriented activities (inventors in patent data) with those of actors involved in science-based activities (authors in scientific publication data). Actors in these lists were identified by their common link with Trieste in the period 2000-2009. As a result of the matching procedure, 129 A-Is were identified: more than one out of four inventors (26%) in the Trieste area is also involved in research-based activities. On average, they each produce about 2 patents (ranging from 1 to 14) and 20 papers (ranging from 1 to 227). As expected, A-Is more frequently work for universities (60%) than in public research organizations - PRO - (18%) and firms (22%). The 63% of A-Is work for organizations located in Trieste.

Data on patents and scientific publications produced by A-Is have been organized into two affiliation matrices of size (129×140) for patents and (129×1509) for publications. Affiliation matrices represent a two-mode network linking individuals to their outputs. These matrices are reorganized into one-mode actors-by-actors networks and transformed to binary data to derive the presence or absence of scientific and technological relations among individual A-Is. From these relations, the multiplex network $\mathbf{Y}$ with $r = 2$ is then defined.

Quite complex mechanisms, determined by both endogenous and exogenous processes, with respect to the network structure, underlie tie formation in multiplex networks. Ideally, these processes can be expressed in terms of p^* model parameters associated with particular network substructures (endogenous factors) and actor covariates (exogenous factors). These substructures are reflected in the model by the value of network statistics which counts both the number of ties in a specific configuration and that of ties by actor covariates. In particular, we formulate several hypotheses in order to describe the possible mechanisms governing the multiplexity of ties in the A-I community according to the observed network and actor characteristics:

- H1: Ties in one network increase the likelihood of tie formation in the other.
- H2: Low tendency toward centralization around prominent A-Is.
- H3: High propensity toward simple and mixed clustering in cohesive subgroups.
- H4: Academic A-Is are more likely to activate multiple ties than non-academic.

- H5: Sharing resources increases the likelihood of multiple tie formation.
- H6: Scientific and technological productivity affects multiple tie formation.

Model results provide evidence of the entrainment of co-invention and co-authorship ties activated in a community in which members mainly act as well-connected peers (H1), according to quite complex forms of transitive closure (H3). Unlike univariate network studies (especially on co-authorship), in this community multiple connections caused by the influence of star actors are quite irrelevant (H2). The significance of the hypothesized endogenous effects supports the idea of a true diffusion process of scientific productions toward applied research fields. Academic A-Is play a key role in this process (H4). Geographical proximity does not seem important for either co-invention or multiple relations, whereas the positive effect of sharing similar resource spaces does affect scientific linkages (H5). Multiple tie formation seems to benefit mainly by the complementarity and non-redundancy of both technical and scientific competences (H6). We have evidence of an interconnected knowledge system, in which scientific achievements are transformed into patentable results (or viceversa). Interestingly, the overlap of both co-invention and co-authorship ties connects otherwise disconnected actors.

The model may be potentially extended to explain more than two relations at once. However, we would like to stress that model estimation, even with only two networks on about 100 actors, is an extremely time-consuming task. In addition, model convergence with adequate goodness-of-fit statistics is very hard to achieve with currently available algorithms. As networks were derived from affiliation data (authors-by-papers and inventors-by-patents), we should add that some of the results must be interpreted carefully. For example, triangulation effects may originate from collaboration on different patents/papers, as well as from collaboration on the same patent/paper. Although the model does not distinguish between the two cases, interpretations of transitivity mechanisms require supplementary exploration. Despite these limitations, the present study provided interesting results in geographically bounded A-Is interactions.

References

1. Breschi M. and Catalini C.: Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. Research Policy **39**, 14–26 (2010)
2. Lee S. and Monge P.: The Coevolution of Multiplex Communication Networks in Organizational Communities. Journal of Communication **61**, 758–779 (2011)
3. Pattison, P. and Wasserman, S.: Logit Model and logistic regression for social networks. II. Multivariate relationships. Ann. British J. Math. Statist. Psych **52**, 169–193 (1999)
4. Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P.: Recent developments in exponential random graph (p*) models for social networks. Social Networks **29**, 192–215 (2007)