A Statistical Mechanics Approach to Immigrant Integration in Emilia Romagna (Italy)

Francesco De Pretis and Cecilia Vernia

Abstract. Integration phenomena are social processes among human beings that take place every day when an autochthone population is experiencing the arrival of new immigrants. Although being a rising phenomenon (involving now over one billion people according to United Nations) which questions societies and policy-makers all over the world, numerical measurements capable to give robust insights over the way immigrant integration occurs are still far from what is usually considered an affordable standard in mathematical and physical sciences. Basing our analysis on previous seminal works, we follow here a statistical physics approach to the analysis of immigrant integration. In specific, we consider a large dataset collected by the Emilia Romagna region office of statistics (Italy), containing information over all marriages occurred amid the regional population during a sixteen years span, from 1995 to 2010. We define as quantifier of integration the percentage of marriages with spouses of mixed origin and we perform several analyses over the dataset, including binning and data fitting. The final outcome consists in an emerging pattern: quantifier's average measurements align around a square root fit when considered with respect to a suitable function of the immigrant density. The theoretical interpretation we offer is that such result agrees with a suitable version of the Curie-Weiss model used in statistical mechanics to describe ferromagnetisms. More explicitly, immigrants living in Emilia Romagna municipalities seem to present mainly imitative behavior's phenomena in making social actions for integration. The result emerged with Emilia Romagna data complies with previous works concerning similar data coming from Spain.

Francesco De Pretis
M2SCS School of Graduate Studies, Università degli Studi di Modena e Reggio Emilia, Italy
e-mail: `francesco.depretis@unimore.it`

Cecilia Vernia
Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università degli Studi di Modena e Reggio Emilia, Italy
e-mail: `cecilia.vernia@unimore.it`

1   Introduction

Integration of immigrants is a political priority in many countries: there are over one billion migrants all over the world, one quarter of which are international migrants [1]. Even though it is not clear how sensitive integration is to an increase of immigrant density and to what extent social interaction goes into higher integration, it is easy to guess that social interaction between immigrants and autochthonous population is a necessary condition for immigrant integration. Curie-Weiss models have been used in the last years in the quest to model social interactions and processes of decision taken by individual human beings [2,3,4,5,6]. In this paper we follow a statistical physics approach to the study of immigrant integration using methods and models already explored in a previous seminal work concerning a large collection of Spanish data [7]. Given a large dataset described in the subsequent section, we focus on a classical quantifier of integration such as the fraction of marriages with spouses of mixed origin (native – i.e. bearing Italian citizenship – and immigrant)

$$M_m = \frac{number\ of\ mixed\ marriages}{number\ of\ marriages}.$$

Within this framework, our goal is a statistical mechanics theory by which the magnitude of the above-mentioned quantifier can be expressed as a function of the density of immigrants, i.e. the ratio between the number of immigrants N_{imm} and the total population $N = N_{imm} + N_{nat}$ where N_{nat} is the number of natives:

$$\gamma = \frac{N_{imm}}{N_{imm} + N_{nat}} \in [0,1]$$

For a better representation of the integration quantifier – based on combinatorial reasoning [7] – we are interested in studying its dependence on the quantity $\Gamma = \gamma(1 - \gamma)$. Afterwards, we seek an empirical function from real data, able to entail the observed collective behavior. As it will be reported in the results section, this work confirms that it is possible to discriminate, using quantitative methods, whether the value of the integration quantifier follows from people acting according to some individual preferences independently of other people (*independent choices*), or whether it follows as a result of social interaction with other ones (*imitative behaviors*). These two opposite cases are described in statistical mechanics theory either as perfect gas of independent particles (in this case, average measurements of the quantifier against Γ follow a linear growth) or interacting theory with possible phase transitions (in this case, average measurements of the quantifier against Γ align around a square root curve).

2   Data Description and Methods

As stated above, the perspective of this work belongs to the statistical mechanics methods used to explain social integration phenomena, starting from the analysis

of real data. More precisely, the work has been centered on a large dataset collected by the Emilia Romagna region office of statistics (Italy), containing information recorded in all Emilia Romagna region municipalities (348 cities) regarding marriages occurred amid the population during a sixteen years span, from 1995 to 2010. In particular, for each municipality, the database provides the reference year, the number of marriages between Italians and foreigners, the number of marriages only between foreigners, the number of marriages only between Italians and the total amount of marriages.

Regarding the sources of data, the Emilia Romagna region dataset was somehow freely downloadable from the regional office of statistics website (data were accessible per municipality at given year, so that techniques of automatic web-contents wrapping have been employed to collect the entire dataset). All data were real (i.e. not estimated) and were subsequently matched with the density of immigrants for each municipality at each given year (information freely retrievable from ISTAT – Italian National Institute of Statistics – sources): the density of immigrants was estimated only for two specific years (1999 and 2000) since the recording of immigrant population was suspended during that period. Given the dimension of the considered dataset, the work can be somewhat inscribed in a *big data exploitation*: to give a rough idea of the computational efforts pursued, around 50.000 data have been processed in order to compute the above described quantifier (y-axis) matched with the density of immigrants (x-axis), producing the scatter plot reported below in Figure 1. It is worth to note that according to prescriptions of a time-independent analysis, data have been plotted together independently from the year they were referring to.
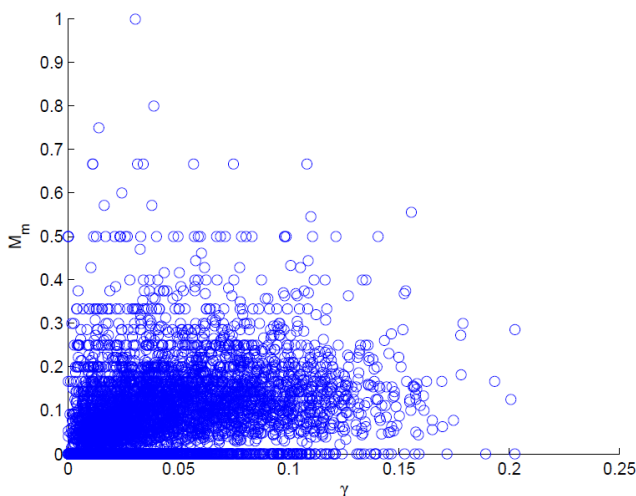
Fig. 1 Raw data versus γ. Blue points represent the fraction of mixed marriages occurred from 1995 to 2010 in all municipalities located in Emilia Romagna region where a percentage γ of migrants is present.

Besides the representation through a scatter plot, we have been interested in the quantifier's average measurements as functions of Γ: since the quantifier M_m is a ratio (i.e. the total number of mixed marriages over the total number of marriages) and we were concerned in looking at patterns in the global scale of the dataset, a natural way to compute averages has been a *mediant* measure. This means that for a given bin of γ, we have computed the ratio between the statistical average of numerators and the statistical average of the denominators. This processing has been performed according to a constant information binning, i.e. each bin contained a fixed number of points. After having compacted the initial information of raw data into bins, a classical procedure of curve fitting has taken place. In particular, following the example of previous work conducted in [7], we have evaluated linear and power functions according to the R^2 coefficient of determination.

3   Results

After having performed the procedures described in the methods section, the following result has been obtained: quantifier's average measurements computed for all data coming from the Emilia Romagna region dataset have fitted around a visible square root pattern with a R^2 coefficient of determination being over 95%, highly reproducing results obtained with similar Spanish data in [7]. Since the analysis of the data density versus Γ shows that only about the 7% of the data are found for Γ greater than 10%, we limit our study below this threshold. It is worth of note that in Emilia Romagna region in 2010 the percentage of immigrants over the total population is about 11%, the highest density of immigrants with respect to any other Italian region.

The result has been verified according to various types of binning (i.e. changing the number of bins) and various families of functions (for instance, linear functions). In the end, a square root fit emerged as the best estimation for the quantifier's average measurements, since with linear and other fittings, the outcomes reported lower R^2 coefficient of determination associated with noisy fits highly depending on the nature of binning.

The mathematical model that supports these results is a generalization of the monomer-dimer model [8] with the addition of an imitative interacting social network component of small world-type [9]. The model, proposed and described in [7], reduces to the classical discrete choice theory [10] (or perfect gas of independent particles) with linear growth of the quantifier as a function of Γ, when imitation is negligible, and to the square root behavior when imitation is dominant. The social network structure explains why the integration starts very close to $\Gamma = 0$ when the choice is dependent on other agent behavior.

Therefore, translated in statistical mechanics terms according to the theoretical interpretation shown in [7], the result of an empirical square root function for the quantifier's average measurements offers an interesting picture of immigrant integration issues in Emilia Romagna region. In specific, even though we do not deal with the possible origins of such cooperative influence, we simply conclude that data suggest that in Emilia Romagna municipalities imitative phenomena mainly take place against the possibility of independent choices carried on by the same immigrants.
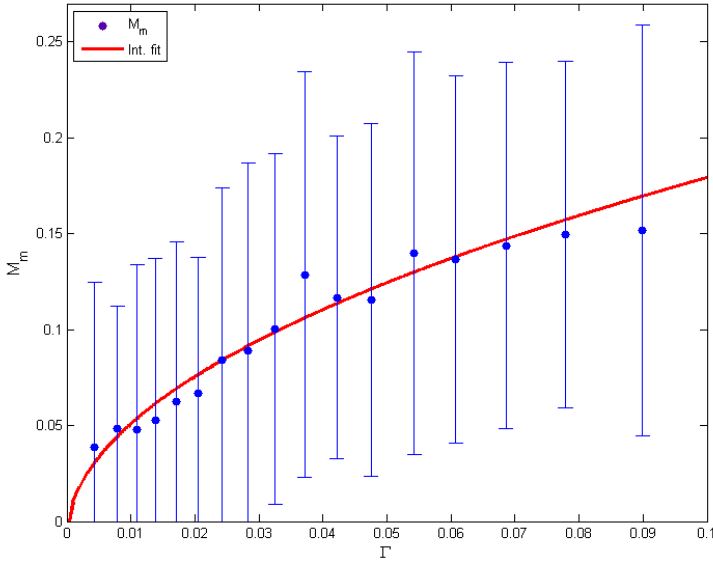
Fig. 2 Emilia Romagna dataset. Dots are average quantities versus Γ, whereas lines denote error bars. Quantifier M_m (blue dots), fraction of mixed marriages occurred from 1995 to 2010 in all the municipalities located in Emilia Romagna region, with the best square root fit (red curve) $a\sqrt{\Gamma} + b$ ($a = 0.5943 \pm 0.0757$, $b = -0.008631 \pm 0.014019$, goodness of fit $R^2 = 0.9529$ computed for $\Gamma < 0.078$). Parameter b evaluation is compatible with the hypothesis that it can be null, as prescribed by the statistical mechanics model we use for results interpretation.

References

1. The Global Approach to Migration and Mobility. EU Report, Commission 743, sec. 1353 (2011)
2. Durlauf, S.N.: Statistical mechanics approaches to socioeconomic behavior. Technical Working Paper, 203, Natl. Bur. Econ. Res (1996)
3. Barra, A., Contucci, P.: Toward a quantitative approach to migrants social integration. Europhys. Lett. 89, 68001, 68007 (2010)
4. Gallo, I.: An equilibrium approach to modelling social interaction. PhD Thesis, Università di Bologna (2009)
5. Contucci, P., Gallo, I., Menconi, G.: Phase transitions in social sciences: two-populations mean field theory. Int. J. Mod. Phys. B 22(14), 1–14 (2008)

6. Barra, A., Contucci, P., Gallo, I.: Parameter Evaluation of a Simple Mean-Field Model of Social Interaction. Mathematical Models and Methods in Applied Science 19, 1427–1439 (2009)
7. Barra, A., Contucci, P., Sandell, R., Vernia, C.: Integration indicators in immigration phenomena. A Statistical Mechanics Perspective (2013),
   `http://arxiv.org/abs/1304.4392`
8. Heilmann, O.J., Lieb, E.H.: Monomers and dimers. Phys. Rev. Lett. 24(25), 1412 (1970)
9. Watts, D.J., Strogatz, S.H.: Collective dynamics of small world networks. Nature 393, 6684 (1998)
10. McFadden, D.: Economic choices. The Amer. Econ. Rev. 91 (2001)