



ARTICLE

DOI: 10.1057/s41599-018-0097-5

OPEN

Social interaction effects on immigrant integration

Elena Agliari¹, Adriano Barra², Pierluigi Contucci³, Andrea Pizzoferrato^{4,5} & Cecilia Vernia⁶

ABSTRACT In recent years Italy has been involved in massive migration flows and, consequently, migrant integration is becoming a urgent political, economic and social issue. In this paper we apply quantitative methods, based on probability theory and statistical mechanics, to study the relative integration of migrants in Italy. In particular, we focus on the probability distribution of a classical quantifier that social scientists use to measure migrant integration, that is, the fraction of mixed (natives and immigrants) married couples, and we study, in particular, how it changes with respect to the migrant density. The analysed dataset collected yearly by ISTAT (Italian National Institute of Statistics), from 2002 to 2010, provides information on marriages and population compositions for all Italian municipalities. Our findings show that there are strong differences according to the size of the municipality. In fact, in large cities the occurrence of mixed marriages grows, on average, linearly with respect to the migrant density and its fluctuations are always Gaussian; conversely, in small cities, growth follows a square-root law and the fluctuations, which have a much larger scale, approach an exponential quartic distribution at very small densities. Following a quantitative approach, whose origins trace back to the probability theory of interacting systems, we argue that the difference depends on how connected the social tissue is in the two cases: large cities present a highly fragmented social network made of very small isolated components while villages behave as percolated systems with a rich tie structure where isolation is rare or completely absent. Our findings are potentially useful for policy makers; for instance, the incentives towards a smooth integration of migrants or the size of nativist movements should be predicted based on the size of the targeted population.

¹Dipartimento di Matematica, Sapienza Università di Roma, Roma, Italy. ²Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy.

³Dipartimento di Matematica, Università di Bologna, Bologna, Italy. ⁴Mathematics Institute, University of Warwick, Coventry, UK. ⁵The Alan Turing Institute, London, UK. ⁶Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Modena, Italy. Correspondence and requests for materials should be addressed to P.C. (email: pierluigi.contucci@unibo.it)

Introduction

Systems made of a large number of components can be suitably analysed with probability theory and statistical mechanics formalism. This comes from the fact that each component can be mapped into a random variable and the behaviour of the entire system encoded on their joint probability distribution. The simplest, although somehow idealised, case is when the random variables are mutually independent. In statistical physics this is often referred to as a *perfect gas* or, in general, as a system of non-interacting particles where their global behaviour is fully and easily deducible from that of the single ones. From the mathematical point of view a perfect gas is described by a joint probability distribution that is the product of the probability distribution of each particle. Under very general assumptions, when the particles are of the same type and have each a regular distribution, their large sum converges to a smooth function of the natural parameters and, suitably normalised, can be proved to have Gaussian fluctuations, according to the well known Central Limit Theorem (Feller, 1960).

Real systems, nevertheless, are rarely described *tout court* by such an elementary scheme. This is because interaction among parts is ubiquitous and the independence among the components is rather an exception than the rule. Still, it has been understood that, even in the presence of interaction, systems display *generically* a smooth behaviour and Gaussian fluctuations, apart from special points in the parameter space. Those points, called *critical points*, display non-Gaussian fluctuations and their exhaustive comprehension and classification is among the main challenges of probability theory and statistical physics (Liggett, 1985; Parisi, 1988). Also within the social science, the relevance of the interaction among agents is a growing focus of research (Scheinkman, 2018; Horst and Scheinkman, 2006; Glaeser and Scheinkman, 2001; Bialek et al., 2014; Brock and Durlauf, 2001a, b; Durlauf, 1999) that is progressively generalising the original social choice paradigm (McFadden, 2001) based on the independent agents assumption.

In this paper, we investigate a data set collected by the Italian National Institute of Statistics (ISTAT) for the years 2002 to 2010 on the *social choice* (Weber, 1978) (for a native) of marrying a person from the host country (i.e., Italy) or from a different one. This quantifier is used, among other classical ones (Portes and Sensenbrenner, 1993; Rannala and Mountain, 1997; Agliari et al., 2014; Barra et al., 2014), to study the level of integration of migrants. The fraction of mixed marriages m_{mix} , collected each year and for each municipality, is studied vs. the fraction γ of immigrants on the total population. While previous studies of analogous databases from Spain (Barra et al., 2014; Agliari et al., 2014; Barra et al., 2016), France, Germany and Switzerland (Agliari et al., 2015), all reported an average smooth law for the evolution of mixed marriages vs. the migrant's percentage (alternating a square-root behaviour vs. a linear growth), the Italian database does not show the existence of an underlying average law if analyzed as a whole: in other terms, it is not possible to map the statistical sample into a well defined function $m_{\text{mix}}(\gamma)$. Since this feature is generally the signature of a mixture of two, or more, different phenomena that need to be disentangled we split the database into *large* and *small* municipalities with respect to a proper tuning of a threshold θ over the population size. In this case, analyzing separately the two resulting ensembles, a clear functional dependence was recovered for each of them matching two behaviours: for large municipalities the quantifier m_{mix} grows linearly with γ (i.e., $m_{\text{mix}} \propto \gamma$), while for small municipalities a *square-root* function emerges (i.e., $m_{\text{mix}} \propto \sqrt{\gamma - \gamma_c}$). In this last case γ_c turns out to be positive and close to zero. By further analyzing the quantifier probability distribution around the critical point, we find that large

municipalities display Gaussian fluctuations while small municipalities fluctuate according to a quartic exponential distribution close to γ_c . To further confirm our findings we performed the same analysis by splitting municipalities according to their relative population densities, rather than their size, hence in low and high-density areas and we found analogous results. While this is partly due to the natural correlation that the larger the municipalities, the higher the density of people they contain, a detailed inspection of this point was necessary (see Clark (1951) and “Analysis of the data distributions” in the [Supplementary Information](#) for a discussion).

The previous results can then be interpreted in terms of a probabilistic model of monomer-dimer type, the mean-field version of a class of statistical mechanical models used in condensed-matter statistical physics to describe the deposit of diatomic molecules on lattices (Heilmann and Lieb, 1972). The dimer corresponds here to a married couple and the monomer to the unmarried person. The monogamic rule of the social setting is the equivalent of the forbidden configuration made of two dimers on the same vertex. Our interpretation is based on a series of rigorous mathematical works (Alberici et al., 2014a, b; Alberici and Contucci, 2014; Alberici et al., 2015, 2016a, b) where it has been shown that, in the presence of an imitative interaction among vertices (monomers), the model displays a phase transition: the dimer densities have a square-root growth by the critical point, where quartic exponential fluctuations are observed at the scale $N^{3/4}$ (rather than quadratic at the usual scale $N^{1/2}$, i.e., the standard Gaussian scenario). This result is also typical of a large class of mean-field ferromagnetic spin models whose behaviour was understood in the works (Ellis and Newman, 1978a, b; Ellis et al., 1980; Ellis and Rosen, 1982). The imitative interaction, from a sociological modelistic point of view (Nowak, 2006; Hauert and Doebeli, 2004), is seen as the trust-bond among two people. In the works (Barra et al., 2014; Gallo et al., 2009) it was seen that this type of interaction is strong enough to cause, when the interaction graph is percolated, a mean-field phase transition with square-root singularity. This allows us to conclude that in small municipalities the trust social network is percolated while it is fragmented in large municipalities thus confirming classical theories on alienation (Durkheim, 1897).

Our conclusions lend to sociological studies and are potentially useful for policy makers. The incentives towards a smooth integration of migrants can in fact be predicted to be proportional to the size of the targeted population for sparse networks (large cities) while are sensibly smaller for percolated ones (Burioni et al., 2015).

Database and observables

The source of our database is the ISTAT (Italian National Institute of Statistics). We consider the yearly collected data in the time window 2002–2010 for 8100 *municipalities* (*comuni*, the smallest administrative units of the Italian territory), distributed all over the country. For every municipality the resident population is provided, divided into native citizens and immigrants. Data for the marriages are split into three types: couples which are composed by two Italians, or one Italian and one immigrant (*mixed marriage*), or two immigrants. ISTAT also provides information about the surface S (in Km^2) of each municipality (ISTAT, 2018), in such a way that the densities (ρ , people per Km^2) for native citizens, immigrants and total population can be deduced.

Labelling with i each municipality, and with t each year we define:

- $M(i,t)$: the total number of marriages (this includes marriages where partners are both natives, or both foreign-born, or mixed)
- $M_{\text{mix}}(i,t)$: the number of mixed marriages;
- $m_{\text{mix}}(i,t)$: the fraction of mixed marriages ($m_{\text{mix}}=M_{\text{mix}}/M$);
- $N_{\text{nat}}(i,t)$: the native resident population;
- $N_{\text{imm}}(i,t)$: the immigrant resident population;
- $N(i,t)$: the overall population ($N = N_{\text{nat}}+N_{\text{imm}}$);
- $\gamma(i,t)$: the fraction of immigrants on the total population ($\gamma = N_{\text{imm}}/N$);
- $\Gamma(i,t)$: the fraction of potential cross-links between the two subsets of the population ($\Gamma = \gamma(1-\gamma)$);
- $S(i)$: the surface of the municipality (independent of time);
- $\rho(i,t)$: the total population density ($\rho = N/S$).

As anticipated, the index of municipalities i ranges from 1 to 8100, while the index t ranges from 1 to 9 for the years from 2002 to 2010. During the considered time window the resident population and the number of marriages evolve as summarized in Fig. 1 (panels a–c).

In the following, we will focus on the averages and on the fluctuations of the (suitably normalized) number of mixed marriages as functions of Γ ,¹ and we can therefore merge the data entries into a unique catalogue, regardless of their coordinates in space and time, ordering them by increasing values of Γ (see Fig. 1, panel d). Such raw data are then properly “binned” over Γ in order to highlight a possible functional dependence. We investigated two binning procedures: *constant information* and *constant step*. With the first the intervals on Γ are chosen in such a way to include a fixed number of points. The resulting set of averaged data will have a non-uniform spacing, with larger width where data are less dense. The second has intervals of the same size but the statistical robustness of each bin varies. We thoroughly checked the quantitative consistence of the results of the two methods emphasising the use of the first for the analysis at small Γ and of the second for the whole histograms and related distributions.

We finally notice that our analysis goes up to $\Gamma \approx 0.1$. Beyond that value there is <5% of the data and, beside us being mainly

interested on the neighbour of $\Gamma = 0$, its statistical robustness would be insufficient for our purposes.

Analysis of mean values

The analysis of the mean values is performed on the quantifier m_{mix} as a function of Γ according to the binning procedures explained in the previous section. The advantage of normalising the mixed marriages with the total number of marriages is two-fold. First, it takes care of the finite size fluctuations due to the fact that the database includes a mixture of municipalities of all sizes. Second, it allows us to momentarily skip the search of the proper *volume* of the phenomenon, i.e., the normalisation scale to observe the intensive quantities and their averages through the law of the large numbers as well as their fluctuations and their limiting behaviour.

Our first attempt to identify a law $m_{\text{mix}}(\Gamma)$, if any, has been to consider the whole database at once, i.e., with all the municipalities included. In this case the average values turned out to be irregular in Γ and unstable with respect to the binning procedures. The lack of a functional relation between the observable and the parameter Γ can be interpreted as the signature of the mixture of two, or more, different laws to be disentangled.

We identified such laws according to two (correlated) parameters: the population size of the municipality and the density (per unit area) of its population. What we found is that $m_{\text{mix}}(\Gamma)$ grows linearly for large size (or high-density) cities

$$m_{\text{mix}}(\Gamma) \propto \Gamma \tag{1}$$

while smaller cities (or less dense ones) display a square-root growth

$$m_{\text{mix}}(\Gamma) \propto \sqrt{\Gamma - \Gamma_c}. \tag{2}$$

The two different growth laws have been observed (Barra et al., 2014; Agliari et al., 2014, 2015; Contucci and Sandell, 2015, 2016) and successfully explained (Alberici et al., 2014a, b; Alberici and Contucci, 2014; Alberici et al., 2015; McFadden, 2001; Brock and Durlauf, 2001a, b; Agliari and Barra, 2011; Barra and Agliari,

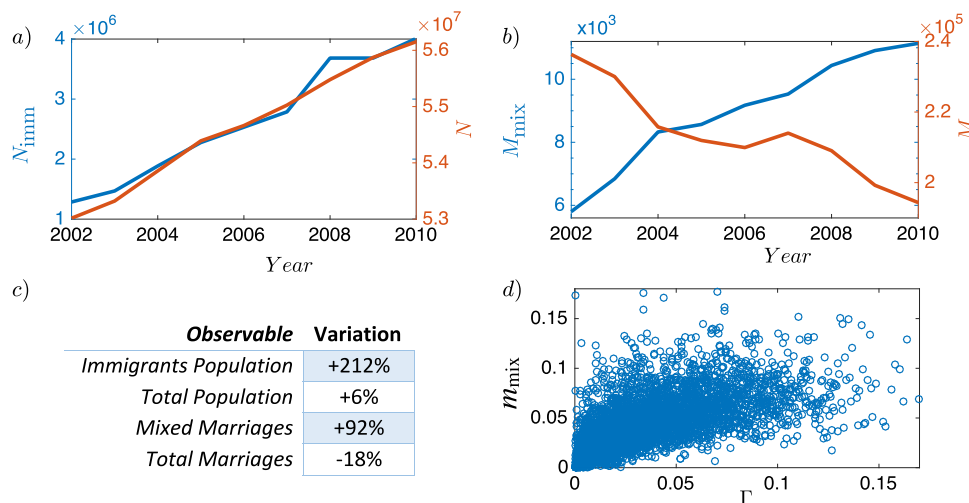


Fig. 1 Annual trends of population and marriage indicators from the data set. In **a** we show the temporal evolution of the number of resident immigrants and of the total Italian population (time frame 2002–2010). As can be seen by direct inspection, the immigration is increasing during the years with a rate higher than the total population growth (see also the table in **c**). Since the municipality area is constant in this time window, we do not show the population density as it is simply obtainable via a rescaling by the total area and thus it has the same monotonicity as the latters. In **b**, we show the same temporal evolution for marriages: note that mixed and total marriages have opposite monotonicity. In **c** we summarize the variations of the observables considered along the time window analyzed. Finally, in **d** we show a scatter plot for m_{mix} vs. Γ

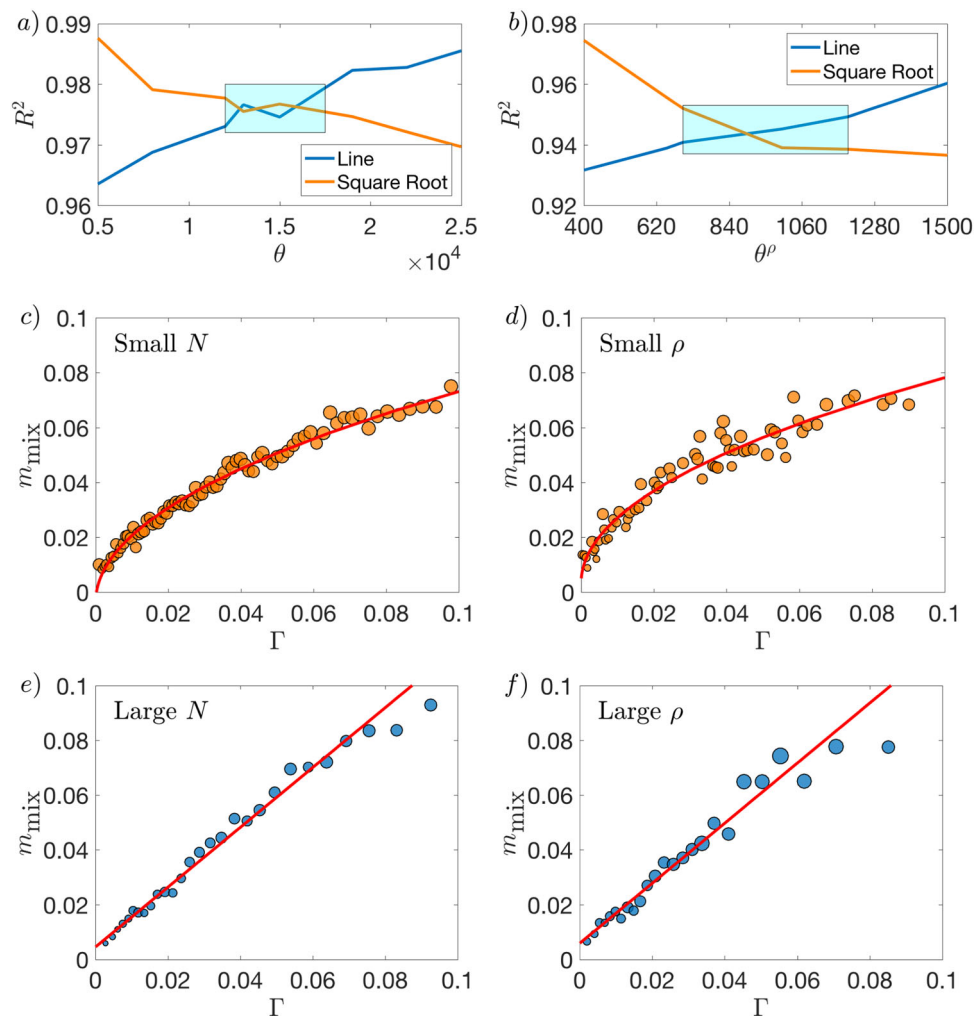


Fig. 2 Analysis of the evolution of the mean value of m_{mix} vs. the fraction of cross-links Γ . In **a** we show the R^2 values obtained when fitting “small” (orange line, square-root fit) and “large” (blue line, linear fit) municipalities, where small and large is meant according to the trial threshold θ over the population N . The dashed zone identifies where the two curves intersect each other, determining the optimized criterion by which we split the whole database in two subsets, one for each type of city. **b** Presents the same information of **a** but focusing on densities rather than their extensive counterparts. That is, the trial threshold θ distinguishes between “sparse” (orange line, square-root fit) and “dense” (blue line, linear fit) municipalities. In **c**, **e** we report data (bullets) and the best fits (solid lines) for the small and the large municipalities, respectively; in the former the best fit is given by a square-root function with $R^2 = 0.99$ and in the latter the best fit is given by a linear function with $R^2 = 0.98$. These results were corroborated by checking that, in a log-log scale, the best fits are provided by linear functions with slope consistent with $1/2$ and 1 , respectively. **d**, **f** are built analogously, using density rather than population to discriminate between small and large municipalities, and the best fits are given by, respectively, a square-root function with $R^2 = 0.98$ and a linear function with $R^2 = 0.96$. Again, we successfully checked fits also in a log-log scale

2012; Barra and Contucci, 2010; Agliari et al., 2008) in terms of different social network structures (see also the section “A review of the statistical mechanical model” in the [Supplementary Information](#) for a short but self-contained treatment). The first type of behaviour can be associated to a lack of citizen’s proximity interaction, namely a lack of a (percolated) *trust network*² of mutual influences (Granovetter, 1983; Watts and Strogatz, 1998). This is also an implicit and indirect confirmation of the general sociological theories about alienation and anomie in large cities (Durkheim, 1897). Conversely, the second type of behaviour typically emerges in the presence of *citizen’s interaction* on a percolated network.

In order to identify the two subsets of municipalities, we analysed the coefficient of determination for the linear and for the square-root fit, as a function of a trial threshold used to split the database into two complementary ones, according to either the population size or the population density. It turns out that

(see Fig. 2, panel a) R_{lin}^2 and R_{sqr}^2 are, respectively, monotonically increasing and decreasing, thus identifying a crossover regime that we take as threshold: for the size $\theta_c \sim 15000$ people and for the density $\theta_c^{(\rho)} \sim 1000$ people per Km^2 . The two thresholds are clearly correlated (see “Analysis of the data distribution” in the [Supplementary Information](#)) since high densities are usually reached only on large cities. It is worth to mention that, as explained in the [Supplementary Information](#) (see “Analysis of the data distribution”), the critical thresholds θ_c and $\theta_c^{(\rho)}$ are not far away from their respective medians. This guarantees that the data sets of each bipartition are robust enough as expected a priori since otherwise we would have seen a clear functional law even without splitting the original one.

With the two databases identified by large and small cities, or cities at high or low density, we can finally analyse the data sets and check the behaviour of the quantifier in each of them: results are shown in Fig. 2 panels c and e, for small and large cities,

respectively. They display a clear square-root scaling in the first case and a linear one in the second. Similarly, for sparse and dense cities, respectively, the results are shown in Fig. 2 panels d and f (see also “Statistical analysis and robustness tests for mean values” in the [Supplementary Information](#)).

To summarise, the analysis above has outlined two key variables (i.e., population size and density), intrinsically connected, responsible for the heterogeneity among municipalities, when looked in terms of mixed marriages. In particular, for both the variables, we found a critical threshold according to which homogenous subsets of municipalities can be determined and a percolative or non-percolative regime for the interaction identified.

Analysis of fluctuations

In the previous section, we analysed the behaviour of the first moment of the fraction of number of mixed marriages and we found the existence of two different laws for the quantity m_{mix} versus Γ . Here, we extend our investigation to its second moment, i.e., the fluctuations, and also to the type of its limiting distribution.

Such analysis, beside being crucially relevant on itself, is of fundamental importance to validate the theoretical picture that we advanced. Our model, in fact, predicts precise quantitative differences in the behaviour of the fluctuations around the critical point Γ_c (see “A review of the statistical mechanical model” in the [Supplementary Information](#)). In the absence of a percolating-interaction, fluctuations should always be Gaussian. When the interaction is strong and percolating instead we expect Gaussian fluctuations away from the critical point, while in the vicinity of the critical point the model predicts a quartic exponential limiting distribution.

In order to compare empirical data with theoretical laws a paramount step to be solved is the identification of the proper “size” of the interacting system. Dealing with a matching problem among two groups of sizes N_{nat} and N_{imm} , the natural candidate definition, to be verified and tested, seems to be

$$\frac{N_{\text{nat}}N_{\text{imm}}}{N} = \Gamma N := \Omega. \quad (3)$$

To this purpose we introduce an intensive random variable, i.e., the ratio

$$\mu_{i,t}(\beta, \Gamma, \Omega) = \frac{M_{\text{mix}}(i, t)}{\Omega(i, t)^\beta}, \quad (4)$$

where the exponent β at the denominator is a trial positive number to be experimentally identified through the Law of Large Numbers. The filtering procedure to be implemented in this case, unlike in the analysis of the previous section where the size problem was avoided by studying the ratio of two extensive observables, must be able to select, within the data set, all those sub-clusters of data that are approximatively at the same Γ and at the same Ω . To this aim we mesh the (Γ, Ω) space in such a way that municipalities falling within the k -th region $K \equiv [\Gamma_k, \Gamma_k + \delta\Gamma] \times [\Omega_k, \Omega_k + \delta\Omega]$ can be considered as *homogeneous*, namely municipalities corresponding to observables $\Gamma(i, t)$ and $\Omega(i, t)$ which fulfill, simultaneously, $\Gamma_k \leq \Gamma(i, t) < \Gamma_k + \delta\Gamma$ and $\Omega_k \leq \Omega(i, t) < \Omega_k + \delta\Omega$, constitute the k -th sample. Of course, the partition of the space (Γ, Ω) , that is, ultimately, the choice of $\delta\Gamma$ and $\delta\Omega$, must provide a proper trade-off, which ensures that each sample is statistically large enough but, still, relatively homogeneous. We can then define the average $\overline{M}_{\text{mix}}(\Gamma_k, \Omega_k)$ over the k -

th sample as follows

$$\overline{M}_{\text{mix}}(\Gamma_k, \Omega_k) \equiv \frac{\sum_{i,t: [\Gamma(i,t), \Omega(i,t)] \in K} M_{\text{mix}}(i, t)}{|K|}, \quad (5)$$

where, with some abuse of notation, in the denominator there is the cardinality of the k -th sample. Basically, $\overline{M}_{\text{mix}}(\Gamma_k, \Omega_k)$ represents the average number of marriages within those municipalities which share the same *effective size* and the same fraction of immigrants. Analogously, the normalised fluctuations (around $\mu_{i,t}$) are defined as

$$\Delta_{i,t}(\alpha, \Gamma_k, \Omega_k) = \frac{|M_{\text{mix}}(i, t) - \overline{M}_{\text{mix}}(\Gamma_k, \Omega_k)|}{\Omega(i, t)^\alpha}, \quad (6)$$

where the exponent α at the denominator is another tuneable parameter to be experimentally fixed through the Central Limit Theorem away from the critical point and to a possibly different law, if any, in its vicinity. Hereafter we shall drop the subindex k for the sake of simplicity. More explicitly, we are left to prove the existence of two suitable exponents $\overline{\alpha}$ and $\overline{\beta}$ such that

$$\mu_{i,t}(\overline{\beta}, \Gamma, \Omega) = \mathcal{O}(1), \quad (7)$$

$$\Delta_{i,t}(\overline{\alpha}, \Gamma, \Omega) = \mathcal{O}(1), \quad (8)$$

namely, when $\alpha = \overline{\alpha}$ and $\beta = \overline{\beta}$, μ and Δ shall not exhibit any dependence on the system size thus making possible direct comparison with statistical-mechanical theories. In particular, we are interested in identifying any possible breakdown of the Eq. 8 as this could provide the signature of a possible critical behaviour (expected as $\Gamma \rightarrow \Gamma_c \sim 0$). For this reason a separate analysis shall be conducted by focusing on the small municipalities database at progressively smaller Γ because those are, respectively, the phase and point at which a phase transition has been found on the previous section. The results for μ are summarised in Fig. 3 where the choice of the trial size is indeed confirmed since, for $\beta = \overline{\beta} \approx 1$, data points corresponding to different sizes Ω merge together. Similarly, the results for Δ are summarised by Fig. 4, where the set of data corresponding to different sizes Ω collapse when $\alpha = \overline{\alpha} \approx 0.5$, confirming the Central Limit Theorem.

Finally, we investigate the behaviour of $\Delta_{i,t}(\alpha, \Gamma, \Omega)$ in the vicinity of $\Gamma = 0$. We expect that in this region there exists a suitable exponent α_c , that keeps the quantity finite for large sizes, namely the fluctuations of M_{mix} pertaining to different values of Ω collapse to a finite value (and do not grow in N). We expect, moreover, that

$$0.5 \approx \overline{\alpha} < \alpha_c < \overline{\beta} \approx 1, \quad (9)$$

our theory suggesting $\alpha_c = 0.75$ (Alberici et al., 2016a). To check this we look at the distribution of the raw data falling in three Γ -bins approaching $\Gamma = 0$ for $\alpha = 0.5$ and $\alpha = 0.75$. The results are shown in Fig. 5. When the normalisation exponent is $\alpha = 0.5$ (left panels) and the data are fitted against a Gaussian distribution, the quality of the fit progressively deteriorates while Γ approaches 0. In particular one sees that the coefficient R^2 goes from 0.981 in panel e to 0.956 in panel a, which suggests that near $\Gamma = 0$ the Gaussian regime for the fluctuations does not hold any longer. When instead we choose $\alpha = 0.75$ and fit the data against a quartic exponential (see the right panels of Fig. 5) we observe an opposite trend i.e., approaching $\Gamma = 0$ the quality of the fit is progressively improving. Namely $R^2 = 0.999$ in panel b, while far from the critical point $R^2 = 0.586$ in panel (f). This picture is in complete agreement with the statistical-mechanics prediction (Alberici et al., 2015, 2016b; Alberici and Mingione, 2017) especially considering the finite size effects that real data carry with

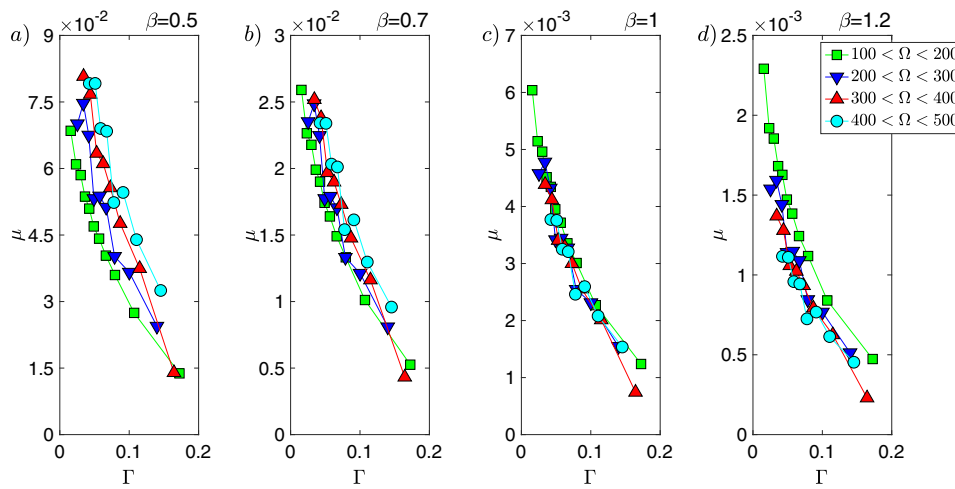


Fig. 3 Collapse of μ and estimate of β . In the analysis presented in this figure we partitioned the (Γ, Ω) space by fixing $\delta\Gamma = 0.008$ and $\delta\Omega = 100$, with $\Omega_1 = 100$ (green squares), $\Omega_2 = 200$ (blue triangles), $\Omega_3 = 300$ (red triangles), and $\Omega_4 = 400$ (cyan bullets). We focused on this range of sizes Ω since it provides a better statistics. For each region we collected the pertaining values for $\mu_{i,t}$, which are then averaged. Different panels correspond to different choices of β : $\beta = 0.5$ (a), $\beta = 0.7$ (b), $\beta = 1$ (c), $\beta = 1.2$ (d). Our estimate $\bar{\beta} \approx 1$ stems from the minimal spread of the curves in **c**

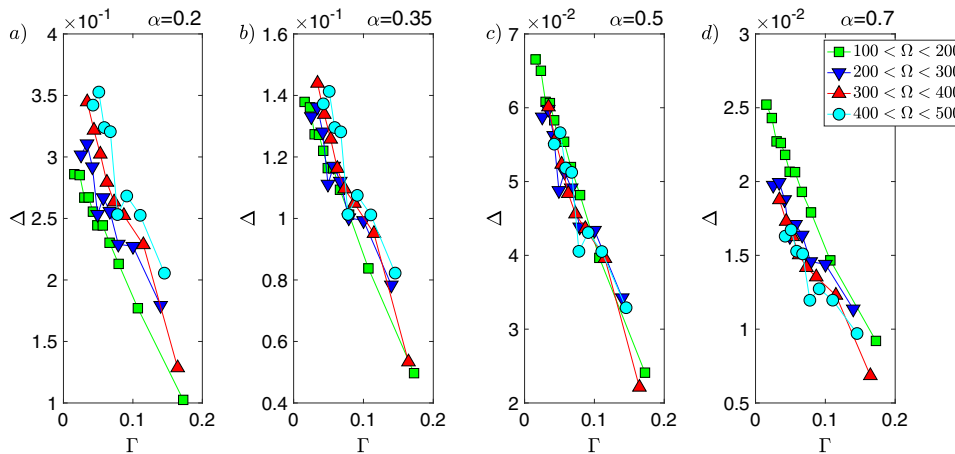


Fig. 4 Collapse of Δ and estimate of α . The observable $\Delta_{i,t}(\beta, \Gamma, \Omega)$ is measured for different choices of Ω and of Γ and the related raw data are properly binned over Γ and over Ω . Different symbols correspond to different choices of Ω . The (Γ, Ω) mesh as well as the choice of symbols is the same as in Fig. 3. Different panels correspond to different choices of α : $\alpha = 0.2$ (a), $\alpha = 0.35$ (b), $\alpha = 0.5$ (c), $\alpha = 0.7$ (d). Our estimate $\bar{\alpha} \approx 0.5$ stems from the minimal spread of the curves in **c**

them. To test the compatibility of those effects with the model we have made numerical simulations, reported in the [Supplementary Information](#) (see “Further robustness tests on fluctuations”), that show how the boundedness of the fluctuations with the normalisation $\alpha_c = 3/4$ is consistent with those found on real data.

Conclusions

The study of migration fluxes and their relative integration has a long tradition in Italy, a country that has been constantly exposed to immigration phenomena. There are therefore excellent accounts from the classical sociological perspectives (Colombo, 2012; Pastore and Ponzio, 2016). Our approach in this study has been based on data and on mathematical modelling and data analysis methods developed within the hard sciences, in particular statistical physics. What we have learnt is that large cities and small villages have very different mechanisms to integrate the immigrants, as far as the mixed marriage observable is concerned. While in large cities the growth of integration follows linearly the increasing presence of immigrants, in villages the functional dependence is of square-root type. This result confirms (see also

(Barra et al., 2014; Agliari et al., 2015)) that the collective behaviour of integration phenomena obeys to either two of the social paradigms: independent agents vs. correlated ones. One further confirmation step was obtained in the present work, i.e., the verification that also at the fluctuation level one can see the difference of the two phases. At very small immigration densities the central limit theorem is violated and a different limiting distribution appears, at much larger scales $N^{3/4}$ instead of $N^{1/2}$, that turns out to be a quartic exponential. The idea that statistical mechanics could shed some light on the social sciences was discussed by Durlauf almost two decades ago (Durlauf, 1999). The results reached in this work provide an instance of concrete evidence in favour of that idea by identifying a matching between data and theory. The conclusion that we reached concerning the percolation of the trust network and the size of the municipality provides very suggestive forecasts on different phenomena. One of them is the vote of the country on sensitive topics like *pro* or *anti* immigration policies. Small villages (percolated case) will display highly polarised votes almost totally concentrated on one of the two positions. Such voting results are quite difficult to

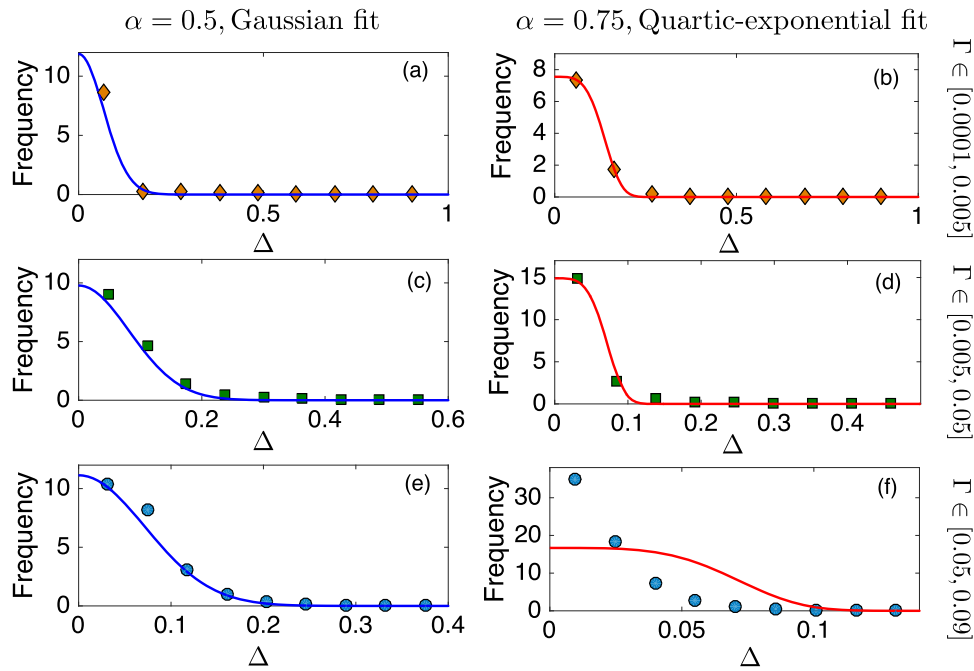


Fig. 5 Histogram of $\Delta_{it}(\alpha, \Gamma, \Omega)$ for small cities ($N < 10000$) for $\alpha = 0.5$ (left panels) and $\alpha = 0.75$ (right panels). Moreover, we consider different intervals of Γ : $[0.0001, 0.005]$ (upper panels), $[0.005, 0.05]$ (central panels), $[0.05, 0.09]$ (lower panels). According to the statistical-mechanics theory, the data in the left panels are fitted with a Gaussian distribution $y = e^{-x^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$ while those in the right panels with a quartic exponential distribution $y = e^{-x^4/\sigma^4}/[2\sigma\Gamma(5/4)]$. The Gaussian fits (blue lines in the left panels) with one free parameter σ give a decreasing coefficient of determination R^2 from **e** to **a** showing that approaching $\Gamma = 0$ a critical point is expected. In particular, the parameters of the fits are: $\sigma = 0.10 \pm 0.03$ $R^2 = 0.956$ (**a**), $\sigma = 0.12 \pm 0.01$ $R^2 = 0.976$ (**c**), $\sigma = 0.10 \pm 0.01$ $R^2 = 0.981$ (**e**). The quartic exponential fits (red lines in the right panels) with one free parameter σ provide instead an increasing R^2 from **f** to **b**, confirming the theoretical scenario that requires that the correct normalization at the critical point has exponent $3/4$. The parameters of the fits are: $\sigma = 0.15 \pm 0.01$, $R^2 = 0.999$ (**b**), $\sigma = 0.078 \pm 0.007$, $R^2 = 0.991$ (**d**), $\sigma = 0.08 \pm 0.01$, $R^2 = 0.586$ (**e**)

predict based on preliminary polls. Large cities instead (fragmented network) will display a voting result that comes from the superposition of the two opinions. A well done survey, therefore, will correctly predict the outcome of the elections.

Received: 22 January 2018 Accepted: 12 March 2018
 Published online: 08 May 2018

Notes

- 1 Since $\Gamma = \gamma/(1-\gamma)$, for small percentages of migrants γ (i.e., in a neighborhood of zero, which is the main focus of the present work), Γ and γ are basically indistinguishable.
- 2 With the term *trust network* we mean not just the usual social network of acquaintances, but rather a subnetwork where only the most significant connections are retained in such a way that the behaviour of nearest neighbours is actually strongly correlated.

References

Agliari E, Barra A (2011) A Hebbian approach to complex-network generation. *EPL (Europhys Lett)* 94(1):10002
 Agliari E, Barra A, Camboni F (2008) Criticality in diluted ferromagnets. *J Stat Mech: Theory Exp* 2008(10):P10003
 Agliari E, Barra A, Contucci P, Sandell R, Vernia C (2014) A stochastic approach for quantifying immigrant integration: the spanish test case. *New J Phys* 16(10):103034
 Agliari E, Barra A, Galluzzi A, Javarone MA, Pizzoferrato A, Tantari D (2015) Emerging heterogeneities in italian customs and comparison with nearby countries. *PLoS ONE* 10(12):e0144643
 Alberici D, Contucci P (2014) Solution of the monomer-dimer model on locally tree-like graphs. rigorous results. *Commun Math Phys* 331(3):975–1003
 Alberici D, Contucci P, Mingione E (2014a) The exact solution of a mean-field monomer-dimer model with attractive potential. *EPL (Europhys Lett)* 106(1):10001

Alberici D, Contucci P, Mingione E (2014b) A mean-field monomer-dimer model with attractive interaction: Exact solution and rigorous results. *J Math Phys* 55(6):063301
 Alberici D, Contucci P, Mingione E (2015) A mean-field monomer-dimer model with randomness: Exact solution and rigorous results. *J Stat Phys* 160(6):1721–1732
 Alberici D, Contucci P, Fedele M, Mingione E (2016a) Limit theorems for monomer-dimer mean-field models with attractive potential. *Commun Math Phys* 346(3):781–799
 Alberici D, Contucci P, Mingione E (2016b) Non-Gaussian fluctuations in monomer-dimer models. *EPL (Europhys Lett)* 114(1):10006
 Alberici D, Mingione E (2017) Two populations mean-field monomer-dimer model. arXiv preprint arXiv:1706.07356
 Barra A, Agliari E (2012) A statistical mechanics approach to Granovetter theory. *Phys A: Stat Mech its Appl* 391(10):3017–3026
 Barra A, Contucci P (2010) Toward a quantitative approach to migrants integration. *EPL (Europhys Lett)* 89(6):68001
 Barra A, Contucci P, Sandell R, Vernia C (2014) An analysis of a large dataset on immigrant integration in spain. the statistical mechanics perspective on social action. *Sci Rep* 4:4174
 Barra A, Galluzzi A, Tantari D, Agliari E, Requena-Silvente F (2016) Assessing the role of migration as trade-facilitator using the statistical mechanics of cooperative systems. *Palgrave Commun* 2:16021
 Bialek W, Cavagna A, Giardina I, Mora T, Pohl O, Silvestri E, Viale M, Walczak AM (2014) Social interactions dominate speed control in poisoning natural flocks near criticality. *Proc Natl Acad Sci* 111(20):7212–7217
 Brock WA, Durlauf SN (2001a) Discrete choice with social interactions. *Rev Econ Stud* 68(2):235–260
 Brock WA, Durlauf SN (2001b) Interactions-based models. *Handb Econ* 5:3297–3380
 Burioni R, Contucci P, Fedele M, Vernia C, Vezzani A (2015) Enhancing participation to health screening campaigns by group interactions. *Sci Rep* 5:9904
 Clark C (1951) Urban population densities. *J R Stat Soc Ser A (General)* 114(4):490–496
 Colombo A (2012) Fuori controllo?: miti e realtà dell’immigrazione in Italia. Il mulino, Bologna
 Contucci P, Sandell R (2015) How integrated are immigrants? *Demogr Res* 33:1271

- Contucci P, Sandell R (2016) How immigrant integration unfolds. Elcano Royal Institute for International and Strategic Studies, ARI 17, Madrid, Spain
- Durkheim E (1897) *Le suicide: étude de sociologie*. Flix Alcan, Paris
- Durlauf SN (1999) How can statistical mechanics contribute to social science? *Proc Natl Acad Sci* 96(19):10582–10584
- Ellis RS, Newman CM (1978a) Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Probab Theory Relat Fields* 44(2):117–139
- Ellis RS, Newman CM (1978b) The statistics of Curie-Weiss models. *J Stat Phys* 19(2):149–161
- Ellis RS, Newman CM, Rosen JS (1980) Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Probab Theory Relat Fields* 51(2):153–169
- Ellis RS, Rosen JS (1982) Laplace's method for Gaussian integrals with an application to statistical mechanics. *Ann Probab* 47:66
- Feller W (1960) *An introduction to probability theory and its applications*. John Wiley and Sons, Inc, New York, NY
- Gallo I, Barra A, Contucci P, (2009) Parameter evaluation of a simple mean-field model of social interaction. *Mathematical Models and Methods in Applied Science* 19:1427–1439
- Glaeser E, Scheinkman J (2001) Measuring social interactions. In Durlauf SN, Young HP (eds) *Social dynamics*. Brookings Institution Press, Washington DC, pp 83–132
- Granovetter M, (1983) The strength of weak ties: A network theory revisited. *Sociol Theor* 1:201–233
- Hauert C, Doebeli M (2004) Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428(6983):643
- Heilmann OJ, Lieb EH (1972) Theory of monomer-dimer systems. *Commun Math Phys* 25(3):190–232
- Horst U, Scheinkman J (2006) Equilibria in systems of social interactions. *J Econ Theory* 130(1):44–77
- ISTAT (2018) <http://www.istat.it/it/archivio/82599>.
- Liggett TM (1985) *Interacting particle systems*. Springer Verlag, New York
- McFadden D (2001) Economic choices. *Am Econ Rev*, 91(3):351–378
- Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314(5805):1560–1563
- Parisi G (1988) *Statistical field theory*. Addison-Wesley, USA
- Pastore F, Ponzio I (2016) *Changing neighbourhoods: Inter-group relations and migrant integration in European cities*. Springer Press Imiscoe Research Series, Dordrecht
- Portes A, Sensenbrenner J (1993) Embeddedness and immigration: Notes on the social determinants of economic action. *Am J Sociol* 98(6):1320–1350
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci* 94(17):9197–9201
- Scheinkman J (2018) Lectures on social interactions. <http://www.princeton.edu/joses/lsi.html>
- Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. *Nature* 393(6684):440
- Weber M (1978) *Economy and society: An outline of interpretive sociology*, vol 1. California University Press, Berkely

Data availability

The datasets analysed during the current study are available on demand in the ISTAT repository: <https://contact.istat.it>.

Acknowledgements

EA, AB, AP are grateful to GNFM-INdAM Progetto-Giovani Agliari-2016; AB further acknowledges support by Salento University; AP further acknowledges support by the Engineering and Physical Sciences Research Council (EPSRC), Grant No. EP/L505110/1, by The Alan Turing Institute EPSRC grant EP/N510129/1 and seed project SF029 “Predictive graph analytics and propagation of information in networks”; CV acknowledges financial supports from Fondo di Ateneo per la ricerca 2015, Università di Modena e Reggio Emilia and FIRB Grant RBF10N90W. PC acknowledges financial support from PRIN project Statistical Mechanics and Complexity (2015K7KK8L). Finally and most importantly we want to dedicate this work to our missed colleague and friend Ignacio Gallo. Ignacio has been a driving force in the mathematical approach to quantitative sociology and his ideas had a profound impact on our understanding of the field.

Additional information

Supplementary information: accompanies this paper at <https://doi.org/10.1057/s41599-018-0097-5>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018