A quantitative approach to the modelling of interacting systems from empirical data: the statistical mechanics perspective and a case study from social sciences

Dr. Francesco De Pretis
M2CSC School of Graduate Studies in
Multiscale Modelling, Computational Simulations
and Characterization in Material and Life Sciences

University of Modena and Reggio Emilia (Italy)

A new approach to social phenomena modelling (in particular, immigrant integration) is proposed since the necessity of defining more effective models, able to explain such phenomena.

Very little is known about the mechanisms that bring about integration.

Elementary questions like how integration responds to an increase in immigration density still miss empirical and theoretical answers.

PLAN OF THE PRESENTATION

- Statistical Mechanics models:
  - mono-populated model (non-interacting systems)
  - bi-populated model (interacting systems)
- Description of empirical data
- Data-Analyses results
- Discussion and future perspectives

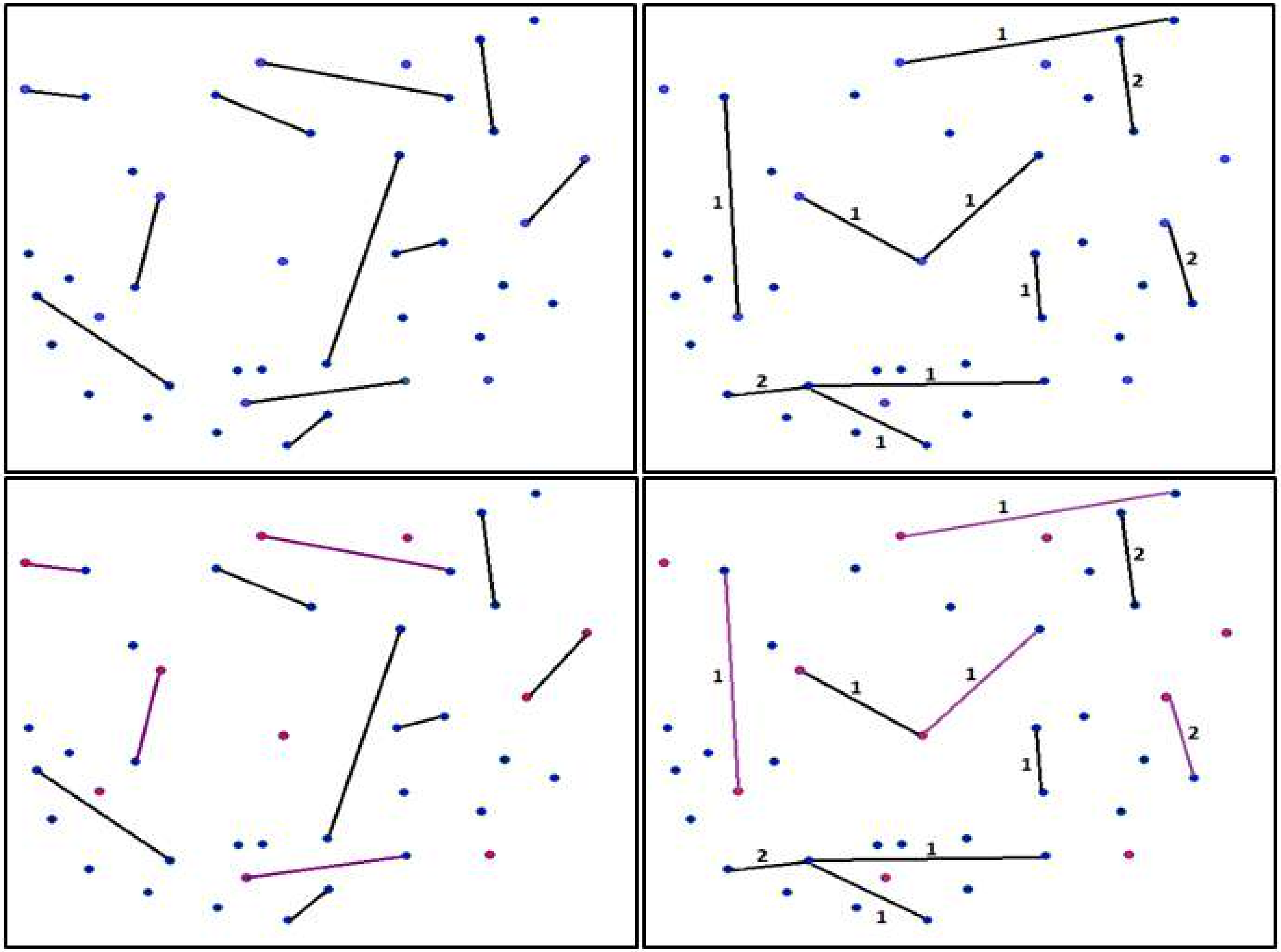
STATISTICAL MECHANICS APPROACH

We focus on classical integration quantifiers Q such as the rate of mixed marriages and mixed newborns in a system including immigrants and natives.

**Our aim is to study such quantifiers Q
as a function of immigrant density γ,
following a statistical mechanics modelling.**

In particular, we want to distinguish whether these phenomena are due to imitative interaction among individuals rather than individual choices.

MONO-POPULATED AND BI-POPULATED MODELS



MONO-POPULATED MODEL: MARRIAGES

Given a population, that is a set of N individuals named $I = \{1, \dots, N\}$ we define a marriage configuration M as the union of two sets S_M and C_M, that is $M = S_M \cup C_M$.

The subset $C_M \subset I^2$ represents all the paired individuals (i.e. possible married couples) of the M configuration.

Each element $(i, j) \in C_M$ is denoted by having the following properties:

- $i \neq j$ (no self-loops)
- $(i, j) = (j, i)$ (symmetric relation)
- $if\ (i, j) \in C_M \rightarrow (i, k) \notin C_M\ \forall k \neq j$ (monogamy constraint)
- (i, j) is connected *only* by one link, symbolizing a possible marriage union.

The subset $S_M \subset I$ represents instead the unpaired individuals (i.e. singles) of the M configuration

MONO-POPULATED MODEL: MARRIAGES

Each element of the set S_M shows a certain inclination to marry versus remaining singles.

Each element of the set C_M has its own likelihood to marry.

The parameter s_i models this tendency for the i-element of the set S_M and so $c_{i,j}$ that one for the couple (i, j) of the set C_M.

Both s_i and $c_{i,j}$ can be thought as weights (positive real numbers) for the i-single and for the (i, j) couple.

MONO-POPULATED MODEL: NEWBORNS

We define a filiation configuration F as the union of two sets U_F and P_F.

The subset U_F represents the unpaired individuals (i.e. undescendents), whereas the subset $P_F \subset I^2$ represents all the paired individuals (i.e. those couples who are characterized by possibly having children) of the F configuration.
For each element $(i,j) \in P_F$, we define a function $l: P_F \rightarrow \mathbb{N}^+$, where $l(i,j)$ counts the number of links, i.e. children belonging to the (i,j) couple.

Each element $(i,j) \in P_F$ is denoted by having similar but not equal properties to those ones previously described for the elements of the C_M set namely:

- $i \neq j$ (no self-loops)
- $(i,j) = (j,i)$ (symmetric relation)
- (i,j) is connected *at least* by one link ($l(i,j) \geq 1$), symbolizing in this case a newborn originated by the couple.

MONO-POPULATED MODEL: NEWBORNS

The number $l(i,j)$ for each couple (i,j) may be modeled as a random variable, likely a Poisson distribution.
The choice of such distribution is reasonable but the following results do not strictly depend on it.

Each element of the set U_F shows an individual tendency to have children.

Each element of the set P_F too.

We call therefore u_i this tendency for the i-element of the set U_F and $p_{i,j}$ that one for the couple (i,j) of the set P_F. Again, both u_i and $p_{i,j}$ can be thought as weights (positive real numbers) for the i-undescendent and for the (i,j) parents.

MONO-POPULATED MODEL: PARTITION FUNCTIONS

$\mathcal{M}$ is the set of *all* marriage configurations
$\mathcal{F}$ is the set of all filiation configurations.

The model is fully defined together with E – the acquaintance matrix of the N individuals of the population – whose elements $\varepsilon_{i,j} \in \{0,1\}$ set the connections among the individuals.

Given such information, we are able to write the partition functions of these systems.

MONO-POPULATED MODEL: PARTITION FUNCTIONS

$$Z^{(\mathcal{M})} = \sum_{M \in \mathcal{M}} \prod_{(i,j) \in C_M} \varepsilon_{i,j} c_{i,j} \prod_{i \in S_M} s_i$$

$$Z^{(\mathcal{F})} = \sum_{F \in \mathcal{F}} \rho(F) \prod_{(i,j) \in P_F} \varepsilon_{i,j} p_{i,j} \prod_{i \in U_F} u_i$$

MONO-POPULATED MODEL: EXPECTED VALUES

Calling K_M the total number of links in the configuration M and defining the frequency as $\nu_M = 2K_M/N$, the expected value of the marriage frequency can be computed as:

$$P_{\mathcal{M}} = \mathbf{Av}\frac{\sum_{M\in\mathcal{M}} \nu_M \prod_{(i,j)\in C_M} \varepsilon_{i,j} c_{i,j} \prod_{i\in S_M} s_i}{Z^{(\mathcal{M})}}$$

where the average operation **Av** is computed on the acquaintance matrix ensemble.

MONO-POPULATED MODEL: EXPECTED VALUES

Analogously, calling K_F the total number of links in the configuration F and defining the frequency as $\nu_F = 2K_F/N$, the expected value of the newborn frequency is

$$P_{\mathcal{F}} = \mathbf{Av}\frac{\sum_{F\in\mathcal{F}} \nu_F \rho(F) \prod_{(i,j)\in P_F} \varepsilon_{i,j} p_{i,j} \prod_{i\in U_F} u_i}{Z^{(\mathcal{F})}}$$

where the average operation **Av** is computed on the acquaintance matrix ensemble.

BI-POPULATED MODEL: INTERACTION

We consider now a similar model where the initial population of N individuals is partitioned into two subsets, representing N_{imm} immigrants and N_{nat} natives.
We model the imitative interaction (for marriages) between the two populations through a suitable mean-field Hamiltonian:

$$H(M) = -J_M \sum_{i \in Nat, j \in Imm} \varepsilon_{i,j} \sigma_i \sigma_j$$

where

$$\sigma_i = \begin{cases} +1 & \text{if } i \text{ belongs to a mixed marriage} \\ -1 & \text{otherwise} \end{cases}$$

and similarly another Hamiltonian is introduced for the newborns.

BI-POPULATED MODEL: EXPECTED VALUES

Following the same path shown for the mono-populated model, we can write the expected value for the mixed marriages frequency:

$$P_{\mathcal{M}}^{(Nat,Imm)} = \mathbf{Av}\frac{\sum_{M\in\mathcal{M}} f_M e^{-H(M)} \prod_{(i,j)\in C_M} \varepsilon_{i,j} c_{i,j} \prod_{i\in S_M} s_i}{Z_H^{(\mathcal{M})}}$$

where the average operation **Av** is computed on the acquaintance matrix ensemble, and $f_M = M_M/K_M$ with M_M the number of mixed marriages and K_M the total number of marriages in the configuration M.

BI-POPULATED MODEL: EXPECTED VALUES

Analogously, we can write the expected value for the mixed newborns frequency:

$$P_{\mathcal{F}}^{(Nat,Imm)} = \mathbf{Av}\frac{\sum_{F\in\mathcal{F}} f_F e^{-H(F)}\rho(F)\prod_{(i,j)\in P_F}\varepsilon_{i,j}p_{i,j}\prod_{i\in U_F}u_i}{Z_H^{(\mathcal{F})}}$$

where the average operation **Av** is computed on the acquaintance matrix ensemble, and $f_F = M_F/K_F$ with M_F the number of children from mixed couples and K_F the total number of children in the configuration F.

BI-POPULATED MODEL: MEAN-FIELD LIMITS

Even if a general solution of this model is not yet available, we can focus on two extreme regimes for the previous expected values P's.

When the imitative interaction is dominant, we have:

$$P(\Gamma) \propto \mathrm{a}\sqrt{\Gamma}$$

whereas when individual choices are dominant, we have:

$$P(\Gamma) \propto \mathrm{a}\Gamma$$

where $\Gamma = \gamma(1-\gamma)$ tunes the cross-links couplings among immigrants and natives and $\gamma = {}^{N_{imm}}/_{N}$ is the immigrant density.

EMPIRICAL DATA

We work on a large ISTAT database containing over 10^6 information concerning marriages and newborns occurred between two populations (immigrants and natives).

A deep look into Italy, considering what was registered in all its 8100 municipalities for 11 years (from 2001 to 2011)

IMMIGRANT DENSITY AND INTEGRATION QUANTIFIERS

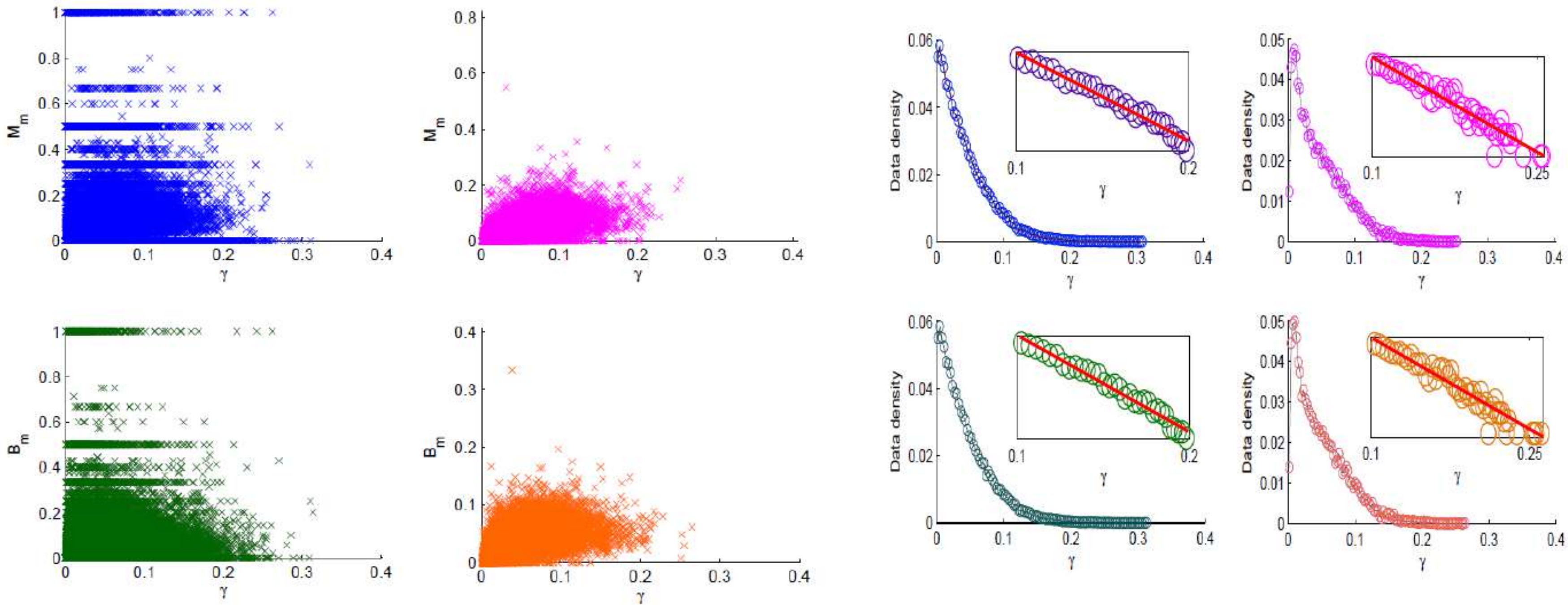
For each municipality at each given year, we compute the following social indicators:

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

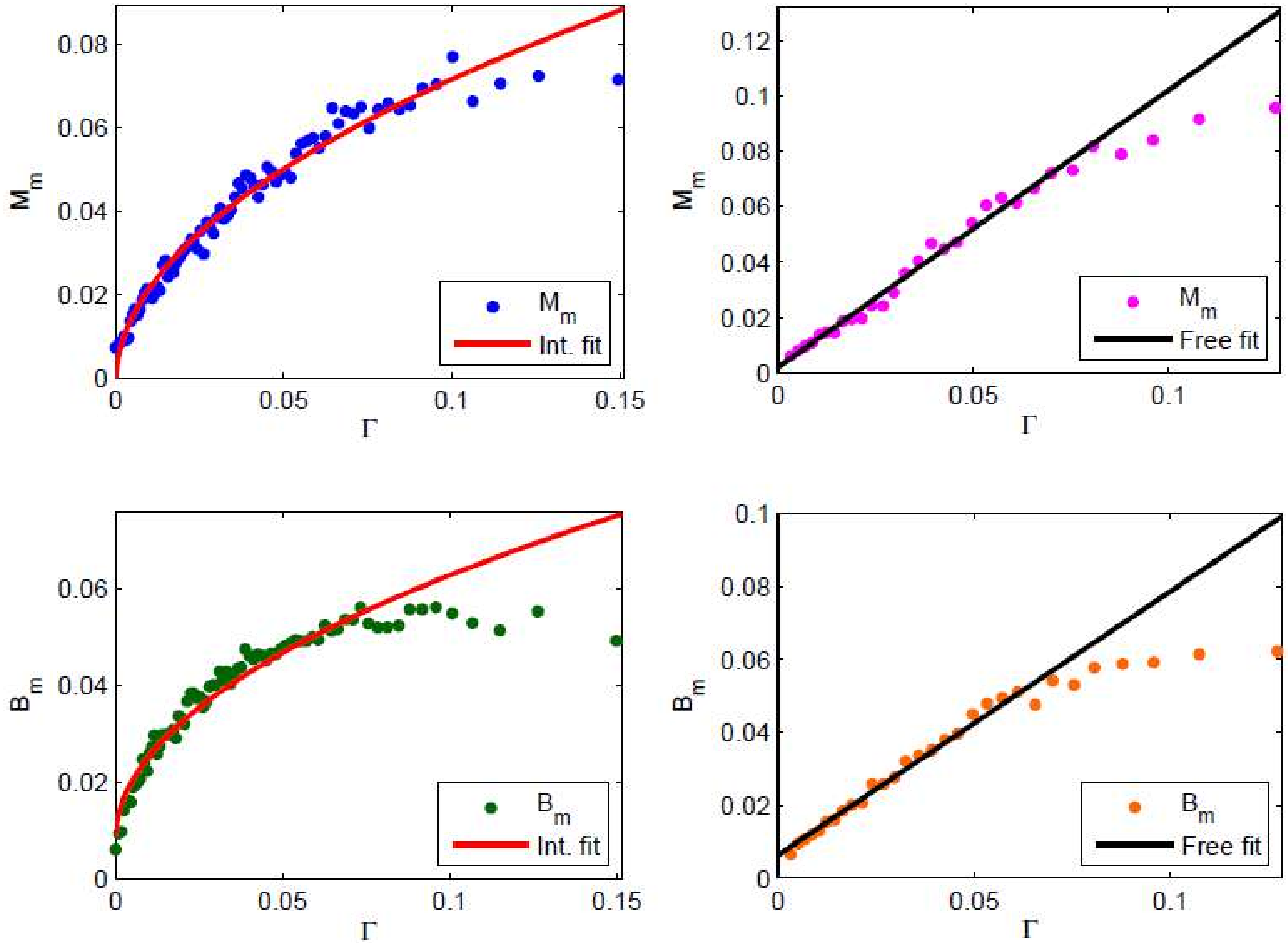
$$M_m = \frac{M_{mix}}{M_{mix} + M_{nat} + M_{imm}} \in [0,1]$$

$$B_m = \frac{B_{mix}}{B_{mix} + B_{nat} + B_{imm}} \in [0,1]$$

RAW DATA CLOUDS AND DATA DENSITY



AVERAGE QUANTIFIERS FITS



DISCUSSION

As seen before, different patterns for average quantifiers of integration arise in small and large municipalities.

Differences in social actions were already perceived by sociologist Emile Durkheim (*anomie*)

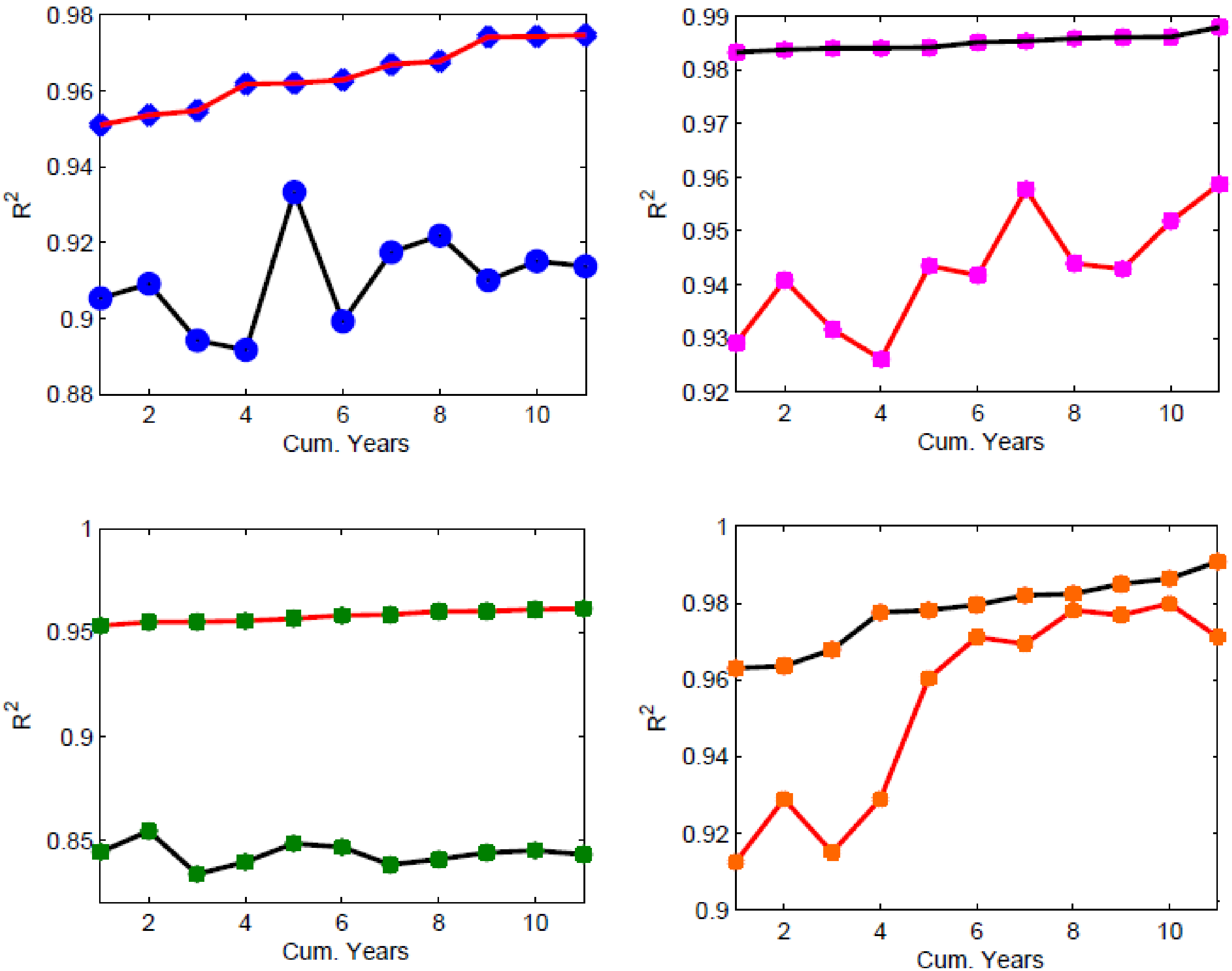
ROBUSTNESS OF THIS APPROACH

Do these empirical laws hold also on a time-scale?

Do imitative interaction phenomena still occur also according to different time-windows?

- Check of predictability: time-dependent analysis

TIME-DEPENDENT ANALYSIS



ROBUSTNESS OF THIS APPROACH

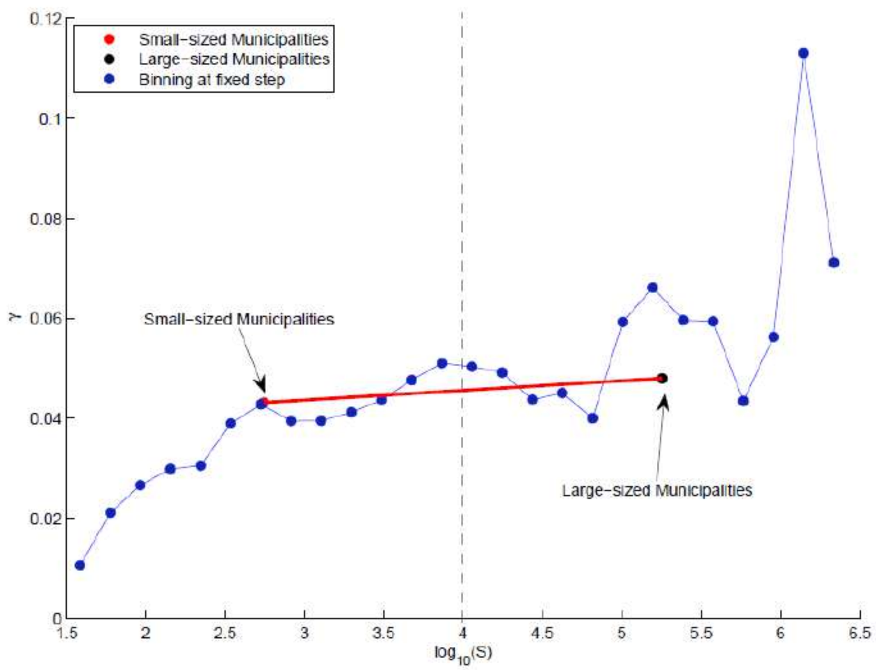
Is there an imbalance in the distribution of immigrants between small and large municipalities?

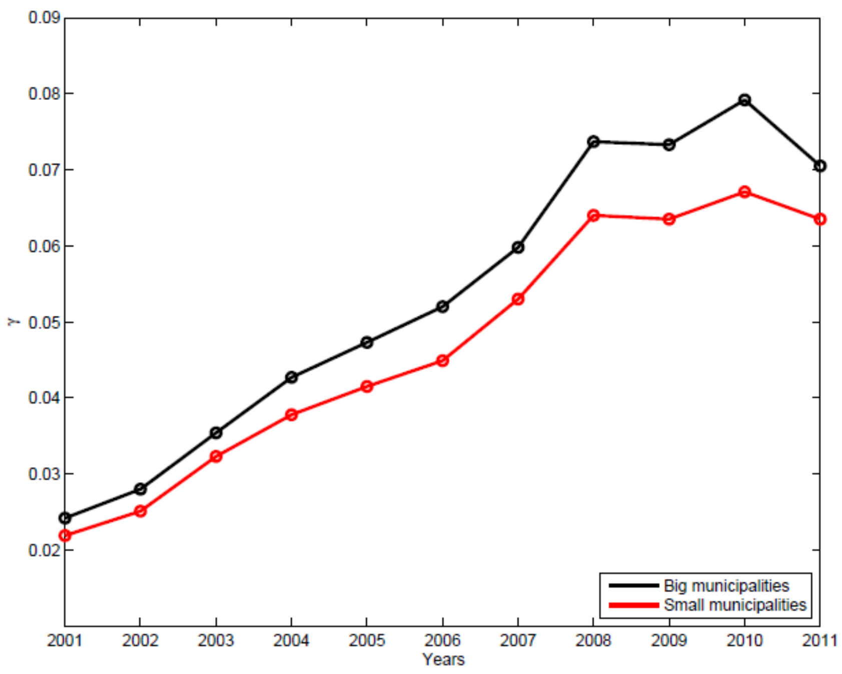
- Check of immigrant density in small and large municipalities.

Is there an imbalance in the proportion of immigrants and natives in small municipalities?

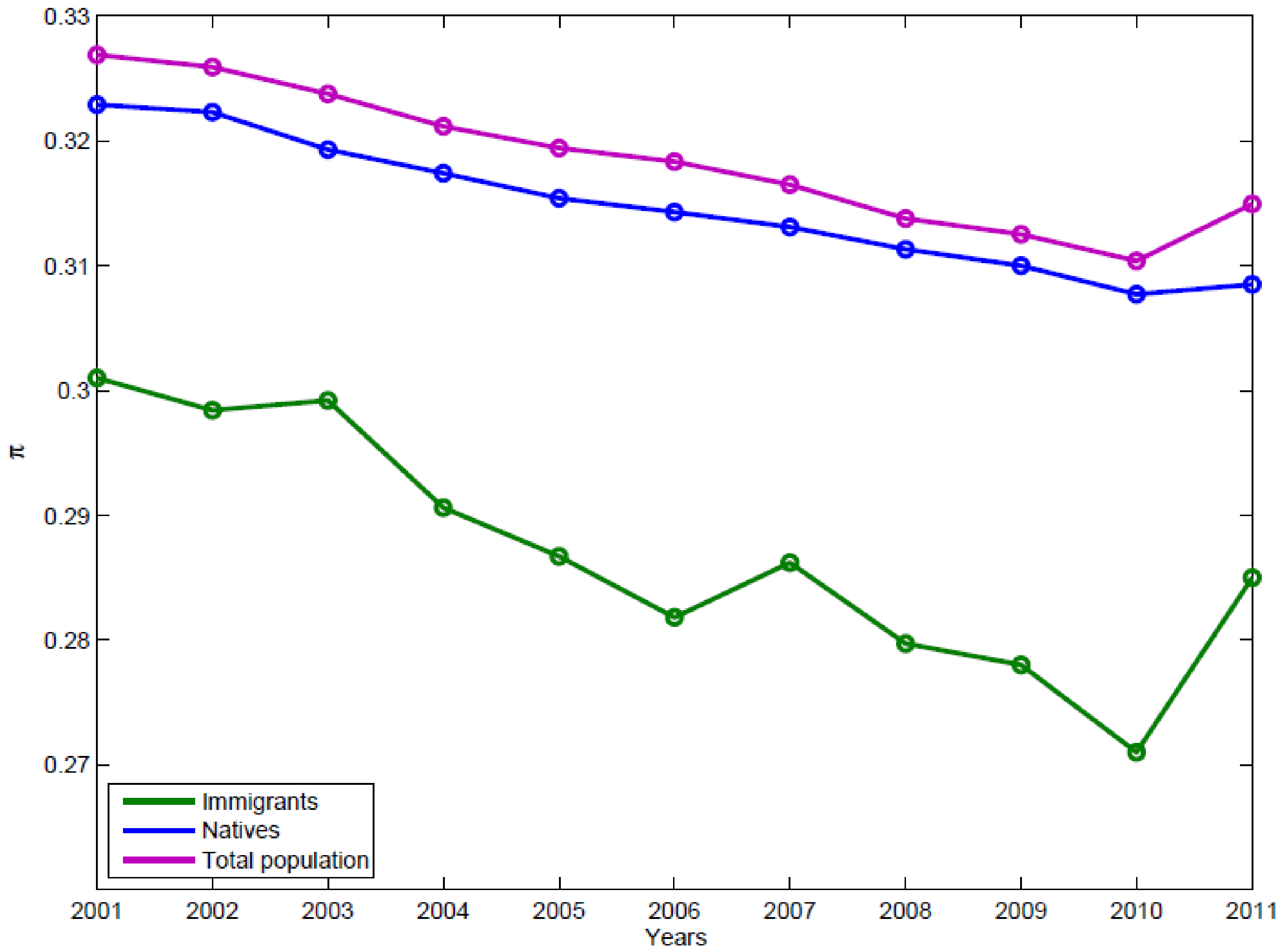
- Check of immigrant and native proportion in small municipalities.

IMMIGRANT DENSITY IN SMALL AND LARGE CITIES





PROPORTION OF IMMIGRANTS AND NATIVES IN SMALL CITIES



FUTURE PERSPECTIVES

- Comparison between the national scale and other macro-geographical areas

- the Emilia Romagna case-study

REFERENCES AND ACKNOWLEDGEMENTS

REFERENCES

An analysis of a large dataset on immigrant integration in Spain. The Statistical Mechanics perspective on Social Action.
A. Barra, P. Contucci, R. Sandell, C. Vernia (2013). *Scientific Reports, Nature, 4:4174.*

A statistical mechanics approach to immigrant integration in Emilia Romagna (Italy).
F. De Pretis, C. Vernia (2013). *Preprint. Accepted to be published in Complex Networks, Studies in Computational Intelligence 424, Springer-Verlag.*

The Statistical Physics approach on immigrant integration in Italy.
P. Contucci, C. Vernia, F. De Pretis (2014). *Preprint.*

ACKNOWLEDGEMENTS

I would like to express my gratitude to my PhD tutor, Prof. Claudio Giberti and to Prof. Pierluigi Contucci and Prof. Cecilia Vernia for the valuables insights given to my research activity.

Thanks for the attention!