Bartoli’s areal norms revisited: an agent-based modeling approach

Dalmo Buzato
Universidade Federal de Minas Gerais (UFMG)

FALE
FACULDADE
DE LETRAS

UFMG

Long paper with full and detailed explanations:

Buzato, Dalmo, and Evandro Cunha. "Bartoli's areal norms revisited: an agent-based modeling approach." Proceedings of the 16th International Conference on Computational Processing of Portuguese. 2024.

Evandro L.T. P. Cunha
Assistant professor at the Faculty of Letters of UFMG
PhD degree in Linguistics from Universiteit Leiden and in Computer Science from UFMG

ADA
Núcleo de Linguística Computacional
FALE/UFMG
ALGORITMOS · DADOS · ANÁLISE LINGUÍSTICA

ADA: Working Group on Computational Linguistics at UFMG's Faculty of Letters

Neolinguistica and Spatial Linguistics

- Linguistics in 19th century
  - What is the place of *change* in language nature?
  - *Junggrammatiker* (lit. young grammarians, neogrammarians)
  - Regularity of sound change (Grimm's law; Verner's law…)

- Against the infallible laws
  - Matteo Bartoli (1873 - 1946)
  - *Neolinguistica* and *Linguistica Spaziale*
  - Probabilistic norms for language change, based on geographical space

Grimm's law VS. Bartoli’s norms

Grimm’s Law: Proto-Indo-European voiceless plosives change into Proto-Germanic voiceless fricatives (p → f)

Sanskrit	Greek	Latin	Gothic	English	German
pad-	**p**od-	**p**ed-	**f**ōtus	**f**oot	**f**uss
pitār-	**p**atḗr	**p**ater	**f**adar [ˈfaðar]	**f**ather (< fæder)	**v**ater [**f**ater]

- Infallible and universal laws for language change

Grimm's law VS. Bartoli’s norms

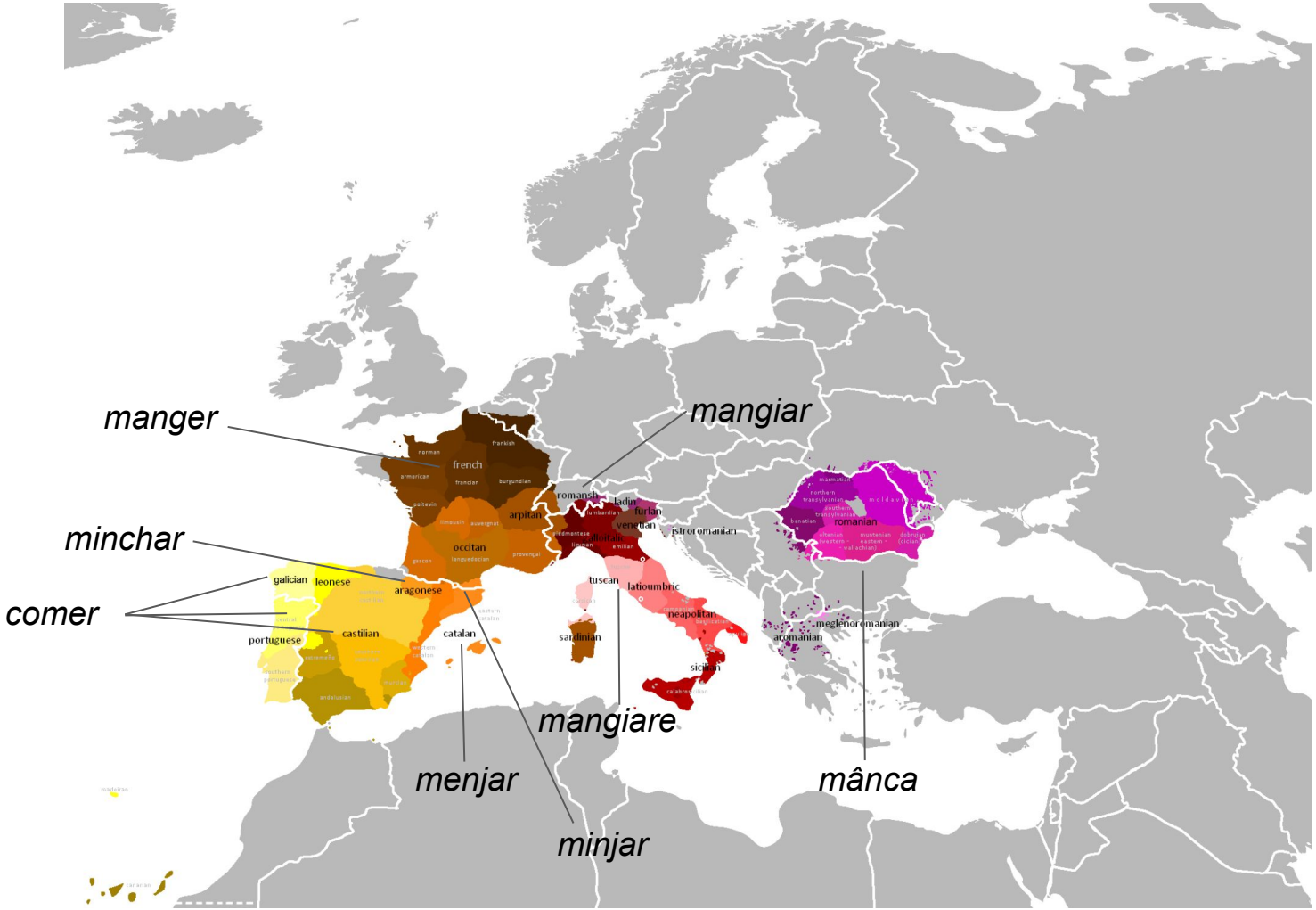
Bartoli’s norms: I The more isolated area usually preserves the earlier stage.

II If one of two linguistic stages is found in peripheral areas and the other in a central area, the stage occurring in the peripheral areas is usually the earlier one.

III The larger area usually preserves the earlier stage.

IV The earlier stage is usually preserved in the later area.

V If one of two linguistic stages disappears or becomes moribund and the other survives, the stage that disappears is usually the earlier one.


manger
mangiar
minchar
comer
mangiare
menjar
minjar
mânca
french
occitan
arpitan
romansh
ladin
furlan
venetian
istroromanian
romanian
galician
leonese
aragonese
portuguese
castilian
catalan
tuscan
latioumbric
neapolitan
sicilian
sardinian
megleno-romanian
aromanian

Bartoli’s second norm

If one of two linguistic stages is found in peripheral areas and the other in a central area, the stage occurring in the **peripheral areas** is usually the **earlier one**.

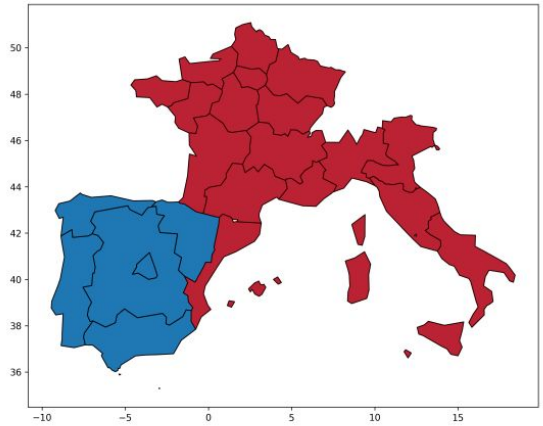


Figure 1: Part of present-day Romance language-speaking Europe that use forms for the verb ‘to eat’ derived from Latin *manducare* (red) and *comedere* (blue)

Thus, according to Bartoli's theory, the verb *comedere* should be **older** than the verb *manducare*, which in fact applies.

Center and periphery in Romance languages

In the case of Romance languages, in their genesis and initial development, central Italy can be considered an important center of innovations due to Rome being the capital of the Roman Empire, and a major population and prestige center.

oldest form: *comedere* (comer)

innovation: *manducare* (minchar, menjar, minjar, mangiare, mânca)

The areas close to the center of innovation adopted the new form, while the peripheral areas kept the more archaic one in their lexicon, since the innovation did not reach there.

Limits

- *comedere* vs. *manducare* is just one example of countless language changes

- Since the publication of Bartoli’s first work with his areal theory, several case studies questioning and validating his hypotheses have been presented;

- Would lexical (words) changes occur in the same way as phonetic (sounds) or grammatical changes?

- How can we prove (or not) Bartoli's theory, taking geographic aspects into account?

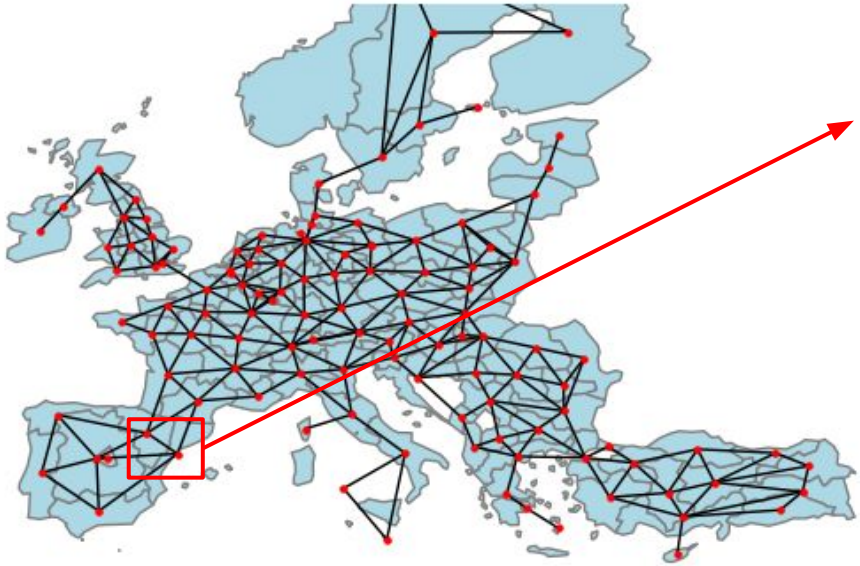
How to build a transmission model of linguistic items?

- Ideally, it would be better to have a computationally tractable file with geographic data of Medieval Europe, but we did not find such an option.
- `Geojson` file of the NUTS 16 classification (First-level Classification of Territorial Units for Statistics)

How to build a transmission model of linguistic items?

- It involved extracting centroids (the average of the coordinates that define the area boundary) to connect the regions and constructing the graph based on the Queen model (where the graph considers two areas as connected if they share a single point on their boundary).

Algorithm



- We take all pairs of nodes that are connected by an edge, and we will refer to these nodes as A and B.
- Two interaction possibilities: A → B and B → A.
- We randomly generate a probabilistic item (PI) ranging from 0 to 1.
- If PI > 0.9 and if the originating node has the innovative item while the destination node does not, then transmission occurs.

Algorithm

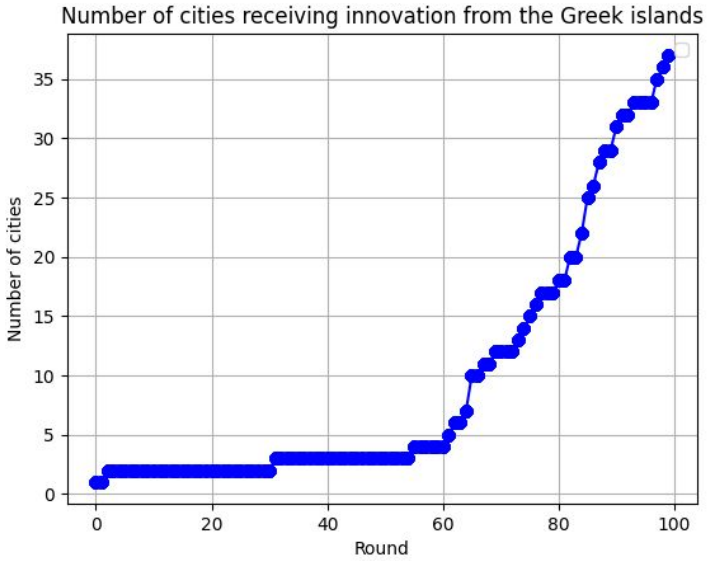
- Initially, **all nodes** are set with the innovation variable equal to 0, meaning that none of them possesses the innovation.
- At the start of each simulation run, **one node** is set with the innovation variable equal to 1, and from that point we can observe how this innovation propagates through the network.

Network details

- **120** nodes
- **254** edges
- Each simulation consists of **100** rounds
- We ran the model **1,000** times for robust quantitative analyses
- To analyze the centrality of each node in the network, we opted to use the ***betweenness centrality*** method in NetworkX library.

Results

- Does innovations originating from different positions in the network (consequently, with different centralities) exhibit different patterns of propagation in equal times?



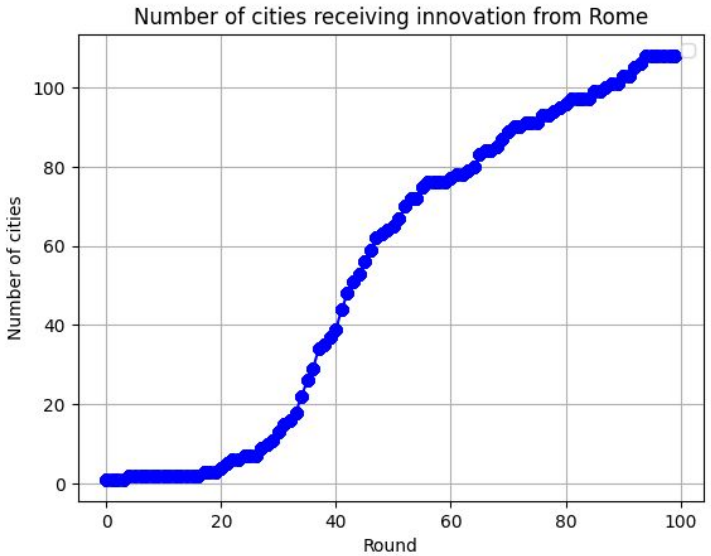
A node near the Greek islands, one of the most isolated locations on the graph

The number of nodes that received the innovation was low compared to the total, with only 37 receiving nodes

Long periods of stabilization in a few isolated locations

Results

- Does innovations originating from different positions in the network (consequently, with different centralities) exhibit different patterns of propagation in equal times?



A node near Rome, a node more central compared to the Greek islands

The trajectory depicted seems highly prototypical of the S-curve definition present in sociolinguistic tradition

Almost all nodes received the innovation from the central location

Results

- Innovations originating from a more central node (in this analysis, Rome) have significantly greater and more consistent reach across the entire network when compared to innovations originating from a more isolated node (in this analysis, the Greek islands).

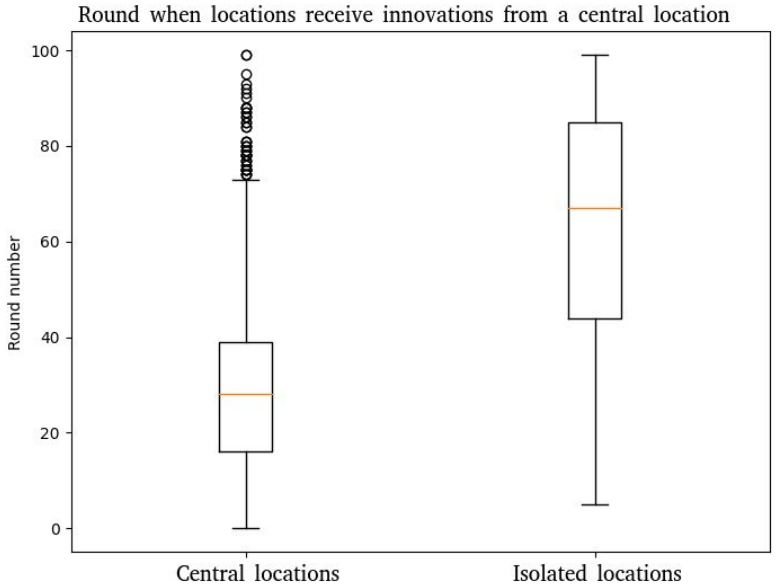
Region of origin	μ	sd
Greek islands	41.27	30.971
Rome	107.48	9.252

Table 1: Mean number of receiving nodes (and standard deviation) according to the innovation's node (representing regions and cities) of origin

When innovation originates from the isolated location, in none of the 1,000 runs did the innovation reach all nodes of the graph!

Results

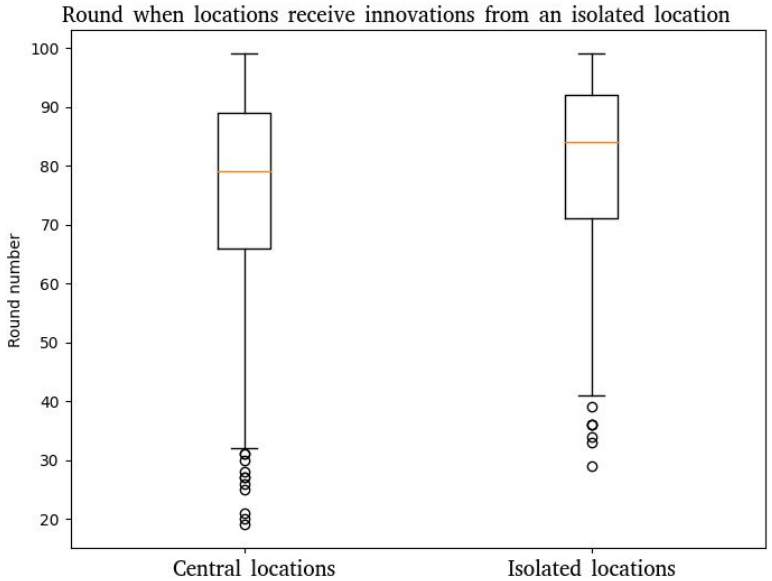
- In which round, on average, does innovation reach central and isolated locations when it originates in other central and isolated locations?



When an innovation originates from a central location, other central locations (μ = 28.76, sd = 16.706), disregarding graph distances, typically **receive the innovation earlier** compared to isolated regions (μ = 63.64, sd = 24.120).

Results

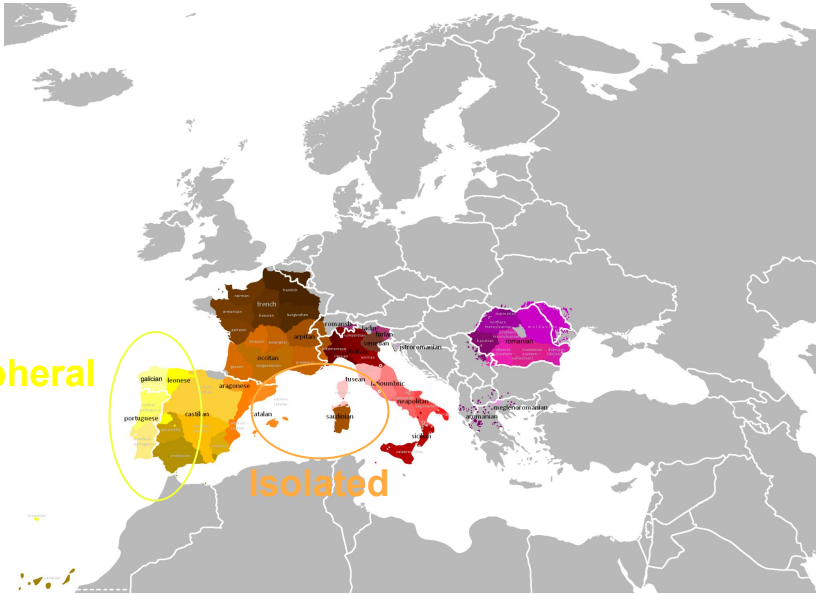
- In which round, on average, does innovation reach central and isolated locations when it originates in other central and isolated locations?



When innovations from isolated nodes **take more rounds to reach central nodes** (μ = 76.39, sd = 15.918) **and also to reach other isolated nodes** (μ = 80.19, sd = 14.562), with no significant difference between them.

Results

- In which round, on average, does innovation reach central and peripheral locations when it originates in other central and peripheral locations?



Innovations take longer to reach peripheral areas like Portugal/Galicia: μ = 56.17, sd = 17.173 when coming from **central regions**; μ = 91.09, sd = 6.561 when coming from **isolated regions**

Balearic definite articles (*es*, *sa*, *ses*, *s'*) derived from the archaic Latin form *ipse/ipsa*, while all other Romance languages (except Sardinian) use articles derived from the later Latin *ille/illa*

Bartoli’s norms and our results

- When innovations arise from central regions, they take much longer to establish themselves in isolated/peripheral areas meaning that these regions become linguistically isolated.
- If we observe this state over time, we will find that more archaic forms are concentrated in isolated regions, while innovations are used in more central areas, as stated in Bartoli’s first norm (“*the more isolated area usually preserves the earlier stage*”) and second norm (“*if one of two linguistic stages is found in peripheral areas and the other in a central area, the stage occurring in the peripheral areas is usually the earlier one*”).

Mozarabic

- Mozarabic was a set of Ibero-Romance varieties that developed in Al-Andalus, the part of the medieval Iberian Peninsula under Islamic control.
- This set of varieties likely became extinct by the end of the 14th century, being replaced by Andalusi Arabic (Muslim-controlled south) and to the Romance varieties (Christian kingdoms in the north).

Mozarabic

- In terms of phonology and morphology, Mozarabic is closer to Latin than other Romance varieties.
- *lo**p**a* (Sp. lo**b**a, ‘she-wolf’), *to**t**o* (Sp. to**d**o, ‘everything’), and *formi**c**a* (Sp. hormi**g**a, ‘ant’).
- In other peninsular Romance languages, changes occurred such as /p/ becoming /b/, /t/ becoming /d/, and /k/ becoming /g/, but not in Mozarabic.

Mozarabic

- Bartoli's fifth norm, which postulates that the variety that becomes extinct in favor of another is usually the oldest.
- But if we look at the past…
  - **Mozarabic**
  - **Emerging romance varieties**
  - **Andalusi Arabic**

Why was Mozarabic the only archaic variant even in the past?

Mozarabic

Mozarabic was **peripheral** due to its use being increasingly concentrated in the south of the peninsula due to the Reconquista.

Isolated due to linguistic, cultural, and political barriers between them and the Christian kingdoms to the north.

As seen in the simulations, the innovations that arrived in the Iberian Peninsula from the north did not reach the Mozarabic-speaking locations due to their extremely **peripheral** and **isolated** characteristics.

Limitations

- Simple rules for interaction between agents;
- The initially created network does not allow verification of Bartoli's third and fourth norms (which involve area size and occupation order);
- The simulation data was applied only to Romance languages.

Bartoli overseas

Map 3. The Northeast Amazon, Northwest Amazon, Southwest Amazon, and Upper Xingu areas of Lowland South America.

20 languages, **8** language families
Including **3** isolated languages, that is, without known relatives.

Our knowledge about the past of these civilizations and languages is extremely limited.

Bartoli's norms, together with analyzes of language contact (ABM included), could help us understand the past of these languages.

References

Jörn Albrecht. 1996. Neolinguistic school in Italy. In E. F. K. Koerner and R. E. Asher, editors, Concise History of the Language Sciences: From the Sumerians to the Cognitivists, pages 243–246. Pergamon.

Matteo Bartoli. 1925. Introduzione alla Neolinguistica: Principi e Metodi. Olschki, Geneva.

Matteo Bartoli. 1945. Saggi di Linguistica Spaziale. Vincenzo Bona, Turin.

Witold Manczak. 1988. Bartoli’s second "norm". In Jacek Fisiak, editor, Historical Dialectology: Regional and Social, Trends in Linguistics. Studies and Monographs, pages 349–355. Mouton de Gruyter, Berlin; New York; Amsterdam.

Roger Wright. 1982. Late Latin and Early Romance in Spain and Carolingian France. University of Liverpool (Francis Cairns, Robin Seager), Liverpool.

FALE
FACULDADE
DE LETRAS

UFMG

Muito obrigado! Thank you!

dalmobuzato@ufmg.br

IFISC (Institute for Cross-Disciplinary Physics
and Complex Systems)

Palma de Mallorca

June 6th, 2024