Analysis of the temporal sequence of co-authorship networks between European countries in the years 1990-2023 from OpenAlex

Anuška Ferligoj, Vladimir Batagelj, Marjan Cugmas, Franc Mali
UL FDV, IMFM, UP IAM

The workshop on "Innovation in dynamic network modelling", Lugano, September 11-13, 2024

Outline

Anuška Ferligoj: anuska.ferligoj@fdv.uni-lj.si

Introduction

- The goal of the study is to identify clusters of European countries with similar scientific collaboration "profiles" inside the group and in other clusters of countries.
- Scientific collaboration intensified significantly with the growth of technology and globalization (Glanzel 2001; Gui et al. 2019).
- Countries are getting involved in a global process of knowledge production. Growth rates and the places of the countries in a global academic network are different (Luukkonen et al. 1993; Wagner, Leydesdorff 2005; Gazni et al. 2012).
- The place of a country in the international scientific collaboration network often connects with policy inside the country (Wagner, Leydesdorff 2005) and with the level of the country's scientific development (Gazni et al. 2012).
- Countries with a high publication output and developed research systems occupy central positions in a global scientific network, and many other countries strive to intensify their scientific collaboration with these nations. Small countries are often in the shadow of the countries with a big publication output.

Research questions

Is scientific collaboration between European countries defined by the 'tyranny of geographical distance' (Castells, 1996)? There are two types of factors affecting geographical proximity on the intensity of scientific cooperation between countries:

- **self-regulatory factors** connected with historical reasons, similarity of languages, etc. (Scharnhorst et al., 2012)
- **external factors** connected by transnationally established research and development strategies and policies (e.g. European Research Area - ERA)

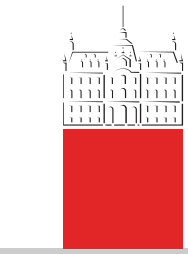
Self-regulatory factors

- Is the similarity of languages an incentive and the diversity of languages is a barrier to scientific cooperation, even though there is English as an 'lingua franca' (Liang et al., 2006)?
- What is the influence of social-institutional differences between European countries on their scientific cooperation (e.g., general historical factors), e.g., between
  - Mediterranean Europe versus Northern Europe,
  - Mittel Europe versus Western Europe (Siune et al., 2009),
  - The core part of Europe versus the peripheral part (Lažnjaka, Švarc, 2011).

External factors

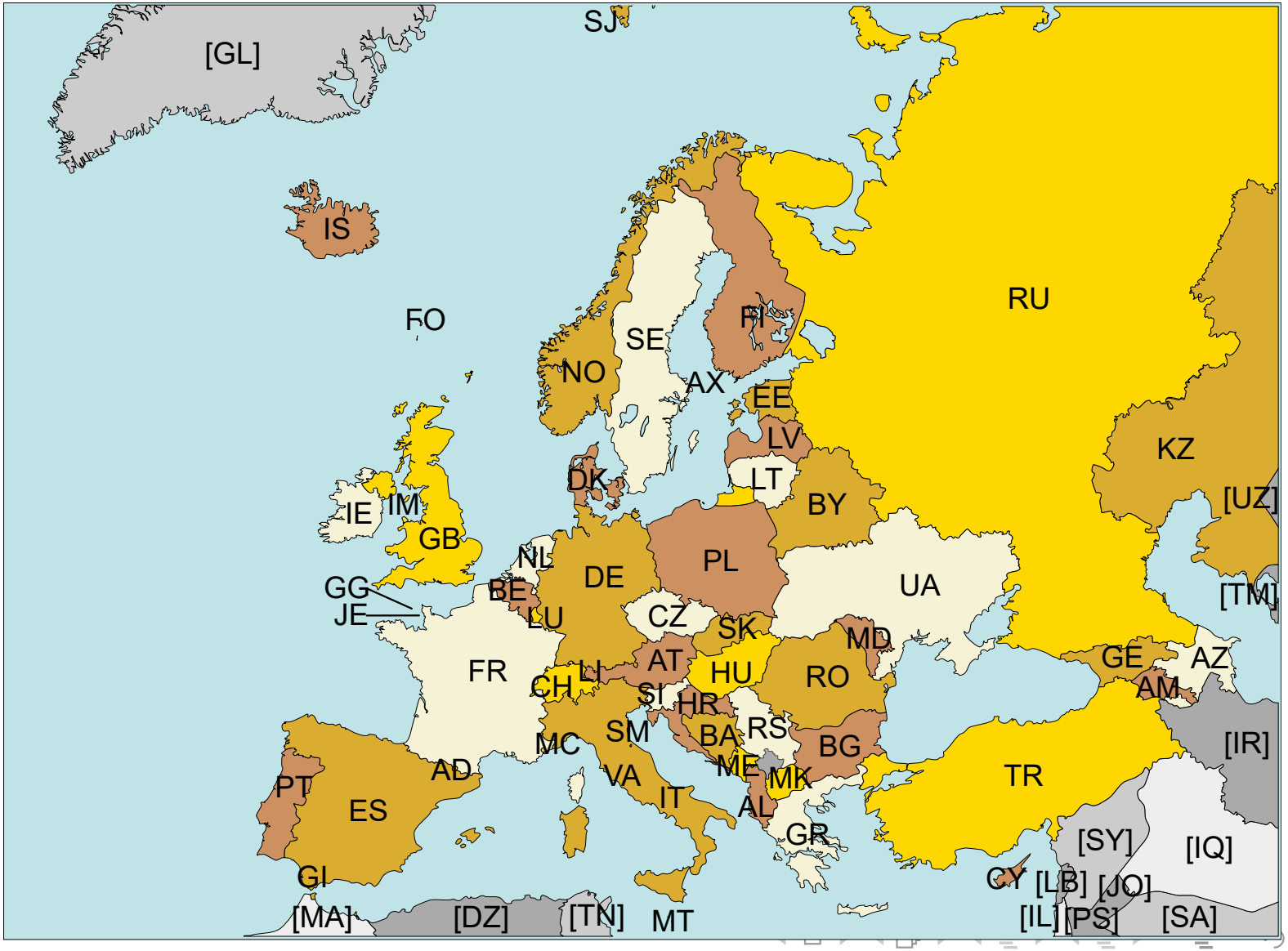
What is the effect of the EU strategy of joint European Research Area?

- Between 1960 and 1980 there were established multi-literary forms of cooperation between the leading scientific and technological forces in leading countries in Europe, by establishing large joint infrastructure research projects (reactors, accelerators, telescopes, laboratories) and financial programs (CERN, 1954; ESO, 1962; ESA, 1964; ILL, 1967; COST, 1974; EUREKA, 1985, ESRF, 1989).
- In the 1980s, the European Commission began to strive for the integration of research potentials in Europe by establishing funding mechanisms known as EU framework programs (EU OPs). Until 2014, there were 7 EU OPs, in 2014 8 EU OPs received the title 'Horizon 2020'. The latest EU OP for 2021-2027 is named 'Horizon Europe'.
- As a part of 'Horizon 2020', the European Commission has established a new mechanism for promoting collaboration in the field of cutting-edge basic science, i.e. the European Research Council (ERC).



European countries according to the ISO standard

ISO 3166-1 alpha-2



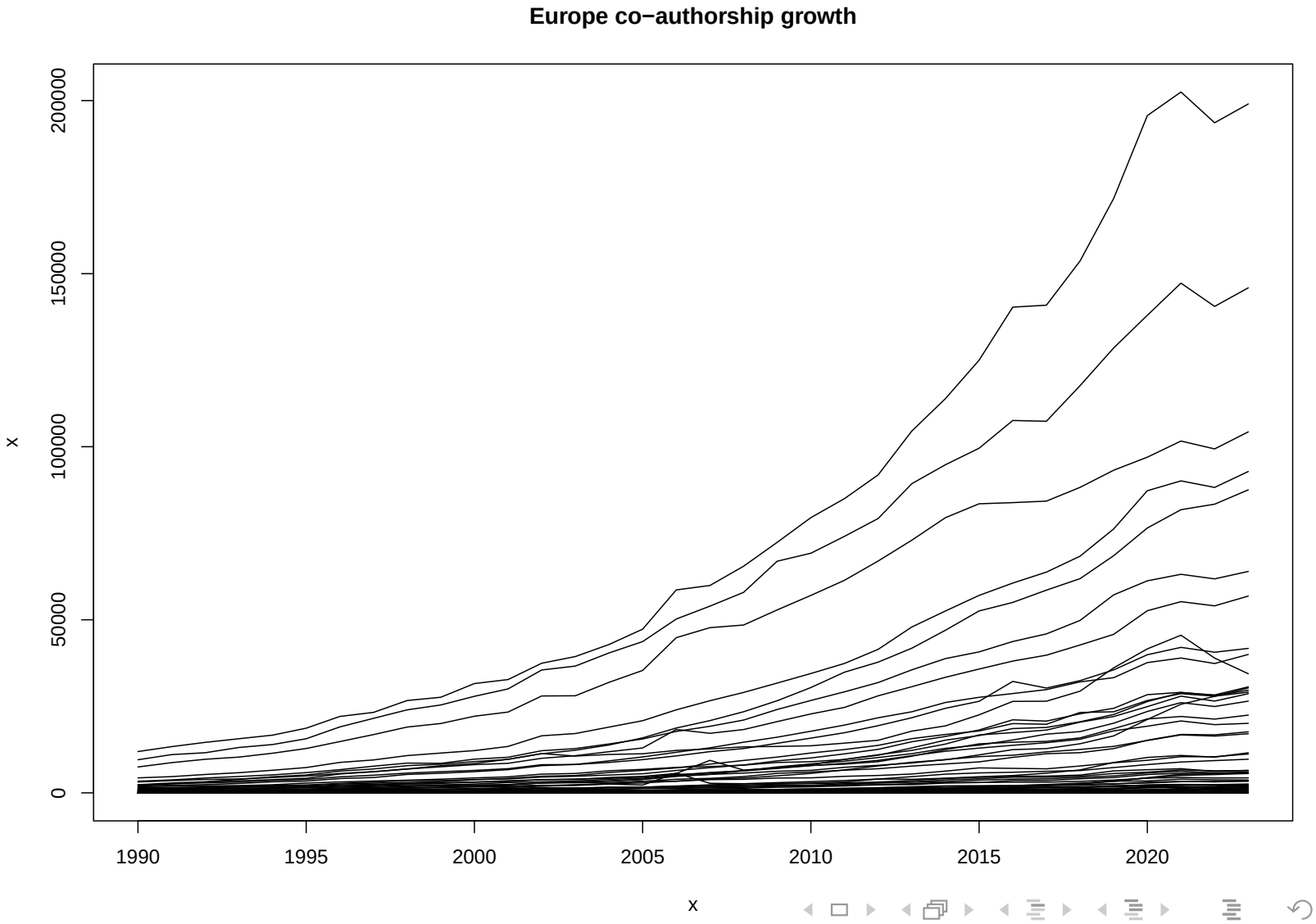
European countries according to the ISO standard

AD Andorra	FI Finland	LT Lithuania	SJ Svalbard+JM
AL Albania	FO Faroe i/DK	LU Luxembourg	SK Slovakia
AM Armenia	FR France	LV Latvia	SM San Marino
AT Austria	GB G Britain	MC Monaco	TR Turkey
AX Aland/FI	GE Georgia	MD Moldova	UA Ukraine
AZ Azerbaijan	GG Guernsey/GB	ME Montenegro	VA Vatican
BA Bosnia+Herz	GI Gibraltar/GB	MK N Macedonia	XK Kosovo
BE Belgium	GR Greece	MT Malta	
BG Bulgaria	HR Croatia	NL Netherlands	
BY Belarus	HU Hungary	NO Norway	
CH Switzerland	IE Ireland	PL Poland	
CY Cyprus	IM i of Man/GB	PT Portugal	
CZ Czech rep	IS Iceland	RO Romania	
DE Germany	IT Italy	RS Serbia	
DK Denmark	JE Jersey/GB	RU Russia	
EE Estonia	KZ Kazakhstan	SE Sweden	
ES Spain	LI Liechtenstein	SI Slovenia	

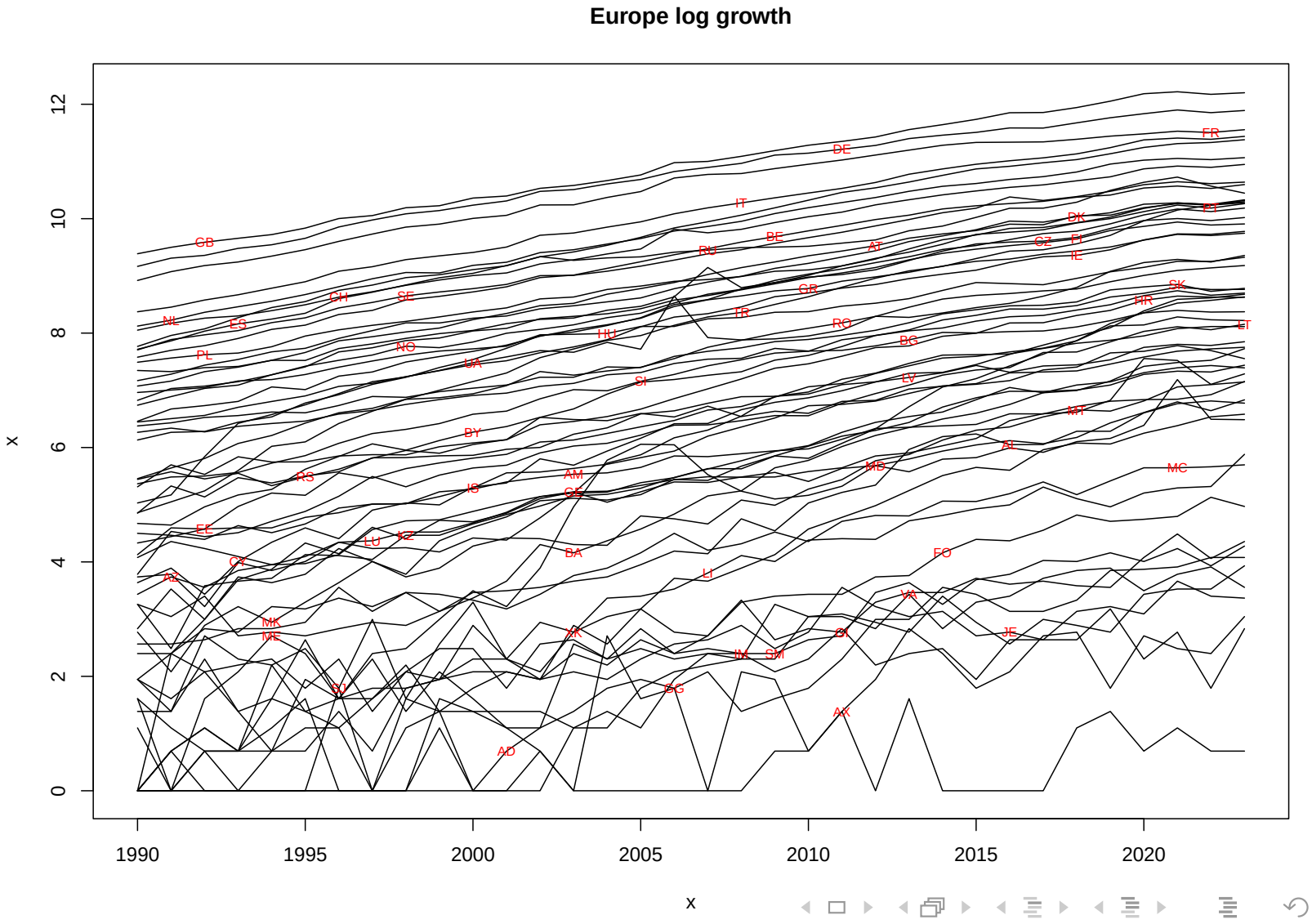
Data - OpenAlex

- The data were obtained using OpenAlex – a fully open catalog of the global research system.
- It is named after the ancient Library of Alexandria and it is made by the non-profit OurResearch.
- Batagelj developed an R package of functions OpenAlex2Pajek for constructing bibliographic networks from selected bibliographic data in OpenAlex.
- Using OpenAlex2Pajek the temporal sequence of co-authorship networks between EU countries in the years 1990-2023 was constructed.

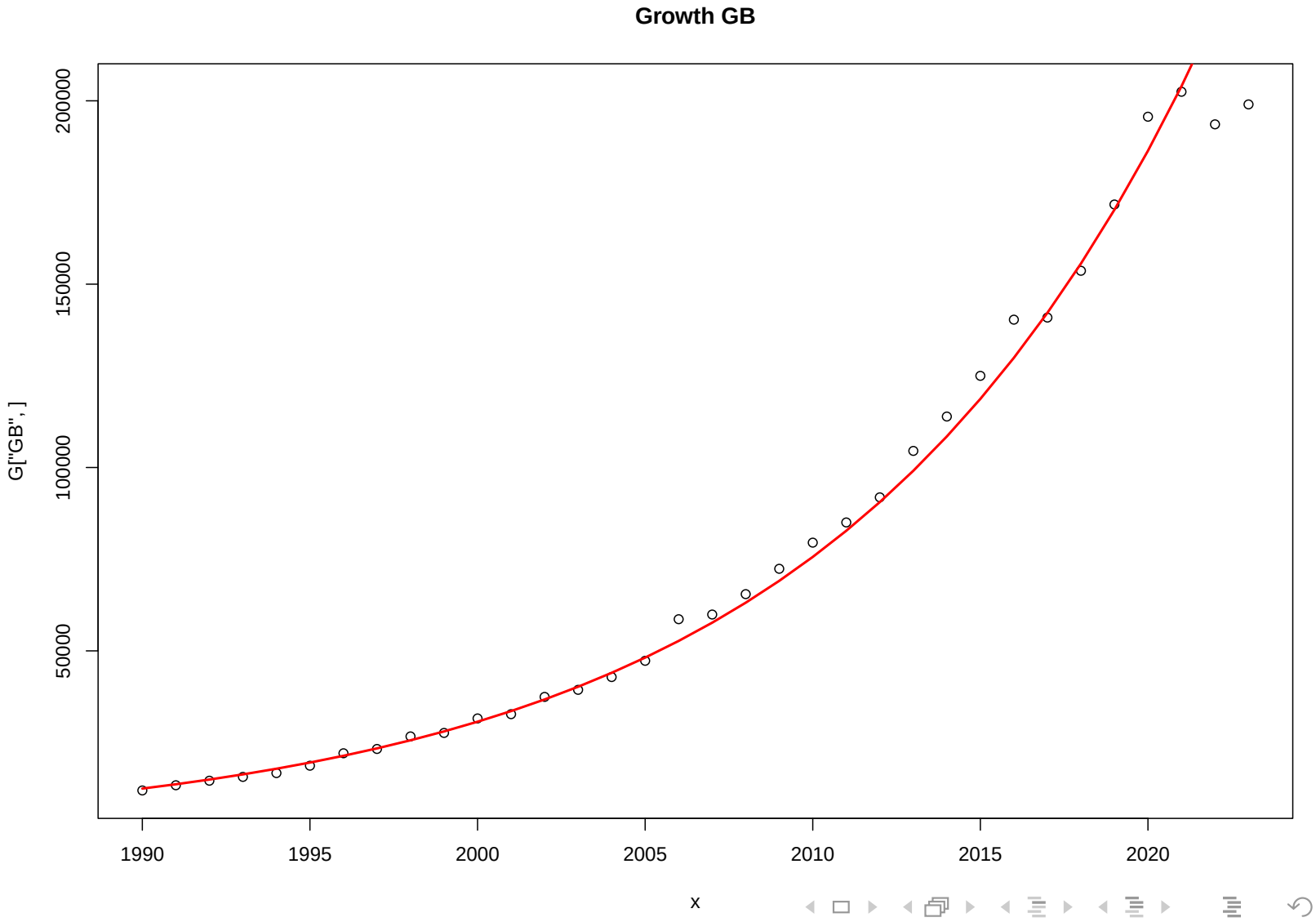
Number of co-authorships for European countries



Number of co-authorships for European countries



Number of co-authorships for European countries



Time periods

The data were organized in three 10-year intervals:

1. **Period 1, 1994-2003: Internet**

   The beginning of the usage of the Internet;

2. **Period 2, 2003-2013: Enlargement of EU**

   The period of the enlargement of EU: the founding members of the EU are Belgium, France, Germany, Italy, Luxembourg, and the Netherlands (1957), in 2004, 10 European countries (Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia) became members of the EU, in 2007 Bulgaria and Romania, and 2013 Croatia;

3. **Period 3, 2014-2023: ERA**

   The implementation of ERA (European Research Area) by many research initiatives in the European Union to speed up the scientific collaboration between EU countries (e.g., the Horizon Europe Programme).

Number of co-authorships between European countries

	number of co-authorships
1994-2003	3 108 474
2004-2013	9 172 425
2014-2023	23 396 350
total	35 677 249

The effects of all three factors (Internet, enlargement of EU, and the implementation of ERA have a cumulative impact which results in exponential growth of the collaboration between European researchers and thus between European countries.

Normalization

- When studying scientific collaborations at the country level, the varying sizes of the countries can make comparisons of collaboration patterns less comparable. To address this issue, network normalization techniques can be applied (e.g., Zitt et al. 2000; Luukkonen et al. 1993; Leydesdorff 2008; Yamashita, Okubo 2006).
- Luukkonen et al. (1993) and Wagner and Leydesdorff (2005) noted the different effects of the normalization approaches. For example, Salton's measure underestimates the collaboration of smaller countries with larger countries.
- Matveeva et al. (2023) studied scientific collaborations among post-Soviet countries, and examined the effect of different network normalizations on the blockmodeling solution. The results show that different normalizations provide distinct perspectives on the structure of scientific collaboration. Country size has the greatest effect when using affinity normalization, while no size effect is observed with activity normalization (e.g. Balassi normalization).

Balassa normalization

- Balassa normalization measures the intensity of collaboration between two countries by taking into account all collaborations in the network.
- It compares observed and expected number of collaborations.
- This normalization is sensitive to the set of countries included in the analysis.
- This normalization consider the sizes of both countries.

Dynamic blockmodels

Cugmas and Žiberna (2022, 2024) compared most of the longitudinal blockmodeling approaches:

- Dynamic Stochastic Blockmodels for Time-Evolving Social Networks (Xu and Hero 2014)
- Stochastic Block Transition Models for Dynamic Networks (Xu 2015)
- Statistical clustering of temporal networks through a dynamic stochastic block model (Matias and Miele 2016)
- Blockmodels for generalized multipartite networks (Bar-Hen et al. 2020)
- Bayesian stochastic blockmodeling (Peixoto 2020)
- A stochastic blockmodel approach for the analysis of multilevel networks (Chabert-Liddell 2022)
- Stochastic blockmodeling for linked networks (Škulj and Žiberna 2021)
- K-means-based algorithm for blockmodeling linked networks (Žiberna 2020)

... Dynamic blockmodels

According to the data of European countries co-authorship (interval scale, approximate normal distributed links, negative links, expected change in a global network structure) we decided to use:

- Dynamic Stochastic Blockmodel (Matias and Miele, 2017)
- K-means-based algorithm for blockmodeling linked networks (Žiberna 2020)

Both approaches are recommended by Cugmas and Žiberna (2022, 2024) according to their simulation study.

Dynamic Stochastic Blockmodel

(Matias and Miele, 2017)

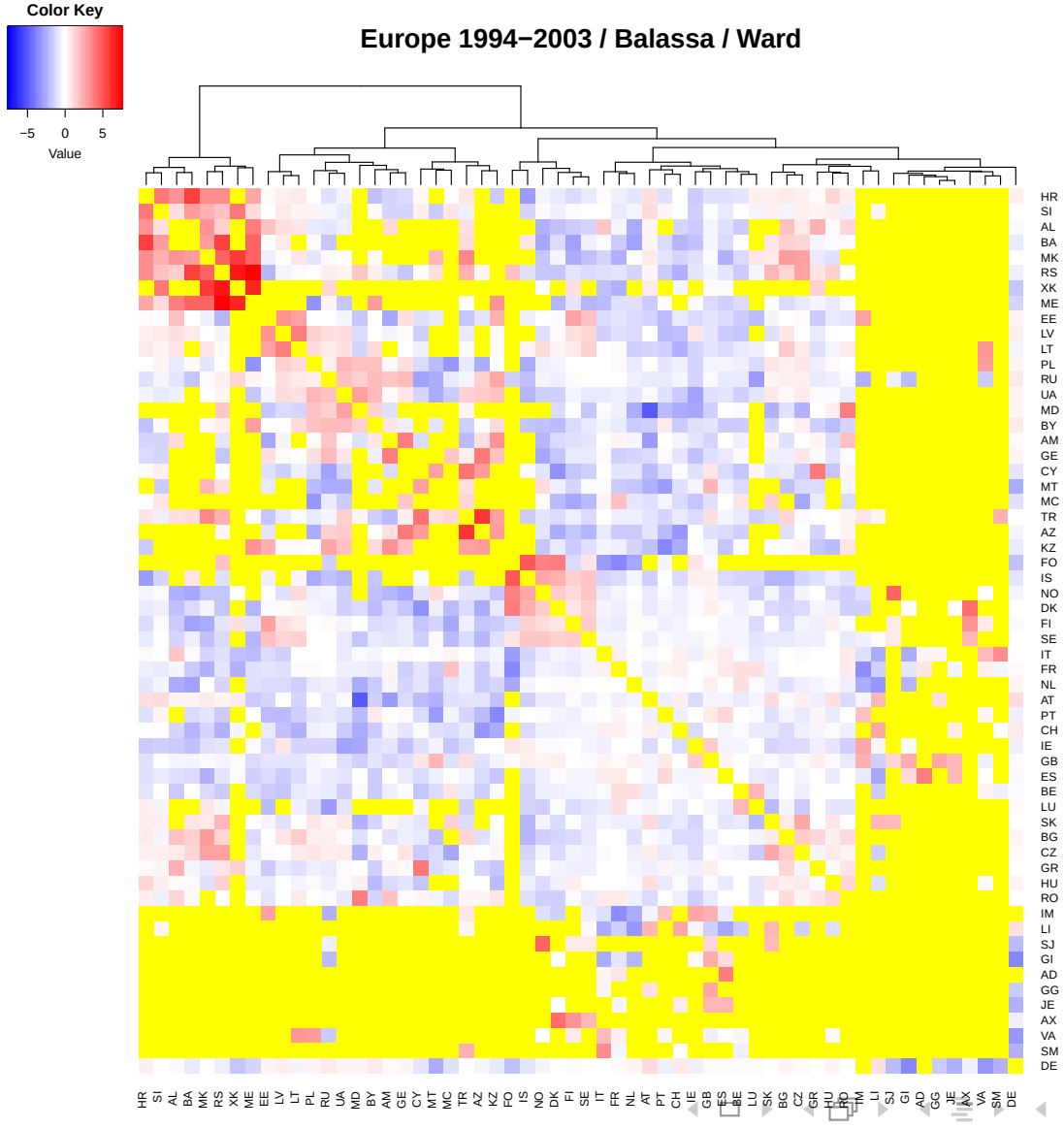
- Matias and Miele (2017) proposed a dynamic stochastic blockmodel allowing control for label switching, that is, they ensure the same cluster has the same label across all time points.
- Their model is based on a stochastic blockmodel (SBM); more precisely, each time point is modelled as an SBM. The tie probabilities at a given time point depend solely on the latent groups/clusters, namely, only on the partition at this time point.
- This indicates that the connectivity parameters (blockmodel) can be different across time points, although changes of connectivity parameters across time are restricted.
- The time points are connected by making the cluster probabilities for each time point (except the first) dependent on membership in the cluster at the previous time point.

K-means-based algorithm for BM linked networks (Žiberna 2020)

- It is k-means-based algorithm for blockmodeling linked networks where linked networks are defined as a collection of one-mode and two-mode networks in which units from different one-mode networks are connected through two-mode networks.
- Examples of linked networks include multilevel networks, **dynamic networks**, dynamic multilevel networks, and meta-networks.
- The algorithm is based on the two-mode k-means (or KL-means) algorithm for two-mode networks or matrices.

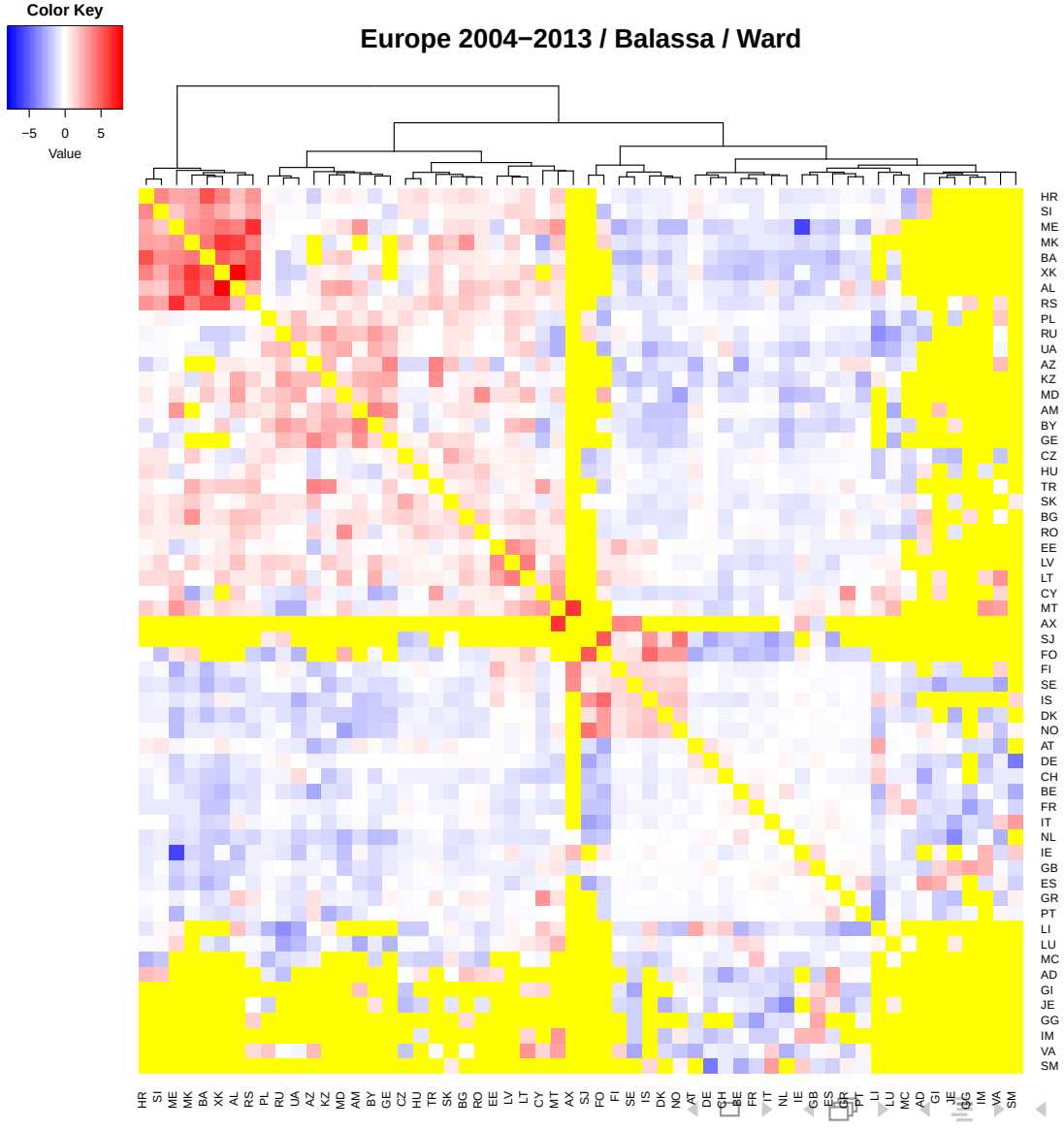
Indirect blockmodeling for 1994-2003

corrected Euclidean distance, Ward method



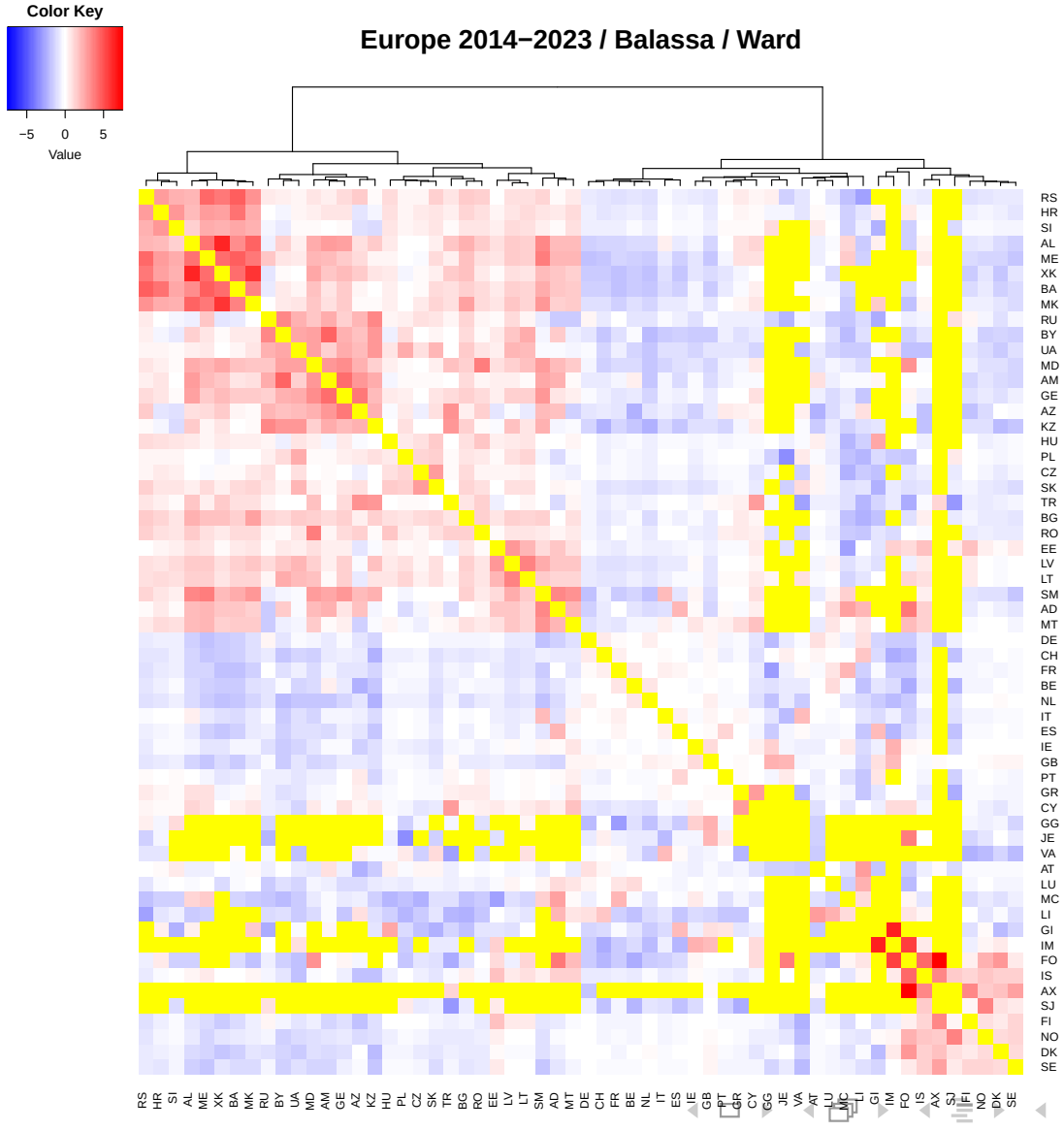
Indirect blockmodeling for 2004-2013

corrected Euclidean distance, Ward method



Indirect blockmodeling for 2014-2023

corrected Euclidean distance, Ward method

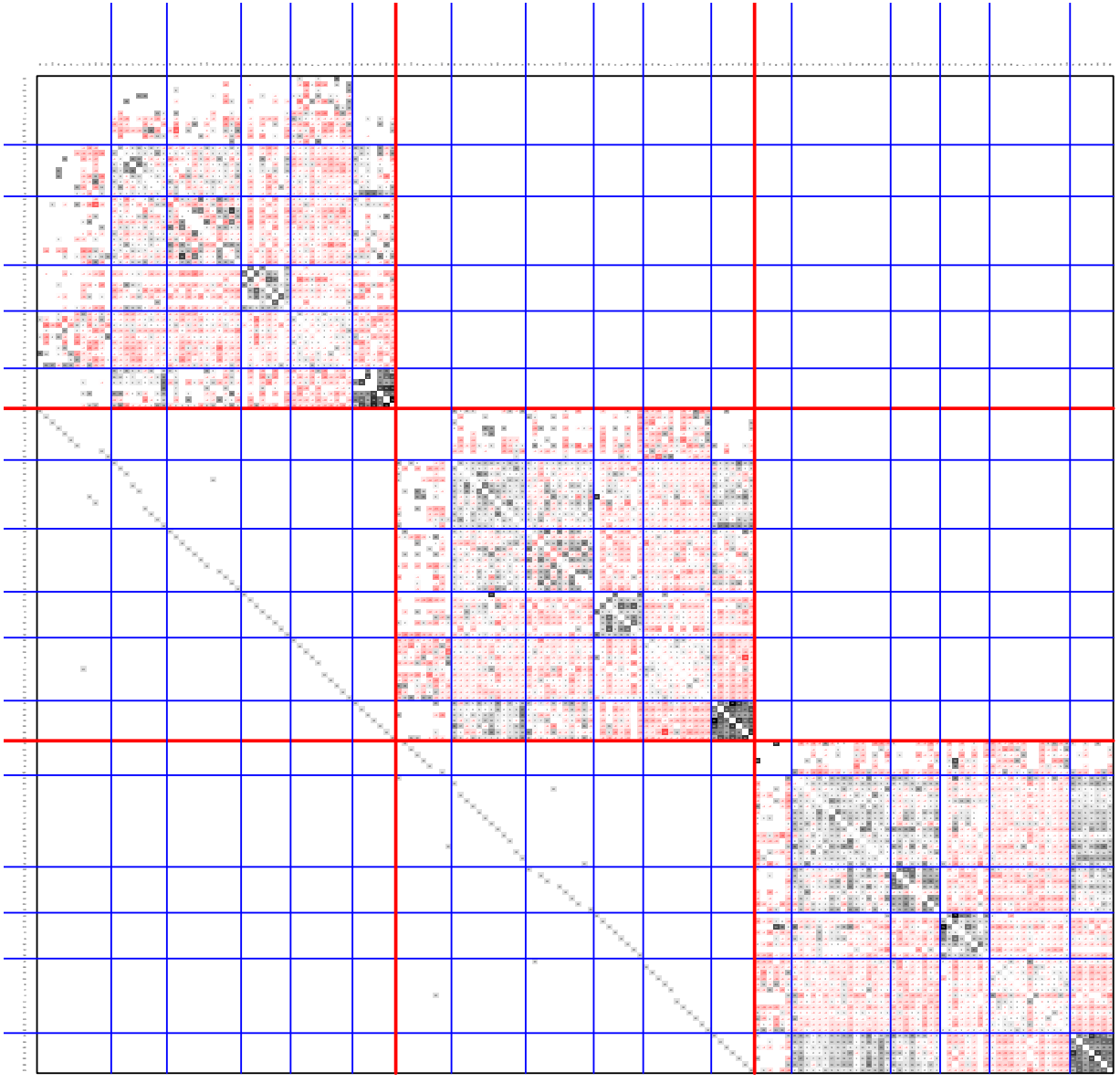


Clusterings into 7 clusters for each time period

1. **Balkan countries** (stable in all periods):
   AL, BA, HR, XK, MK, ME, RS, SI
2. **Eastern countries**:
   1994-2003 : BG, CZ, HU, RO, SK, IM, LI
   2004-2013 : BG, CZ, HU, RO, SK, EE, LV, LT, TR, CY, AX, MT
   2014-2023 : BG, CZ, HU, RO, SK, PL, TR
3. **Mostly Post-Soviet countries** (decreasing in time):
   1994-2003 : AM, BY, GE, RU, UA, MD, PL, EE, LV, LT
   2004-2013 : AM, BY, GE, RU, UA, MD, PL, AZ, KZ
   2014-2023 : AM, BY, GE, RU, UA, MD, AZ, KZ
4. **Scandinavian countries** (incresing in time):
   1994-2003 : DK, FO, FI, IS, NO, SE
   2004-2013 : DK, FO, FI, IS, NO, SE, SJ
   2014-2023 : DK, FO, FI, IS, NO, SE, SJ, AX, GI, IM
5. **Western countries and small countries** :
   1994-2003 : AD, SM, SJ, AX, GI, GG, JE, VA, DE
   2004-2013 : GB, LU, IE, LI, MC, GR, PT, ES
   2014-2023 : GB, LU, IE, LI, MC, GR, PT, AT, CY, GG, JE, VA
6. **Weastern countries**:
   1994-2003 : FR, NL, BE, IT, CH, AT, ES, PT, IE, GB, LU
   2004-2013 : FR, NL, BE, IT, CH, DE, AT
   2014-2023 : FR, NL, BE, IT, CH, DE, ES
7. **Mix**:
   1994-2003 : CY, MT, MC, TR, AZ, KZ
   2004-2013 : GG, JE, VA, GI, AD, SM, IM
   2014-2023 : EE, LV, LT, SM, AD, MT

Results of Dynamic Stochastic Blockmodel

black - more than expected co-authorship, red - less than expected



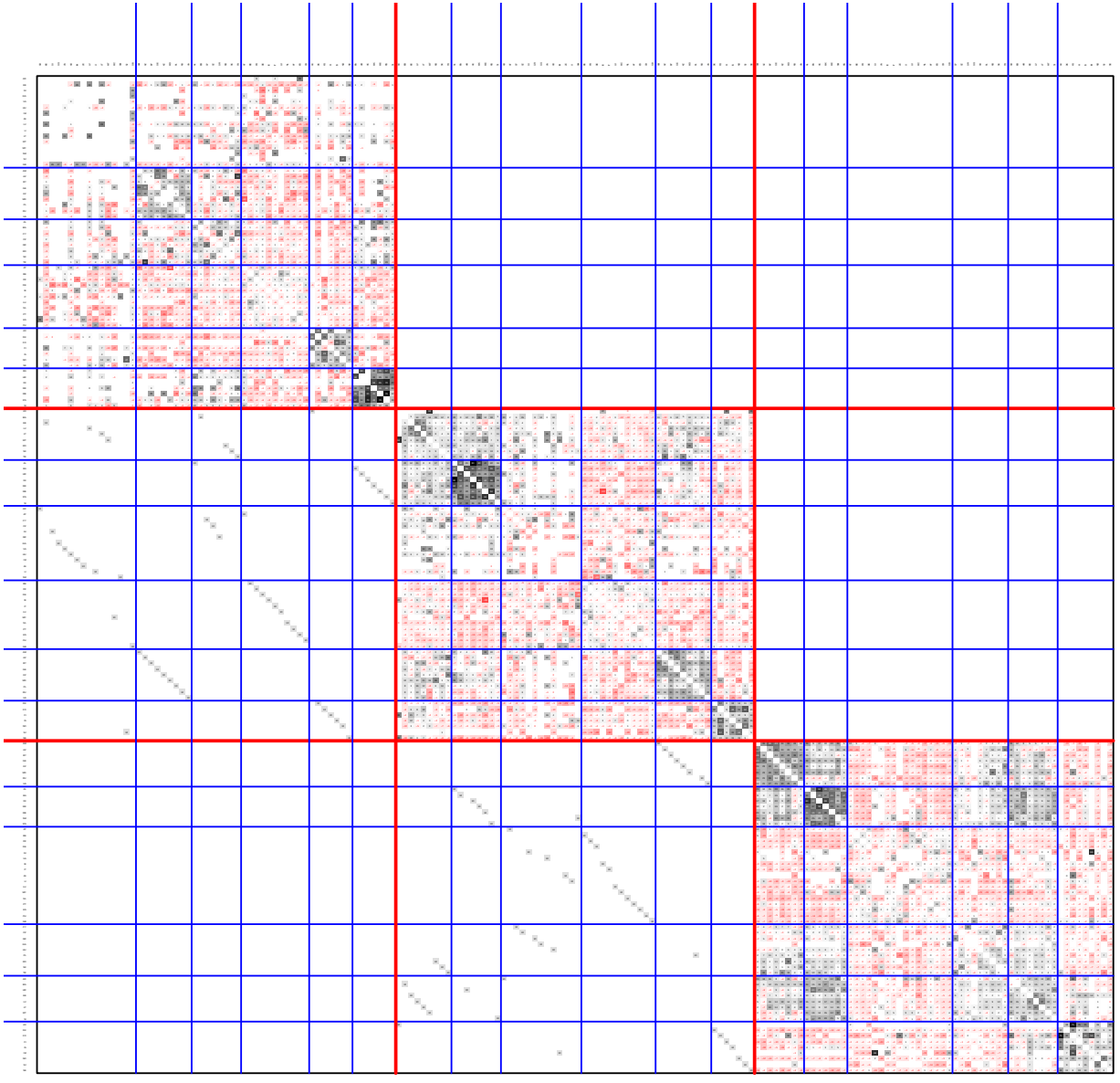
Clusterings of Dynamic Stochastic Blockmodel

1. **Small, less collaborative countries** (decreasing in time):
   1994-2003 : GI, GG, VA, IM, JE, MC, SM, AD, LI , LU, MT, MD
   2004-2013 : GI, GG, VA, IM, JE, MC, SM, AD, LI
   2014-2023 : GI, GG, VA, IM, JE, MC, SM
2. **Eastern countries** (incresing in time):
   1994-2003 : BG, CZ, EE, LV, LT, PL, RO, SK, SI
   2004-2013 : BG, CZ, EE, LV, LT, PL, RO, SK, SI, HU, MD, MT
   2014-2023 : BG, CZ, EE, LV, LT, PL, RO, SK, SI, HU, MD, MT, CY, SM, TR
3. **Mostly Post-Soviet countries** (decreasing in time):
   1994-2003 : AM, AZ, BY, GE, GR, KZ, RU, TR, UA AT, CY, HU
   2004-2013 : AM, AZ, BY, GE, GR, KZ, RU, TR, UA, AT, CY
   2014-2023 : AM, AZ, BY, GE, GR, KZ, RU, TR, UA
4. **Scandinavian countries** (stable in all periods):
   AX, DK, FO, FI, IS, NO, SJ, SE
5. **Western countries** (incresing in time):
   1994-2003 : BE, FR, DE, IE, IT, NL, PT, ES, CH, GB
   2004-2013 : BE, FR, DE, IE, IT, NL, PT, ES, CH, GB, LI, LU
   2014-2023 : BE, FR, DE, IE, IT, NL, PT, ES, CH, GB, LI, LU, AT
6. **Balkan countries** (stable in all periods):
   AL, BA, HR, XK, MK, ME, RS

The description of the blockmodels over time

- The highest collaboration is between the Balkan countries in all three time periods. It is higher than expected also within Scandinavian countries, post-Soviet ones, and a bit less within Eastern countries in all three time periods. The collaboration within the Western cluster is as expected in all periods.
- In 1994-2003 the collaboration between clusters of countries was less than expected, the only exceptions were Balkan countries which collaborated more than expected with Eastern countries and also post-Soviet ones. In time this intensifies.
- In time also the collaboration between Eastern and post-Soviet intensifies.
- A quite interesting result is that the Western countries collaborate with the countries of other clusters less than expected in all three time periods.

Results of K-means-based algorithm for blockmodeling linked networks



Clusterings of K-means-based algorithm for blockmodeling linked networks

The obtained clusters are very similar

1. **Mostly small, less collaborative countries**
2. **Eastern countries** (incresing in time)
3. **Mostly Post-Soviet countries** (decreasing in time)
4. **Scandinavian countries** (stable in all periods)
5. **Western countries** (incresing in time)
6. **Balkan countries** (stable in all periods)

But the stability of clusters over time is much smaller. Quite stable clusters are: Balkan, Skandinavian, and Post-Soviet.
The most unstable is Eastern cluster.
Using Matias and Miele approach, only 10 countries switch the clusters, in Žiberna's approach 26 countries (see the non-diagonal blocks).

Conclusions

- The effects of all three factors (Internet, enlargement of EU, and the implementation of ERA) have a cumulative impact which results in exponential growth of the collaboration between European researchers and thus between European countries.
- The scientific collaboration between European countries is strongly affected by geographical proximity. This effect is not decreasing over time. Self-regulatory factors (historical reasons, language similarity) have a strong effect.
- The Eastern and Western clusters increase with the inclusion of small countries.
- The collaboration is higher than expected between Balkan, Eastern and post-Soviet countries and in time it intensifies. The collaboration between Western countries and all others is lower than expected in all time periods.
- The three approaches give similar results. By the indirect blockmodeling approach, the interval of possible number of clusters was obtained (6 - 8 clusters). The Dynamic Stochastic Blockmodel procedure is the most appropriate for the presented problem and data.