

ECONOMIC
RESEARCH
F O R U M



منتدى
البحوث
الاقتصادية

2015

working paper series

**UNDERSTANDING DEMOCRATIC TRANSITION
USING SELF-ORGANIZING MAPS:
A SPECIAL FOCUS ON ARAB SPRING COUNTRIES**

**Houda Haffoudhi, Racem Mehdi
and Gam Abdelkader**

Working Paper No. 958

**UNDERSTANDING DEMOCRATIC TRANSITION USING
SELF-ORGANIZING MAPS: A SPECIAL FOCUS
ON ARAB SPRING COUNTRIES**

Houda Haffoudhi, Racem Mehdi and Gam Abdelkader

Working Paper 958

October 2015

Send correspondence to:

Houda Haffoudhi

FSEGN, University of Carthage

h_haffoudhi@yahoo.fr

First published in 2015 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Copyright © The Economic Research Forum, 2015

All rights reserved. No part of this publication may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without permission in writing from the publisher.

The findings, interpretations and conclusions expressed in this publication are entirely those of the author(s) and should not be attributed to the Economic Research Forum, members of its Board of Trustees, or its donors.

Abstract

The aim of the paper is to identify, in the light of the Arab spring, the conditions favorable for democratic transition and to analyze the specific experience of Arab countries concerned by regime change. We use a large set of indicators and the self-organizing map (SOM) approach as an alternative analytical tool. Countries are then mapped considering data related to 33 indicators from 121 countries, four years (1984, 1991, 2002 and 2013), including economic, social, demographic, societal and political variables, as well as indicators related to institutional framework. Mapping allows us to link countries configuration to their democratization levels. Then, the SVM clustering methodology is used to cluster the prototypes produced by the SOM in order to discriminate the future successes and to determine the most influential variable on democratic success. Countries neighborhoods are based on similarity in particular characteristics. Our mapping highlights especially population age structure, globalization, health indicators, education and women participation in the society as the most important variables determining each country location which could be related to democracy level and its change over time. In Addition, unemployment, corruption, democratic accountability, and law and order, are the most influential variables distinguishing failure from democratic transition success.

JEL Classification: P16, C45, C53

Keywords: democracy, democratic transition, Arab spring countries, Kohonen Self-Organizing Maps, Support Vector Machine (SVM)

ملخص

الهدف من هذه الورقة هو تحديد الظروف المواتية للانتقال الديمقراطي في ضوء الربيع العربي، وتحليل تجربة محددة من الدول العربية المعنية بتغيير النظام. نستخدم مجموعة كبيرة من المؤشرات ونهج خريطة التنظيم الذاتي (SOM) كأداة تحليلية بديلة. ثم يتم تحليل البلدان باستخدام البيانات المتعلقة بمؤشرات 33 من 121 بلداً، أربع سنوات (1984، 1991، 2002 و 2013)، بما في ذلك المتغيرات الاقتصادية والاجتماعية والديموغرافية والاجتماعية والسياسية، وكذلك المؤشرات المتعلقة الإطار المؤسسي. رسم الخرائط يسمح لنا لربط التكوين البلدان إلى مستويات الديمقراطية الخاصة بهم. ثم، يتم استخدام منهجية تجمع SVM لتجميع النماذج التي تنتجها SOM من أجل تمييز النجاحات في المستقبل وتحديد المتغير الأكثر تأثيراً على النجاح الديمقراطي. نستند على الدول المتشابهة في خصائص معينة. ونسلط الضوء على رسم الخرائط وخاصة هيكل السكان من حيث العمر، والعولمة، ومؤشرات الصحة والتعليم ومشاركة المرأة في المجتمع. ومن أهم المتغيرات تحديد موقع كل بلد والتي يمكن أن تكون ذات صلة بمستوى الديمقراطية والتغيير مع مرور الوقت. وبالإضافة إلى ذلك، البطالة، والفساد، والمساءلة الديمقراطية، والقانون والنظام، هي المتغيرات الأكثر تأثيراً التمييز بين الفشل والنجاح في التحول الديمقراطي.

1. Introduction

According to Huntington (1991), the last democratization wave began in 1974 with the fall of southern European military dictatorships, which paved the way for successful democratic transitions¹ in this region. After Portugal, Spain and Greece, the democratization wave extended to Latin America, Asia, Central Europe and Africa. The number of countries regarded by Freedom House as democracies tripled from 39 in 1974 to a high of 123 in 2005. However, all countries that experienced democratic transition during this period didn't share the same fate. Some countries succeeded in their democratic transition and became consolidated democracies; others remained unstable democracies; while others returned to autocracy.

In December 2010, a popular protest movement erupted in Tunisia and spread to Egypt, Libya, and across the Middle East. When it started, this political turmoil has been perceived as a new democratization phase affecting the Arab World and a new possibility to extend democracy to countries where autocracy is deeply anchored. Again, it appears that the seed of democracy cannot germinate in all countries. Although Arab Spring countries began in roughly similar places, they move along three different paths after their regime change. The first one is the progressive and continuous democratic transition path in which Tunisia is engaged. In the second path, we find Egypt, which has swung back and forth between autocracy and more democratic rule. The third path results in the state collapse as in Libya and Yemen.

Given these observations, the key question we face is the following: why did some countries move from autocracy to democracy, while others continue to maintain or cannot extricate themselves from autocratic rule? The aim of the paper is to identify favorable conditions for triggering democratic transition in light of the Arab Spring and to explain the specific experience of Arabic countries concerned with regime change. The economic literature points to a large number of factors that could potentially induce democracy and many of them have been investigated through numerous empirical work.

In a series of papers and a subsequent book², Acemoglu and Robinson have developed a powerful framework to analyze the economic origins of dictatorship and democracy. This work has become a major reference point in the field of political economy. In an extreme bounds analysis, Gassebner et al. (2013) find that among 59 factors considered in the literature to explain the emergence and survival of democracy, past transition is one of the most robust determinant of the establishment and consolidation of democracy. They also show that GDP per capita acts only on the survival probability of democracy, but not its emergence, and Muslim countries are less likely to become democracies³. Finally, the influence of military origin of the head of state and the level of democracy in neighboring countries seems to be robust.

In this paper, we use an analysis by the neural network method (Kohonen, 2005) and use a set of 33 indicators. This set includes economic, social, demographic, societal and political variables, as well as indicators related to institutional framework. The database contains a large number of correlated data, sometimes incomplete and inconsistent. This kind of data is well suited for use in neural networks that have some advantages in comparison with the conventional methods (i.e., econometric regressions). Some of these advantages are the non-linear model capability, generalization from examples, robustness to noise and ease of use.

¹ Democratic transition is defined as a switch from autocracy to democracy. According to Przeworski et al. (2000), a democracy is a political system in which key government offices are filled through contested elections, (i.e., the executive and legislative members come to or stay in power after elections in which more than one party has a chance of winning office). In other words, "democracy is a system in which incumbents lose elections and leave office when the rules so dictate" (Przeworski et al. 2000, p. 54).

² Acemoglu and Robinson (2006).

³ As Gassebner et al. (2013) show, this effect is driven by the oil rich Arab countries. This is also the likely reason that the coefficient on fuel exports is statistically insignificant, the two variables being highly collinear.

The use of unsupervised neural model is best suited when outputs are not available or when clustering of input data according to similarity is of interest. Recent advances in clustering with the use of unsupervised neural networks are described in Du (2010). A model which is used widely in this area is the SOM (Martin-del-Brio and Serrano-Cinca, 1993; Kohonen, 1998). SOM is a useful tool because it attains the capability of projecting multidimensional data to a lower number of dimensions, usually two and is less computationally expensive making it more attractive in big data sets. A comparison between SOM and the traditional hierarchical methods used in most of the studies (Mangiameli et al., 1996) has pointed out the superiority of SOM by comparing their accuracy and robustness on 252 datasets capturing dispersion, outliers, irrelevant variables and non-uniform cluster densities.

Our contribution seeks value added in two distinct dimensions: First, extending data set to institutional variables in addition to economic, social, demographic, societal and political factors; and second analyzing and clustering country democracy level, by using the SOM approach.

The rest of the paper is organized as follows. We begin by giving a survey of the literature. Next, we will present our methodology and the motives of our choice. Then, we will define our sample. Later we will present our findings and finally discuss and conclude.

2. The Determinants of Democracy: A Brief Survey of the Literature

Building on ideas that go back in time as far as Aristotle, Lipset (1959) argue in a seminal article that there is a strong, positive association between income levels and democracy in the very long run. The Lipset's modernization hypothesis receives empirical support from Barro (1999), who finds that GDP per capita is a positive determinant of democracy in a large sample. Boix (2011) finds evidence of a positive effect of income per capita on democratization over a long period. Treisman (2011) find evidence of a positive effect of income on democracy, more strongly over the medium run than over the short run. However, the validity of Lipset's hypothesis remains on debate since some empirical works find that GDP per capita is no longer a significant determinant of democracy (Acemoglu, Robinson and Yared, 2005, 2008).

Lipset (1959) also argues that education engenders democracy since educated people are more likely to resolve their differences through negotiation and voting than through violent disputes. A body of work on the political economy of education provision sheds light on the political impact of education. It is based on the connection between education and political participation. According to Bourguignon and Verdier (2000), the educated elite, who rule the country, evaluate the incentives of subsidizing education for the poor. On the one hand, increasing education for the non-educated may raise output due to a technological externality. On the other hand, the elite loses political control by extending education to the masses. Bourguignon and Verdier (2000) show that a more equalitarian distribution of education will lead a society to democratize sooner. Campante and Chor (2012) explain how educated individuals tend to develop greater political engagement. Educated individuals are, on average, more productive in both the production and political spheres. In response to an increase in human capital, a rational citizen will thus choose to expand both production and political effort to make use of her increased capacity. There is a body of work giving empirical support to this hypothesis. Barro (1999) finds that primary schooling is a positive determinant of democracy. Murin and Wacziarg (2013) also show that primary schooling, more so than GDP per capita, and more so than secondary and tertiary education, has been a major factor in democratic transition. Castelló-Climent (2008) find a significant effect of education attained by the majority of population on implementation and sustainability of democracy, rather than the average years of schooling. However, Acemoglu, Robinson and Yared (2005) find that there is no evidence that countries increasing their education are more likely to become democratic.

Acemoglu and Robinson (2001) offer the important, and less than obvious prediction that the relation between inequality and democracy will have an inverse-U shape: at high levels of inequality the elite represses, while at low levels they can accommodate redistributive demands within the authoritarian system. It is thus at intermediate levels of inequality that transitions to democracy are most likely to occur. According to Feng and Zack (1999), democracy emerges when per capita income grows such that a sufficiently large number of middle-wealth agents become politically active and agitate for democracy. Preferences for freedoms, government expenditures on education and the police are included among the factors that are in favor of democracy.

Beyond the growth of middle-wealth agents, demographic factors play a role in the democratization process through youth bulges. In autocratic regimes where large youth groups are excluded from political participation, youth may engage in violent conflict behavior in an attempt to force democratic reform when they aspire for political engagement (Urdal, 2006). In other words, a large proportion of young people in a society may be a cause of destabilization of the pre-existing political regime. This phenomenon is amplified when limited employment opportunities or decreasing wages induce dissatisfaction that lead young people to become the leading group in inducing a need for change. At the same time, the youth bulges have an opposite effect on democratization processes. As young men tend to be responsible for a disproportionate amount of the violence that occurs, it is more difficult to establish democracy in countries where this section of the population is relatively large.

Acemoglu and Robinson (2006) show that globalization promotes democracy through several ways. International financial integration makes the elite feel more secure about democracy since it makes it harder to tax them and reduces the extent to which democracy can pursue highly majoritarian policies. Since trade liberalization has redistributive effects, it affects the inequality between capital owners and labor owners in a way which depends on the country position on the inverted U-shaped relationship between inequality and democratization.

Acemoglu and Robinson (2006) add a set of noneconomic structural factors which contribute to democratization and consolidation, such as civil society and political institutions. A developed civil society plays a key role in challenging the political system in place and transition will be easier since well-organized citizens make it difficult to use repression. Acemoglu and Robinson (2006) suggest that political institutions have a crucial impact on democratization processes. When the elite ruling class can use repression and political stakes are high (rents, tax revenues, etc.), the elite prefer to use repression in order to avoid democratizing, given that democracy will affect their interests. However, if economic institutions limit policies and thus reduce political stakes, change in political institutions may be easier (Acemoglu, 2006).

3. Self- Organizing Maps Analysis

Following the pioneering work of Kohonen (1982), our study employs the self-organizing maps (SOM) approach for analyzing data, which covers indicators capturing determinants of democracy emphasized by the theoretical literature previously presented. Kohonen's (1982) paper was extended by his subsequent work and that of other scholars (Kaski and Kohonen, 1995; Kaski et al., 1998; Kohonen, 2001). Hua et al. (2009) demonstrate the usefulness of SOMs using Central Intelligence Agency World Fact Book Data. To the best of our knowledge, this is the first time this technique is used in a political regime issue.

We recall here the theoretical foundations of the SOM, based on Kohonen (1995), before applying the technique in the following section. The two main objectives of the SOM algorithm are vector quantization and vector projection. The vector quantization aims at summarizing the data by dividing a large set of data points into groups having approximately the same number

of points closest to them. The groups are then represented by their centroid points which typically are vectors obtained as the mean of the points of the respective groups (e.g., the k-means algorithm (Tan, Steinbach, & Kumar, 2006)).

A typical way to assess the quality of the resulting quantization is to calculate the mean quantization error (MQE) (Kohonen, 2001; Polzlbauer, 2004). The mean quantization error (MQE) is calculated by averaging all euclidean distances between the different input vectors and their respective closest neurons. A low MQE value indicates that a good representation of the input by the SOM is achieved. The interpretation of the obtained value depends on the scale of the input variables. The second objective is vector projection in which the dimensionality of the data points is reduced by projection onto lower dimensional maps (e.g., the Principal Component Analysis (Jolliffe, 2005)). Typically, a projection to two dimensional maps is performed in order to be able to visualize and represent the different variables on classical reporting supports.

The projection is performed with the neurons obtained after the quantization phase. In a case of a good projection onto the two-dimensional maps, neurons close to each other in the high dimensional space should be mapped to position close to each other in the low dimensional space. The combination of vector quantization and projection enables the exploration of the data and to use techniques like visual correlation analysis or clustering analysis in an intuitive manner while keeping a mapping between the input vectors and the neurons in the low dimensional space.

SOMs represent a special case of neural networks, an exploratory data analysis technique where multi-dimensional data are projected onto a two-dimensional space to allow for clear visualization of the data and easy identification of groups with similar characteristics. Kohonen maps can be thought of as a factor analysis combined with a cluster analysis. A major advantage of Kohonen maps is the self-organizing property of the map, which makes estimated components vary in a monotonic way across the map (Deichmann et al., 2007).

As a brief overview, the SOM algorithm can be briefly and intuitively described as a special case of a competitive neural network, where output nodes compete to become the winning node. The winning node is the one that carries the highest value for a certain score function. This node becomes the center of a neighborhood and attracts similar neurons to it. The weights of neighboring nodes are adjusted via a linear combination of the input vector and the current weight vector in order to improve the score function. Convergence occurs when little or no change arises in the vector of weights. After convergence, the estimated components in the vector of weights arrange themselves onto the hexagonal lattice in a structured manner (Larose, 2005). For a thorough explanation of how to interpret SOM output, we refer the reader to Kaski and Kohonen (1995), Kohonen (2001), Deichmann et al. (2007), and Hua et al. (2009).

The SOM is a two-layer unsupervised neural network that maps multidimensional data onto a two dimensional topological grid or map (Kohonen, 2001). The data is grouped according to similarities and patterns found in the dataset, using some form of distance measure, usually the Euclidean distance. The result is displayed as a series of nodes or points on the map which can be divided into a number of clusters based upon the distances between the nodes. As the SOM is unsupervised, no target outcomes are provided, and the SOM is allowed to freely organize itself based on the patterns identified, making the SOM an ideal tool for exploratory data analysis. "Exploratory data analysis methods, like SOM, are like general-purpose instruments that illustrate the essential features of a data set, like its clustering structure and the relations between its data items" (Kaski and Kohonen, 1996). Thus, the SOM can be said to perform visual clustering of data.

4. Data Description

Our empirical analysis uses data collected from various sources. After excluding countries for which there are missing data, the selected sample consists of 121 countries. The set of the selected indicators includes economic, social, demographic, societal, political and institutional variables:

- As economic indicators, we consider GDP per capita, unemployment, trade openness and a globalization index.
- The set of social indicators covers three dimensions: health, measured by life expectancy at birth and mortality; education, captured by the educational attainment for total population; and internet access.
- The demographic factors encompass the population structure by age and the urban population.
- Societal indicators include religion (Protestant, Catholic, Muslim, Orthodox), colonization legacy (British colonial legacy or not), fractionalization (ethnic fractionalization, linguistic fractionalization and religious fractionalization) and set of variables reflecting women's participation in society (women's economic, social and political rights).
- The set of political and institutional indicators cover military in politics, external conflict, monarchy indicator, electoral self-determination, political participation rights, worker's rights, bureaucracy quality, corruption, law and order, democratic accountability and government stability.

The indicators are listed with more details in Table 1. Their selection is based on previous studies dealing with democratization. Following Deichmann et al. (2007), a temporal dimension is added to the present analysis through four years: 1984, 1991, 2002 and 2013. In the case of Ghana, for example, these time periods are labeled on our maps as gha, Gha, GHa and GHA for 1984, 1991, 2002, and 2013 respectively. By doing this, we can trace the movement of Ghana's position over time.

In order to avoid the elimination of a whole country's data due to missing values of one or two observations, we treat missing values with interpolation and by calculating means for three-year periods or by considering the value of a year after or before the missing observation. This is important because many of the data are not reported on an annual basis in some countries.

For democracy measure, we refer to POLITY IV to deduce democracy scores for each country per the following rules:

- Score = 1 for very stable autocracy,
- Score = 2 for stable autocracy,
- Score = 3 for unstable autocracy,
- Score = 4 for new autocratic country,
- Score = 5 for new partially democratic country,
- Score = 6 for unstable partial democracy,
- Score = 7 for stable partial democracy,
- Score = 8 for very stable partial democracy,
- Score = 9 for new fully democratic country,
- Score = 10 for unstable full democracy,
- Score = 11 for stable full democracy, and
- Score = 12 for very stable full democracy.

When POLITY value is a negative or zero, the country is considered an autocracy; when POLITY value is between 1 and 6, the country is considered a partial democracy; when

POLITY value is between 7 and 10, the country is considered a full democracy. A very stable regime means that a country has the same Polity value throughout the data set; a stable regime means that a country remains in the same category (autocracy, partial or full democracy) throughout the data set but Polity value changes or that a country has shift from a category to another one in one direction; a new autocratic, a partially or fully democratic country means that a newly established country (country with no Polity value before) chooses an autocratic, partially or fully democratic regime.

5. Empirical Results

A number of packages are available to construct Kohonen maps, but for the purpose of our analysis we use the SPSS modeler that produces clear and easily understood graphs.

5.1 Exploration of SOM Maps

The Kohonen analysis yields several plots and component plots. Each country on the map is positioned in that cell of a grid whose estimated vector is closest to its own vector of indicators. The size of the map is influenced by the purpose of the study. If clustering is desired, a small map is created (Deboeck, 1998). If visualization and detail are desired, a larger map is created. We choose a large map, with a large number of nodes available to be fine-tuned to specific countries, allowing for greater detail and accuracy.

We start by training a self-organizing map of 15 by 15 neurons. With this size, it can be expected that each neuron will be the Best Matching Unit for, at most, a few observations and this allows a clear visualization of the map. Maps (Figure 1 and 2) describe countries' positions according to their own configuration defined by economic, social, demographic, societal, political and institutional factors in 1984, 1991, 2002 and 2013. For each individual position, we introduce the democracy score.

In figure 1, we can see that developed countries are likely to be projected on the right-down corner of this map. For these countries, the observations from the four years lie close to each other. The observations are either projected on the same neuron or on adjacent neurons.

The starting point of our analysis is 1984. This year captures the democratic transition in Southern Europe (Portugal, Spain and Greece) and Latin America (Argentina, Bolivia, Dominican Republic, Ecuador, El Salvador, Honduras, and Peru) that occurred from the year 1974 (i.e., the beginning of the last wave of democratization according to Huntington (1991)). In the map, Honduras, Bolivia, Argentina, Ecuador and Peru are very close to Spain, which appears as an epicenter for countries involved in this democratization phase. The 'wave' metaphor, to borrow Huntington's term, can be used here to denote the spread of democracy in countries with relatively similar situations.

The second phase of democratization is incorporated in the mapping of countries' positions in 1991. After the fall of the Berlin wall and the collapse of Soviet Union, several countries abandoned communism and joined the ranks of democracy. Estonia, Latvia, Lithuania, Slovenia, Ukraine, Armenia, Moldova and Russia formed a homogenous group accessing to democracy, as they are mapped in the same neuron in 1991. Among former communist European countries engaged in a democratic transition, two are located not far from this group of countries: Romania and Bulgaria; and two others are mapped with countries regarded as well-established democracies (in the right down corner of the map): Poland and Hungary. Note that Mongolia belonged to the second group of former communist countries.

The second phase of democratization has deeply affected Latin America. Brazil, Chile, Guatemala, Nicaragua, Panama, Paraguay and Uruguay experienced democratic transition in this phase. Except for Uruguay, we observe a shift in countries' location toward positions reflecting conditions close to those favorable for democratic transitions during the previous

phase (i.e., positions around Spain). We clearly see a democratic impulsion in Latin America countries when they converge towards a given configuration.

Three Asian countries in our sample have been involved in the second phase of democratization: Bangladesh, Philippine and Republic of Korea. The democratization of Bangladesh occurred together with a change of its localization in the map, which tends to move towards the position of a democratic country, like Turkey. The Philippine's position has also moved closer to Latin American countries; Philippines and Guatemala are plotted in the same neuron). Finally, the Republic of Korea did not need to change its situation to be able to become a democracy since the map reveals closeness to well-established democracies.

We can draw a third phase of democratization when we focus on the year 2002. After moving over time inside the area favorable for a democratic transition in Latin America, Mexico became finally a democratic country. A democratic transition emerged in Guyana when it attained a position close to well-established democracies. The Latin American transition is considered the longest and the deepest wave of democratization in history.

It was in Africa where this third democratization phase has the deepest impact. Several African countries acceded to democratic regime. The democratic transition in Mozambique took place in the area that captures conditions conducive to democratic transitions for Latin America countries. Guyana, Ghana and Malawi became democratic countries after moving toward well-established democracies. Note that Guyana and Ghana shared the same configuration both in 1984 and in 2002. With Ghana and Malawi, we find other African countries that were engaged in a democratic transition situated in a delimited area in the lower part of the map: Kenya, Sierra Leone, and Lesotho. A second group of African countries involved in democratic transition is mapped in the left upper corner of the map: Mali, Senegal, and Niger.

We remark that East European countries, which begin in roughly similar places, move along two different trajectories: Slovakia, Estonia and Latvia move toward full democracies, whereas Romania, Moldova, Ukraine, Armenia and Bulgaria remain near Russia.

Remarkably, the countries organized themselves in such a manner that they do not only mimic their relative geographical locations. Countries with pair adjacent to those of democratic ones will follow them in their democratic transition. SOM analysis reveals that democratic countries have similar socio-economic conditions. Democratic contagion is not delimited geographically. Countries with adequate social and economic conditions saw democratization.

In 2013, we recorded the last phase of democratization. In fact, additional African countries acceded to democratic regimes. Like in Mozambique, the democratic transition in Gabon, Ivory Coast and the Democratic Republic of Congo occurred near the area capturing conditions conducive to democratic transitions for Latin American countries. Liberia became a democracy after moving over time toward a position close to well-established democracies. It shares the same neuron with Jamaica. The Zimbabwe position in the map was specific when it abandoned its autocratic regime.

The Arab Spring happened during this last phase of democratization. To highlight their democratic transition, we consider figure 2. The MENA region is divided into two groups, which remain the same during the considered period. The first group is situated in the lower left corner (Qatar, Kuwait, Arab Saudi, Jordan, Egypt, Bahrain, Iraq, and the UAE). The second group is mapped in the upper part (Tunisia, Libya, Syria, Yemen, Turkey, Iran, and Algeria). The two groups reveal a distinction based on their colonial backgrounds. The first group has a British colonial heritage, whereas the second presents different colonial background.

The democratic transition appears in 2002 for Algeria and in 2013 for Yemen, Tunisia and Iraq. Although Libya and Syria saw popular uprisings, they failed to join the ranks of democracies

and suffered from state collapse. Egypt saw nonviolent democratic protests but fell back to a military autocracy. In 2013, Tunisia was situated between Algeria, Turkey, Albania and Senegal whereas Yemen is isolated in the map. This mapping reveals a close configuration between Tunisia and these democratic countries which may explain the trajectory taken by Tunisia compared to other Arab Spring countries. In the lower left corner, we find Iraq which is mapped with Sudan.

5.2 Discussion results

Recall that a country's position on the maps is deduced from various indicators capturing determinants of democracy emphasized by theoretical literature. The most important variables determining our mapping encompass population age structure, globalization, health indicators, education and women's participation in society. In other words, when countries are mapped together this is due to the fact of closeness in demographic structure, degree of integration in the world economy, social conditions, and status of women.

We focus here on the impact of population age structure on political change as the most important determinant of country pairs in the SOMs. We show previously that a country may move toward a position associated to a configuration close to conditions conducive to democratic transitions in other countries. A statistical analysis of our data reveals that one of these conditions is a pear-shaped population pyramid or mature population structure. Then, countries engaged in a democratic transition catch up with countries previously presenting this population structure (i.e., principally democratic countries). It is the case of Latin American and some African countries, which converged to a structure characterized by a mature population when they became democracies. Former communist countries are initially mapped near to Greece because of their close population structure. In subsequent periods, they moved differently depending on their demographic structure change. Estonia, Latvia joined the well-established democracies while Ukraine and Romania remained near to Greece.

Population age structure is extremely important in light of its impact on political change. If the number of youths surpasses the other age groups in a country, it has serious implications on its political stability (youth bulge thesis)⁴. If the State fails to provide jobs, education and economic sustenance, it may face a protest movement inducing regime change. Richard Cincotta shows that countries tend to become more democratic when the median age increases. He argues that "the dissipation of a large youth bulge tends to yield relative political calm" Cincotta (2013). On the other hand, democratic gains under youth-bulge conditions "face unfavorable odds." Using age-structure data, he assesses the fragility of existing liberal democracies and forecasts when new ones will emerge. According to him, North Africa is at the forefront of democratization. This region is "pretty much in that category of relatively low fertility, between two or three children per women" notes Cincotta. "Egypt has come down but not nearly as much as some had hoped," which may explain its early reversion to authoritarianism." "Countries at the beginning of this age-structure transition have a very poor chance of A) making liberal democracy, and B) keeping it."

Cincotta studies revolutions between 1972 and 1989, and finds that oppressive autocracies with a median population age of between 25 and 35 had the best chances of becoming democracies. This idea is confirmed by our sample where countries can follow democratic ones when they achieve similar age structures. According to Cincotta, Tunisia,- with a median age of 30 - is the Arab Spring country most likely to hold a democracy permanently. Egypt and Libya have median ages of 25 and 26, respectively, giving them a fighting chance of moving to democracy

⁴ Gaan (2015) explains how youth bulge in Afghanistan can greatly affect its transition to liberal democracy in the context of American withdrawal from Afghanistan and failure of the government to address the socio-economic challenges the people face in society.

in the next few years, But Syria and Yemen - at 21 and 17, respectively - will be lucky to end up with even partial democracies.

Age is an important indicator but not a cause for stable transitions. Older populations tend to be associated with mature and then permanent democratic transition. Mature societies are associated to urbanization, higher income, women's rights and education. In addition, as the median age goes up all these factors reinforce each-other.

5.3 Component planes description

While maps provide general indications about the degree of closeness between countries, they don't allow us to deduce detailed information about democracy. To overcome this limit, component planes are used to gain deeper insight into the data.

Component planes can be created for each input variable and show the weights that connect each neuron with the particular input variable. Then, we consider democracy score in addition to other variables to make a Kohonen SOMs.

The component plane for the democracy score is shown in Figure 3. In this figure, the color gradation describes the countries' distribution from "very stable autocracy" (light shades) to "very stable full democratic" (dark shades). We can observe that the lower left corner of the map contains the countries considered to have the world's most closed political regimes. The most democratic countries are projected at the opposite side. They are all located in the dark red-colored region at the extreme right side of the map. It is important to note that some Latin American and African countries are also projected on the upper and lower right corner of the map indicating a high level of democracy, similarly to European and North American countries.

While, following Kaski and Kohonen (1996), no geographical data were entered into the model, it is noteworthy that countries organized themselves in a manner that mimics their relative geographical location. Geographical neighbors and adjacent country pairs are often located in the same region of the map. This is particularly true for Latin American countries located in the upper-right corner of the map, MENA countries in the lower-left corner and African countries which are concentrated in the 2nd diagonal of the map.

Component planes for other variables are shown in Figure 4. For each variable, there are four component planes in order to show its evolution.

The first two component planes provide information about linguistic and religious fractionalization. The dark spot displays countries with high fractionalization. It appears that there is an inverse correlation between "democracy" and "fractionalization."

The four following component planes are related to ICRG indicators, "government stability" and "law and order," and to "life expectancy" and "urban life." The dark colored spots indicate the area in which countries have high scores on institution and life quality.

The component plane of GDP per capita reveals the absence of impact of GDP per capita on countries' positions. This observation is in line with the ambiguous results of empirical works testing the modernization hypothesis. The same observation appears with internet users. Concerning the unemployment component plane, no impact of this variable is observed before 2002. It becomes important for all country positions in 2002.

The same kind of analysis can be performed for each of the other input variables. A high degree of correlation with the democracy component plane appears for military in politics, population 65+ or globalization. The coincidence between these component planes and the component plane of democracy indicates that in such countries democracy goes hand in hand with a military politician, old population and openness.

We showed that SOMs are a suitable tool for the exploration of data sets. If the SOM grid itself consists of numerous neurons, analysis can be facilitated by clustering similar neurons into groups (Vesanto and Alhoniemi, 2000). Then, we performed the traditional SVM clustering method on the trained map of 15 by 15 neurons. The result of this procedure is given in next section.

6. SVM Clustering Analysis

Several unsupervised labeling techniques have been proposed in the literature to cluster similar neurons and to identify those variables that characterize clusters (Brown et al. 2000, Zomer et al. 2004a). Support vector machines (SVMs) are able to create a nonlinear boundary for discrimination between two classes. SVMs have been applied to a wide variety of classification problems because of good discrimination ability (Freedom House, 2005). A small number of samples in training set which lie near the decision boundary (Support Vectors) are used to determine the classification (SVs) (Zomer et al. 2004b). If the decision boundary between two classes in the data space is not well defined by a linear function, then an appropriate kernel function can be used to transform the data into a higher dimensional feature space (Amari and Wu 1999). More details about the SVMs can be found in the literature (Zomer et al. 2004b).

In this section, we establish predictions on the probability of success of democratization. Models that are derived can be used for different goals. For autocratic countries, they can be applied to obtain an evaluation of the probability to engage a democratization process. For countries with failed or successful democratization, we can adopt these models to investigate which countries obtain results significantly different from the model predictions and study the reasons for these deviations.

6.1 Evaluation framework

We performed the SVM clustering method on the trained map of 15 by 15 neurons. The result is given in figure 5 for the global period and in figure 6 for 2013 observations. We have to choose the most performed model. In order to improve the accuracy of our analysis, several models are developed. Two important points for the selection of an accurate SVM model have to be fixed: the choice of the input variables, which are being considered in the score, and the tuning parameters (linear, sigmoid, RBF and linear SVM).

The choice of the input variables has a decisive influence on the performance results and is not independent from the choice of the classification technique. These variables normally have to comply with the assumptions of the applied classification technique. Since the SVM needs no restrictions on the quality of input variables, it is free to choose them only according to the model accuracy performance. Then, we consider three situations: considering all variables without democracy score, considering all variables, and considering only SOM. The three models for the different situations were trained on data over the time period 1984-1991-2002-2014.

The input variables selection methodology applied in this paper is based on the following empirical tools. In order to compare the performance of models, we need an evaluation framework that computes the models' capacity to predict democratic transition. As we have class information, we mainly use classification performance measures. We classify the outcomes into combinations of predicted and current democratization result using a contingency matrix where 0 means failure and 1 means success.

Based on the matrix, we compute ratios for measuring performance and polynomial algorithm. A very good BALANCED prediction performance of 83% appears⁵.

⁵ Performance ratio = (good prediction for 0 + good prediction for 1)/total prediction (e.g., 83% = (134+247)/457)

6.2 Results

Table 2 shows that the polynomial SVM model provides better forecasting accuracy than the other models. Figures 5 and 6 present polynomial SVM clustering for the global period and for 2013 respectively. Using a Kohonen pair, each country has its own weights according to the difference between the value of their own variables and those of the support vectors of the training data sample.

The estimation of score function and its validation are based on all available variables of democratic and non-democratic countries. A country is classified as non-democratic if it fails in the democratization process - its score is 0. A country is classified as democratic if it succeeds in its democratization process. The study is conducted over a global period and on each year. In each period, a country is regarded as non-democratic if it has been the subject of a democratization failure within one-year since the observation date. Democratic countries are those which have not gone toward authoritarian regimes within one-year after the observation date.

The SVM threshold can be represented within a bi-dimensional graph. A grey line on maps delimits the frontier of democratization success and failure. In the upper-right side of maps, we find countries which will succeed in their democratization. The lower-left part encompasses countries that will fail in their democratization process. Circles on maps indicate probabilities of success and failure.

Tunisia doesn't have a configuration allowing it to join the ranks of democracies before 2013 (figure 5). Tunisia is mapped near to Algeria, Turkey and Kyrgyzstan and could be a democratic country with a low probability. Democratization in Algeria should succeed with a high probability. This country is mapped with Bangladesh and Indonesia. It also appears that Egypt would be able to escape authoritarianism in 1984 but not after.

According to 2013 observations (figure 6), we can argue that the probability of becoming a democracy is one-half in Tunisia. It is very unlikely that Syria, Egypt, Libya, Yemen and Iraq succeed in the way of democratization. African countries may also fail, such as Ivory Coast, Mozambique, Mali, Niger, Senegal and South Africa.

In order to identify the future successes and to determine the most influent variables on democratization success, we use SVM. The polynomial algorithm allows us to identify "unemployment" as the most influent variable distinguishing failure from success followed by "corruption," "democratic accountability," and "law and order".

In fact, democracy overnight may be not the best way to transform authoritarian regimes. Democracy building, like market type reforms, should be gradual, rather than of a shock therapy type, and should go hand in hand with the strengthening of law and order. Democracy, participation in decision making and civil society are precious developmental goals in themselves and they should not be compromised by bad implementation. In his study of Arab political reform, Daniel Brumberg wrote: "It is far from clear how to reform liberalized autocracies... Encouraging rapid change, such as completely free elections, might invite radical forces and even a retreat to full autocracy." (Brumberg, 2003).

7. Conclusion

In this paper, a data mining approach for the analysis of democracy was presented. In the first part, the powerful visualization possibilities of self-organizing maps were used to study the interconnections between 33 variables and the democracy score. The use of multi-year data sets allowed us to visualize the democratic transition over time for 121 countries.

Our results show that there are economic, social, demographic, societal, political and institutional configurations conducive to the switch from autocracy toward democracy. These configurations are close to those existing during the first phase of the last Huntington democratization wave in Spain and in Latin American countries, which have become democracies in this period. Subsequent movements toward democracy in the other Latin American countries and in some African countries are associated with an evolution in their configuration, which consisted in a convergence with favorable configurations for democratic transition.

It appears that MENA countries are divided into two stable groups. In each group, there are close economic, social, demographic, societal, political and institutional configurations. The first group encompasses Yemen, Turkey, Tunisia, Algeria, Iran, Syria, Libya and Morocco while the second group includes Gulf countries, Iraq, Egypt and Jordan. During the last phase of democratization, Iraq, Yemen, Algeria and Tunisia became democratic countries. Among these countries, Yemen and Tunisia were part of the Arab Spring. However, they took different trajectories. Tunisia has succeeded in establishing a new democratic government, whereas Yemen is currently suffering State collapse. We argue that the success of the Tunisian democratic transition could be explained by a configuration close to democratic countries as Algeria, Turkey, Albania and Senegal.

The most important variables that the configuration of each country associated to a given democracy level are population age structure, globalization, health indicators, education and women's participation in the society.

In the second part, we use SVM clustering in order to identify the future successes and to determine the most influential variables on democratization success. It allowed us to identify unemployment as the most influential variable in distinguishing failure from success, alongside corruption, democratic accountability, and law and order. This analysis predicts success for Tunisia with a low probability, but a high probability of failure for Egypt, Iraq, Syria, Libya and Yemen. Tunisia is clustered with Kyrgyzstan, Algeria and Bangladesh. It also predicts failure for numerous African countries with unemployment and low institution quality, which may be regarded as impediments to maintaining democracy.

SVM presents a disadvantage, common to non-parametric techniques, of lack of transparency of results. SVMs cannot represent the score of all countries as a simple parametric function of all the available variables, since its dimension may be very high. It is neither a linear combination of a single variable nor does it have another simple functional form. The weights of each variable are not constant. Thus, the marginal contribution of each variable to the score is variable. We also considered short-term forecast horizons, which present the limits of data quality. Data would be poor, since most countries do not change regimes rapidly, if they are on the path towards democratization. Moreover, countries that move to democracies already show some difficult years before success.

References

- Acemoglu, D., (2006). "Modeling Inefficient Institutions", *NBER Working Paper*, 11940.
- Acemoglu, D., Robinson, J.A. (2006). *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.
- Acemoglu, D., Johnson, S. and Robinson, J. A. (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), pages 1369-1401.
- Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2005). "From Education to Democracy?" *American Economic Review*, 95(2), 44–49.
- Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). "Income and Democracy," *American Economic Review*, 98(3), 808–842.
- Amari S. and Wu, S. (1999). "Improving Support Vector Machine Classifiers by Modifying Kernel Functions," *Neural Networks*, 12(6), 783–789.
- Barro, R. J. (1999). "The Determinants of Democracy," *Journal of Political Economy*, 107, 158–183.
- Boix, C. (2001). "Democracy, Development, and the International System," *American Political Science Review*, 105(4), 809–828.
- Bourguignon, F., Verdier, T. (2000). "Oligarchy, Democracy, Inequality and Growth," *Journal of Development Economics*, 62, 285–313.
- Brown M. P. S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M. and Haussler, D. (2000). "Knowledge-Based Analysis of Microarray Gene Expression Data by Using Support Vector Machines", *Proceedings of the National Academy of Sciences*, 97(1), 262–267.
- Campante, F.R. and Chor, D. (2012). "Schooling, Political Participation, and the Economy," *Review of Economics and Statistics*, 94(4), 841-859.
- Castelló-Climent, A., (2008). "On the Distribution of Education and Democracy," *Journal of Development Economics*, 87(2), 179–190.
- Cincotta, R.P., (2013). *Environmental Change and Security Program Report*, popline.org.
- Deboeck, G. (1998). "Financial Applications of Self-Organizing Maps," *American Heuristics Electronic Newsletter*, January, 1-7.
- Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., Fischer, N., Schwarzwaelder, K., Siegler, M.L., de Ridder, D., Pike-Overzet, K., Howe, S.J., Thrasher, A.J., Wagemaker, G., Abel, U., Staal, F.J., Delabesse, E., Villeval, J.L., Aronow, B., Hue, C., Prinz, C., Wissler, M., Klanke, C., Weissenbach, J., Alexander, I., Fischer, A., von Kalle, C., Cavazzana-Calvo, M., (2007). "Vector Integration is Nonrandom and Clustered and Influences the Fate of Lymphopoiesis in SCID-X1 Gene Therapy," *Journal of Clinical Investigation*, 117(8), 2225-32,
- Du, K.L. (2010). "Clustering: a neural network approach," *Neural Networks*, 23, 89-107.
- Feng, Y. and Zak, P. J. (1999). "The Determinants of Democratic Transitions," *Journal of Conflict Resolution*, 43(2), 162-177.
- Freedom House (2005). *Freedom in the World Country Ratings*.

- Gaan, N. (2015). "Youth Bulge Constraining and Reshaping Transition to Liberal Democracy in Afghanistan", *India Quarterly: A Journal of International Affairs*, 71(1), 16-36.
- Gassebner, M., Lamla, M. J. and Vreeland J. R. (2013). "Extreme Bounds of Democracy," *Journal of Conflict Resolution*, 57, 2, 171-197.
- Hua, G., Skaletsky, M. and Westermann, K. (2009). "Exploratory Analysis of CIA Factbook Data Using Kohonen Self-Organizing Maps," *CS-BIGS*, 3, 48–59.
- Huntington, S. (1991). *The Third Wave: Democratization in the Late Twentieth Century*, Norman: University of Oklahoma Press.
- Jolliffe, I. (2005). "Principal Component Analysis," in Encyclopedia of Statistics in Behavioral Science, B. S. Everitt and D. Howell (eds), Wiley.
- Kaski, S. and Kohonen, T. (1996). "Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World," Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, England, Singapore: *World Scientific*, 498-507.
- Kaski, S., Kangas, J., and Kohonen, T. (1998). "Bibliography of Self-Organizing Map (SOM) Papers: 1981-1997," *Neural Computing Surveys*, 1, 102-350.
- Kohonen, T. (1982). "Self-Organized Formation of Topologically Correct Feature Maps," *Biological Cybernetics*, 43, 59-69.
- Kohonen, T. (2001). *Self-Organizing Maps*, Third extended edition, Springer.
- Lipset, S. M. (1959). "Some Social Requisites of Democracy: Economic Development and Political Legitimacy," *American Political Science Review*, 53, 69–105.
- Mangiameli, P., Chen, S. K. and West, D. (1996). "A Comparison of SOM Neural Network and Hierarchical Clustering Methods," *European Journal of Operational Research*, 93, 402-417.
- Murtin, F. and Wacziarg, R. (2013). "The Democratic Transition," *Journal of Economic Growth*, 19(2), 141-181.
- Pözlbauer, G. (2004). "Survey and comparison of quality measures for SOM," in Proceedings of the 5th "Workshop on data analysis", Paralic J., Pözlbauer G. and Rauber A. (eds), Elfa Academic Press, 67-82
- Ponthieux, S. and Cottrell, M. (2001a). "Neural Classification and "Traditional" Data Analysis: An Application to Households' Living Conditions," in Proceedings of the International Work-Conference on Artificial and Natural Neural Networks (IWANN), 738-745.
- Ponthieux, S. and Cottrell, M. (2001b). "Living Conditions: Classification of Households Using the Kohonen Algorithm," *European Journal of Economic and Social Systems*, 15, 69-84.
- Przeworski A., Alvarez, M.E., Cheibub, J. A. and Limongi, F. (2000). *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*, New York: Cambridge University Press.
- Tan P.N., Steinbach, M. and Kumar, V. (2006). *Introduction to Data Mining*, Pearson Addison Wesley
- Treisman, D. (2011). "Income, Democracy and the Cunning of Reason", *NBER Working Paper*, 17132.

- Urdal, H. (2006). "A Clash of Generations? Youth Bulges and Political Violence," *International Studies Quarterly*, 50, 607–629
- Vesanto, J. and Alhoniemi, E. (2000). "Clustering of the Self-Organizing Map," *IEEE Transactions on Neural Networks*, 11(3), 586–600.
- Zomer S., M. D. N. Sánchez, R. G. Brereton and Pavón, J. L. P. (2004a). "Active Learning Support Vector Machines for Optimal Sample Selection in Classification," *Journal of Chemometrics*, 18(6), 294–305.
- Zomer S., R.G. Brereton, J.F. Carter and Eckers, C. (2004b). "Support Vector Machines for the Discrimination of Analytical Chemical Data: Application to the Determination of Tablet Production by Pyrolysis-Gas Chromatography-Mass Spectrometry," *Analyst*, 129, 175–181.

Appendix

Table 1. List of Variables

Indicator	Code	Source
GDP per capita	GDPpc	
Trade openness	TRADEOP	
Mortality	Mortality	
Urban population	Urban	
Life expectancy	Life expec	World Bank
Population share 0-14	POP_0-14	
Population share 15-64	POP_15-64	
Population share 65+	POP_65+	
Internet users	INTERNET	International Telecommunications Union Dataset
Globalisation index	GLOB	KOF GLOBALIZATION INDEX
Educational Attainment for Total Population	EDUC	Barro and Lee data
Ethnic fractionalization (combined linguistic and racial)	frac_eth	
Linguistic fractionalization	frac_lan	Democracy Time-series Dataset
Religious fractionalization	frac_rel	
Protestant	Prot (dummy)	
Catholic	Cath (dummy)	Democracy Time-series Dataset
Muslim	Muslim (dummy)	
Orthodox	Orth (dummy)	
British colonial legacy	BRITCOL	Democracy Time-series Dataset
Monarchy	ROYAL	Cheibub and Gandhi (2009)
Electoral Self-Determination	ELECS	
Worker's Rights	WORKER	
Women's Economic Rights	WECON	CIRI database
Women's Political Rights	WOPOL	
Women's Social Rights	WOSOC	
Unemployment	UNEMPLOYMENT	
External conflict	EXTERNAL CONFLICT	
Military in politics	MILITARY	
Bureaucracy quality	BURQUAL	ICRG data
Corruption	CORR	
Democratic Accountability	DEMACC	
Government stability	GOVSTAB	
Law and Order	LAWORD	

Table 2: Performance Tests

		svm		
svm_rbf	0	1	total	performance
0	120	45	165	0,727272727
1	42	250	292	0,856164384
			457	0,809628009
svm_poly	0	1	total	performance
0	134	31	165	0,812121212
1	45	247	292	0,845890411
			457	0,833698031
svm_sigmoid	0	1	total	performance
0	66	99	165	0,4
1	18	274	292	0,938356164
			457	0,743982495
svm_Linear	0	1	Total	performance
0	118	47	165	0,715151515
1	40	252	292	0,863013699
			457	0,809628009

Figure 1: SOMs Based Upon 33 Variables and Non-Intervening Democracy Score

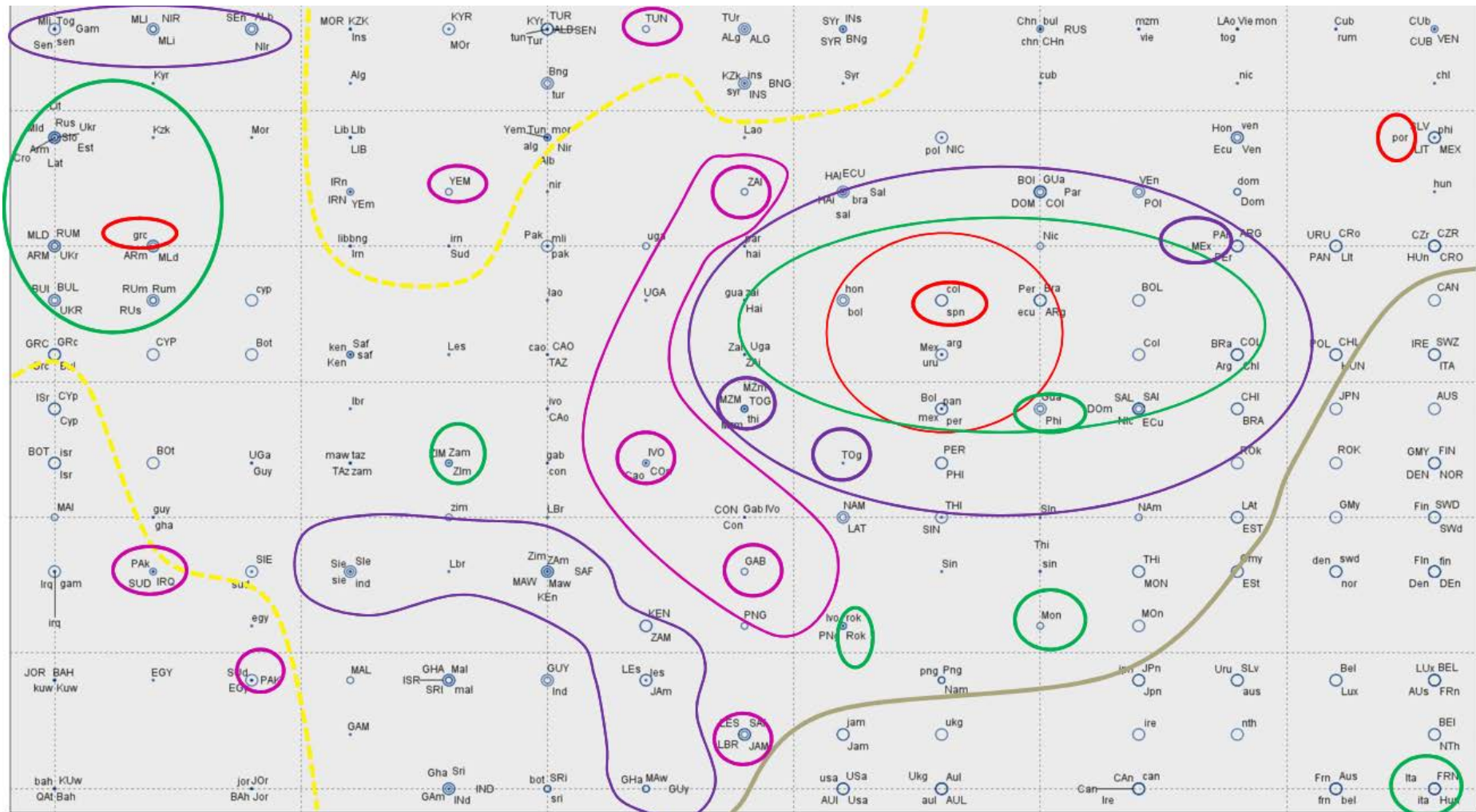


Figure 2: Yearly SOMs Based Upon 33 Variables and Non-Intervening Democracy Score

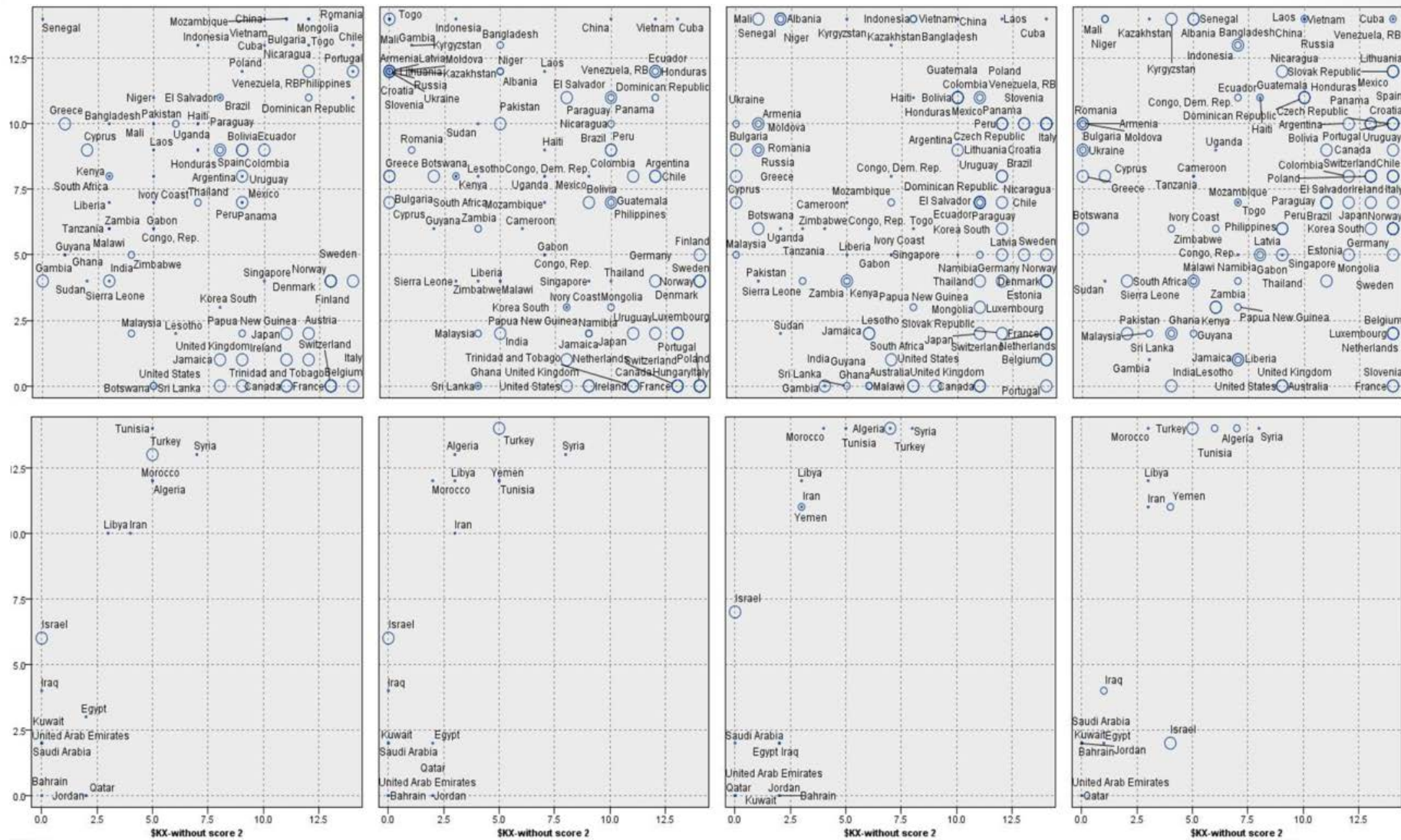


Figure 3: Democracy Score Component Plane



Figure 4: Component Planes

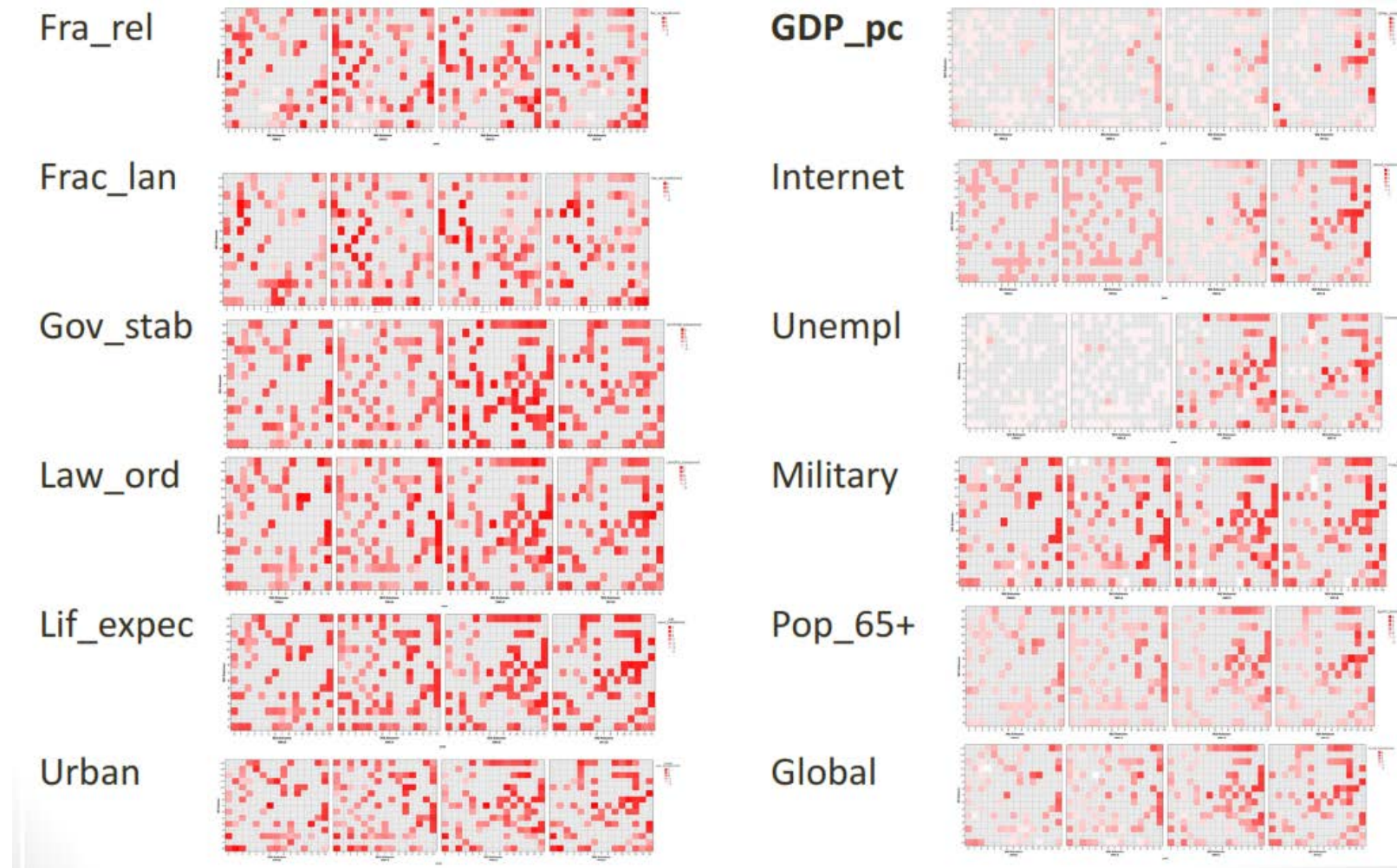


Figure 5: SVM on SOMs

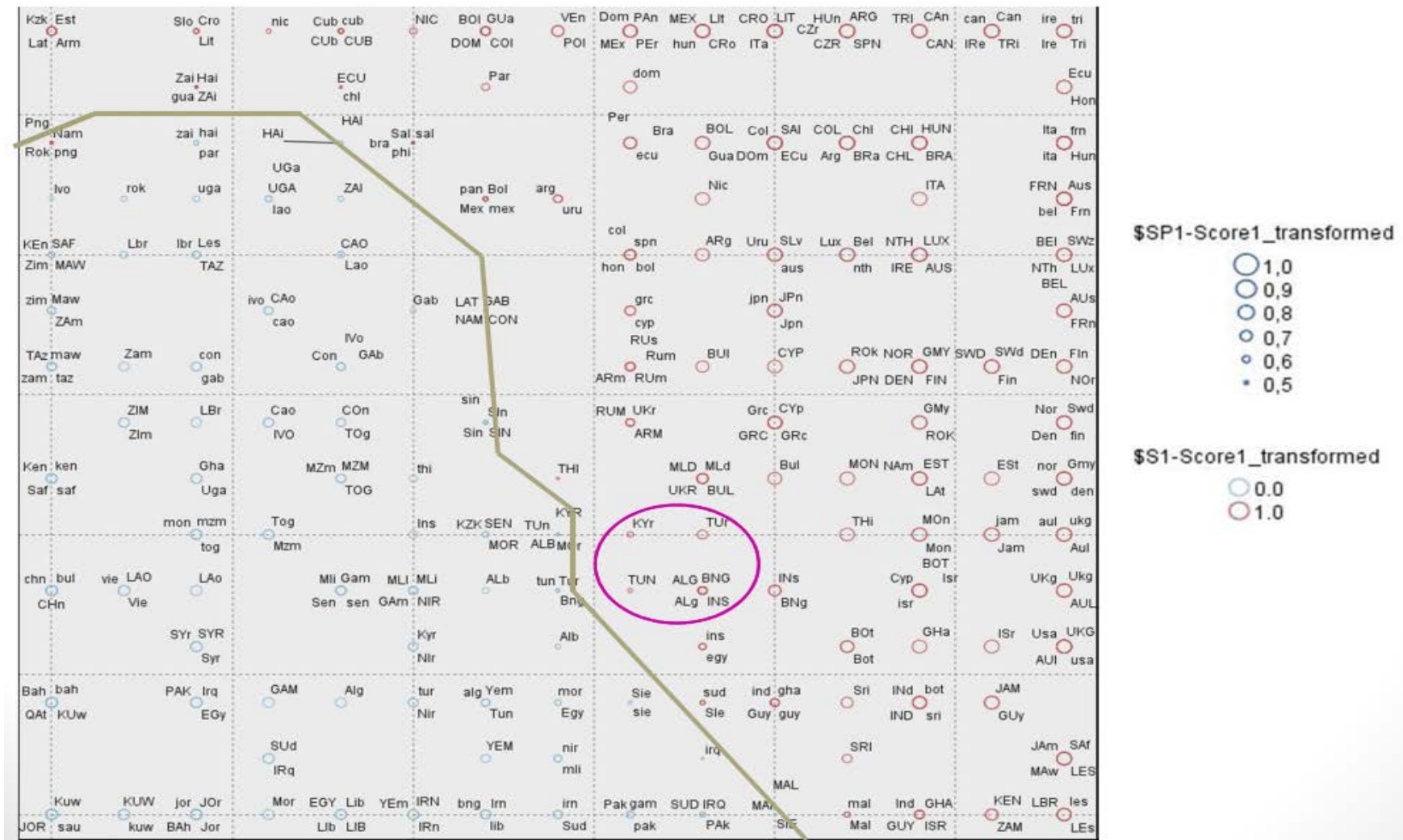


Figure 6: SVM on 2013 SOM

