



Investigating Sequences in Ordinal Data: A New Approach with Adapted Evolutionary Models

Patrik Lindenfors, Fredrik Jansson,
Yi-ting Wang and Staffan I. Lindberg

December 2015

Working Paper

SERIES 2015:18

THE VARIETIES OF DEMOCRACY INSTITUTE



UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a new approach to conceptualization and measurement of democracy. It is co-hosted by the University of Gothenburg and University of Notre Dame. With a V-Dem Institute at University of Gothenburg with almost ten staff, and a project team across the world with four Principal Investigators, fifteen Project Managers (PMs), 30+ Regional Managers, 170 Country Coordinators, Research Assistants, and 2,500 Country Experts, the V-Dem project is one of the largest ever social science research-oriented data collection programs.

Please address comments and/or queries for information to:

V-Dem Institute

Department of Political Science

University of Gothenburg

Sprängkullsgatan 19, PO Box 711

SE 40530 Gothenburg

Sweden

E-mail: contact@v-dem.net

V-Dem Working Papers are available in electronic format at www.v-dem.net.

Copyright © 2015 University of Gothenburg, V-Dem Institute. All rights reserved.

Investigating Sequences in Ordinal Data: A New Approach with Adapted Evolutionary Models*

Patrik Lindenfors
Assistant Professor
Stockholm University

Fredrik Jansson
Postdoctoral Research Fellow
Linköping University

Yi-ting Wang
Assistant Professor
National Cheng Kung University

Staffan I. Lindberg
Professor, of Political Science
Director, V-Dem Institute
University of Gothenburg

* This research project was supported by Riksbankens Jubileumsfond, Grant M13-0559:1, PI: Staffan I. Lindberg, V-Dem Institute, University of Gothenburg, Sweden; by Swedish Research Council, 2013.0166, PI: Staffan I. Lindberg, V-Dem Institute, University of Gothenburg, Sweden and Jan Teorell, Department of Political Science, Lund University, Sweden; by Knut and Alice Wallenberg Foundation to Wallenberg Academy Fellow Staffan I. Lindberg, V-Dem Institute, University of Gothenburg, Sweden; by University of Gothenburg, Grant E 2013/43.

Abstract

This paper presents a new approach for studying sequences across combinations of binary and ordinal variables. The approach involves three novel methodologies (frequency analysis, graphical mapping of changes between “events”, and dependency analysis), as well as an established adaptation based on Bayesian dynamical systems. The frequency analysis and graphical approach work by counting and mapping changes in two variables and then determining which variable, if any, more often has a higher value than the other during transitions. The general reasoning is that when transitioning from low values to high, if one variable commonly assumes higher values before the other, this variable is interpreted to be generally preceding the other while moving upwards. A similar reasoning is applied for decreasing variable values. These approaches assume that the two variables are correlated and change along a comparable scale. The dependency analysis investigates what values of one variable are prerequisites for values in another. We also include an established Bayesian approach that models changes from one event combination to another. We illustrate the proposed methodological bundle by analyzing changes driving electoral democracy using the new V-Dem dataset (Coppedge et al. 2015a, b). Our results indicate that changes in electoral democracy are preceded by changes in freedom of expression and access to alternative sources of information.

Introduction

One of the thorniest issues in social science analysis is to establish causal relationships for observational data when the options of using controlled or natural experiments are not available. We do not purport to have a solution to the general problem, but here suggest a set of novel analyses and adaptations of analyses that together can establish sequences between sets of ordinal variables, provided some key assumptions are met. One of the necessary, if not sufficient, criteria of causal relationships is that the causal factor X has to exist or change before the caused factor Y . However, there are many known exceptions to this general rule, e.g. the anticipation of an event such as elections may result in a series of effects before the event itself. Given this caveat regarding interpretation, the proposed approach focuses on identifying sequences of events in manifest observational data.

Sequences are also critical to understanding many social processes such as regime transitions, onset of wars, legislative procedures, international bargaining, and institutional development. A great deal of work in political science seeks to investigate the order of event occurrence or utilizes historical processes to explain political outcomes. However, the sequences of complex social processes usually involve hundreds of related variables with a large number of characteristics, and today's standard techniques for time-series, cross-sectional (TSCS) analysis for observational data, are not very apt for this sort of problems. First, they do not solve the causal inference problem to a greater or lesser degree than the approach suggested here. Second, they are designed for giving insights into the average effect of x_i on y given the conditions \mathfrak{z}_i , sometimes taking interaction effects into account. However, the social and political processes that we as social scientists are typically interested in, such as democratization, are rarely even approximations of such simplifications. Rather, complex and often long series of sequentially related variables are in play and contribute to the outcome.

We do not purport to be able to offer a complete and final solution to the disentanglement of such processes and explaining the outcomes, but offer a first step towards that goal. We suggest that the approach detailed below can establish descriptive sequences in terms of necessary conditions among, in principle, an unlimited number of variables over any stretch of time, given that adequate data is available and that there are, in fact, sequential relationships to be found. Thus, for example, the approach can establish that among, say, five variables x_1, \dots, x_5 and the outcome y_1 , in terms of either which of the variables “moves first” and/or in terms of which variable reaches a “high” or “full” level before others do, the sequence is $x_2 - x_4 - x_1 - x_5 - x_3$ and then y_1 . Naturally, this is not equivalent to establishing that this

sequence is a causal chain, or that the chain causes the outcome. However, to the extent that one can establish that this sequence across time and space always, or almost always, precedes the outcome, we have arguably come a long way in terms of arriving at a general understanding of and explanation to such a social process compared to where we are today. Until now, we have not been able to provide evidence of such sequences at all across time and a large number of units, other than by individual case analysis found for example in historical sociology and in-depth case-study approaches.

There exist a number of approaches to identify sequences in ordinal and categorical time series data, many more or less inspired by evolutionary biology. Noteworthy are e.g. social sequence analyses that are inspired by DNA sequence analyses (e.g. Abbott 1995; Abbot & Tsay 2000, Gauthier et al. 2010, Casper and Wilson 2015), qualitative comparative analysis (QCA) that is inspired by studies of evolutionary sequences (Ragin 1987; Rihoux & Ragin 2009), and time-series cross-section methods (Beck 2008). There also exists a more novel approach using Bayesian modelling to construe dynamic systems indicating flow of change (Ranganathan et al. 2014; Spaizer et al. 2014). All these methods have their pros and cons, where method choice is dependent on the format of the data and the specific question of interest.

One of the above-mentioned methods, the social sequence approach, has been utilized more frequently in political science studies. Scholars have adopted the technique to explore the careers of social movements activists (Fillieule & Blanchard 2012), voting behavior (Buton et al. 2012), crisis bargaining patterns (Casper & Wilson 2015), and the evolution of regime types (Wilson 2014). This approach identifies the temporal order of discrete events across observations, and uses an algorithm to compare and then cluster similar sequences. It mainly focuses on describing and exploring temporal developments between different states of single variables. When analyzing multiple variables, one has to combine the categories of different variables into a single variable (Gauthier et al. 2010). This approach is thus not suitable for data involving more than one variable measured on ordinal or continuous scales. For example, does the level of liberal rights protection increase before democratic quality improves? Does the strengthening of rule of law have to precede the increased competitiveness of multi-party elections for a successful democratic transition? Utilizing the conventional social sequence approach may ignore the ordinal nature of the vast majority of variables that we as social scientists depend on.

We here suggest a combination of novel and pre-existing methods suitable for investigating sequences of events between two ordinal variables that change along a similar scale, with a similar step size, and with each categorical score indicating similar magnitude. These methods include frequency counts, a graphical representation of observed changes that can be

compared to expected changes, dependency analyses inspired by the logic of Qualitative Comparative Analysis (Ragin 1987, Rihoux and Ragin 2009) as well as evolutionary biology (Sillén-Tullberg 1993), and Bayesian dynamical systems (Spaiser et al. 2014; Ranganathan et al. 2014). This combination of methods is especially useful when determining reform sequences or investigating if there are differences in reform paths between, for example, successful and unsuccessful attempts at, for example, democratization. By combining a number of two-variable sequence analyses, it becomes possible to describe reform sequences of a large number of variables.

What makes our suggested approach different from Granger tests for time-series data (Granger 1969), and the standard technique of lagging variables in TSCS analyses? There is one critical difference between Granger tests and the use of fixed time lags in TSCS and the combination of tools proposed here. None of the approaches presented below depend on specifying a specific time interval. The one, two, sometimes five year lags typically used in TSCS analyses are arbitrary and lack theoretical justifications. We do not know, for example, if improvements in civil liberties such as the freedom of discussion should be expected to be associated with improvements in, say, how clean elections are, one, five, ten or more years down the road. Furthermore, we have no theoretical reasons to (as standard techniques force us to) assume that the time lag between changes in x and changes in y is expected to be constant across countries and over time. On the contrary, in some areas we have empirically based intuitions suggesting that we should expect time-variant time lags. More generally, any statistical method that is based on an arbitrary choice of unit and constant choice of lags between events (one year? several years? months? days?) is fundamentally flawed.

Take suffrage as an example. This was one of the main contentious issues in the 19th and early 20th century. The improvements in freedom of discussion, organization, and the right to demonstrations led to protracted processes in many countries that eventually led to full suffrage. However, in the latter half of the 20th century, democratization processes starting with an opening up of the public sphere have usually (when successful) led to a rapid and immediate extension of full suffrage.

The methods proposed here do not require us to make any assumptions about either cross-sectional or time-invariant distances in time between the occurrences or changes in the status of variables. As explained further below, we reshape the observational data into “events” and focus on changes between these. An “event” simply means the combination of a fixed value i on variable x and a fixed value j on variable y for any period of time. This structuring of the data

makes it possible to map and analyze identical sequences of change, even if the length of time that country A and country B spend in various “events” varies across time.

Analyses based on the proposed approaches below can in principle be conducted for data measured at any level (interval, ordinal, binary) but in practice it requires ordinal or binary variables in order to be interpretable. The analysis is also easier to interpret if all variables in a particular analysis have the same level of measurement, but this is not required. Among the methods below, only the dynamical systems analyses require continuous data.

From the suggested frequency and dependency analyses, combining a series of such bivariate analyses (by running all variables against all), one can establish long series of sequences involving hundreds of variables. The result is a detailed and empirically based “map” of which aspects of a phenomenon occur before other aspects. In other words, we are now capable of providing the first solution to presenting detailed sequences of democratization and other similar phenomena.

With such a set of tools at hand, we can as social scientists for the first time potentially provide answers to questions like: If we want to support a democratic development in Egypt today, where should we start? Should the efforts be directed at strengthening rule of law, or rather focus on supporting a freer media landscape, or perhaps the development of more institutionalized political parties? These are questions to which we have until now been unable to provide much of an informed answer. The combination of the analyses below promises us to be in a much better position to answer such questions with a strong empirical foundation.

Methods

The analysis of sequences is a tricky issue and the approach we have developed builds on combining several related methods that in combination provide a basis for such assessments.

1. Frequency Analysis

To explore the temporal relationship between two ordinal variables we first investigate whether one of them in general tends to be larger than the other (indicating that its values are ahead of the other). We therefore construct a frequency table including all possible combinations of the values of these two variables. An example of such a frequency table for two sample variables, Variable A and Variable B, is shown in Table 1. To calculate the frequency of each combination, for each country, we first combine several yearly observations when the values of both variables do not

change into one observation and count them as one “event”, regardless of how many years that combination is stable. We then count the occurrences of each combination.

Table 1.1 Example of a frequency table of observed combinations of values of two variables

		<i>Variable A</i>			
		<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>
<i>Variable B</i>	<i>0</i>	4	3	1	0
	<i>1</i>	0	3	5	1
	<i>2</i>	0	0	3	4
	<i>3</i>	0	0	0	3

We then calculate the percentage of observations where one variable is greater than the other and compare this observation to the number of observations where the other variable is greater (and where they are both of the same magnitude). For example, in Table 1, Variable A is higher than Variable B in 14 cases, both are equal in 13 cases, while Variable B is never larger than Variable A. We can then do the same for Variable A and Variable C, for example:

Table 1.2 Another combination of variables

		<i>Variable A</i>			
		<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>
<i>Variable C</i>	<i>0</i>	0	0	0	0
	<i>1</i>	3	3	0	0
	<i>2</i>	2	4	1	0
	<i>3</i>	5	4	2	3

Finally, we can (with any number of variables) construct a relative frequency table to systematize the number of combinations.

Table 1.3 Example of relative frequencies table

x= B or C	when A > x	when A = x	when A < x
B	52%	48%	0%
C	0%	26%	74%

Provided that both variables vary along the same scale, in the transitions, Variable A is thus always larger than or equal to Variable B, in absolute terms. At the same time, Variable C is always larger than or equal to Variable A. Assuming that we have a more or less complete dataset on these variables (covering all relevant units and time stretch) and the relationship also holds true for the comparison between Variable B and Variable C, this would constitute strong descriptive evidence that there exists a sequence Variable C => Variable A => Variable B.

2. Graphical Investigation of Changes

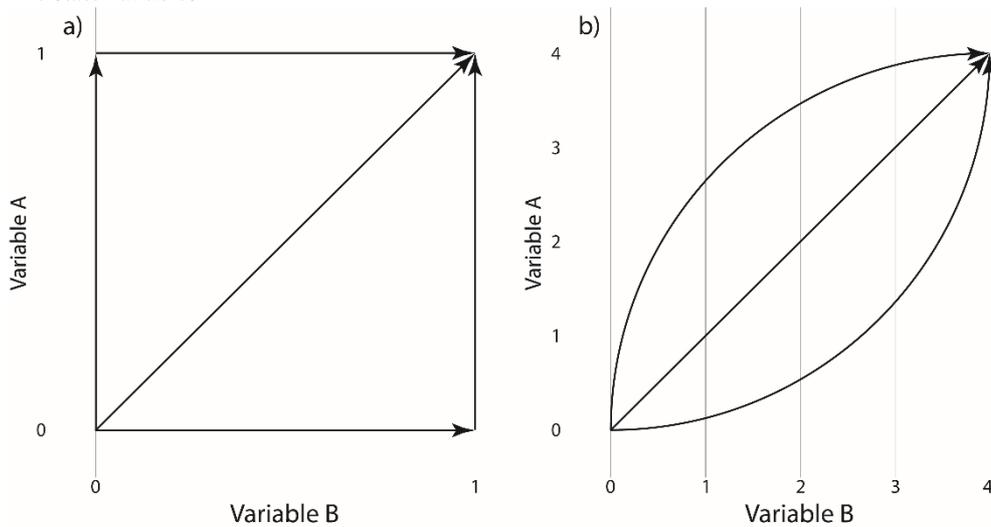
Frequency analyses allow us to explore how often one variable is larger than another, but they do not really clarify how this comes to be. To further investigate the exact pathways for how variables change, we use a graphical approach as described below. Figure 1 shows the main idea behind this approach.

For two binary variables, both of which take the values as 0 and 1, there are three ways for two variables to change from (0, 0) to (1, 1). Either Variable A can change first to (1, 0), both can change simultaneously to (1, 1), or Variable B can change first to (0, 1). Thus, to determine how often Variable A becomes 1 before Variable B, one simply counts the number of changes from (0, 0) to (1, 1) that go via (1, 0) and compare this number to the number of changes from (0, 0) to (1, 1) that are direct, or via (0, 1) (Figure 1a). This counting results in a frequency table similar to Table 1, but for binary characters with only four possible combinations.

For ordinal variables the situation is more complex. Consider two comparable ordinal variables (Variable A and Variable B) varying along the same scale with the same step size – in our example all variables can take on the integer values 0, 1, 2, 3 and 4. The question of interest is again if one variable tends to be larger (or smaller) than the other after a change. Partly, the general thinking is similar – if Variable A tends to be larger than Variable B then changes ending above the diagonal will be more common than changes ending below the diagonal (Figure 1b), enabling the similar frequency analysis as described above.

Partly, however, the situation is entirely different, because a multitude of other potential paths are possible. For example, one variable may become larger first at lower values while the other becomes larger first at higher values; variables may go in both directions; or variables may be unrelated but one variable has a skewed distribution, a pattern that would result in more changes ending on one side of the diagonal even if no correlation exists between the two variables. We may not observe processes from the beginning, but in the middle of transitions, which would mean that we miss parts of the process for each country. If, for example, A tends to be larger than B in the beginning of a transition, and the opposite in the end, then we would underestimate the occurrences of $A > B$ if we lack more beginnings than ends of transitions. Further, and most importantly, the variables may not be comparable – how does the value 2 for Freedom of expression truly compare to a value 2 of Alternative source information? Thus, as a second type of analysis, it is important to visually inspect the movements using a graphical method – because of this, we term the approach graphical rather than statistical.

Figure 1: Potential pathways of change in (a) two binary variables and (b) two comparable ordinal five state variables.



To create plots indicating temporal changes in two variables, we first construct a table listing all observed changes in values of the two variables, then produce a figure mapping all these changes (see the Results section). We report two types of figures. In the first, we map all changes using arrows indicating movement between states, where the thickness of each arrow is proportional to the number of changes that have occurred along that particular path. We also add circles in the graphs to indicate the number of times that particular combination of states is observed in the data, i.e. the size of each circle is proportional to the number of observations of that particular combination. The latter are the same as the numbers reported in the frequency tables discussed above.

The other type of figure reveals reform paths that are more popular than expected by utilizing observed data (indicated by the circles in the first type of figures) to calculate a table of expected values using chi-square methodology from the distribution of each variable. Graphing the difference between the table of observed and the table of expected values reveals reform paths that are more popular than would be expected by the distribution of the two variables alone (see the Results section).

If one wants to draw conclusions about sequences from these graphs, the proposed approach assumes that variables change along a similar scale, with a similar step size, with each categorical score indicating a similar magnitude, and that no parts of the transition processes are systematically underestimated. It may be tempting to use observed and expected tables discussed in the previous paragraph to also calculate the chi-square statistic, but since the two tables may differ for reasons other than the existence of popular reform paths, a significance value is not necessarily meaningful – hence the emphasis here on visual inspection. If significance tests are

deemed desirable (we urge caution since this requires that many assumptions are fulfilled), we instead recommend quasi-symmetric model tests for comparing off-diagonal values in square tables (Agresti & Kateri 2002) utilized on the observed frequencies.

3. Dependency Analysis

To explore whether certain values of one variable are systematically conditional on certain values of other variables in the existing data, we have developed a method here termed dependency analysis. The method is inspired primarily by “the contingent states test,” which is an established method developed to investigate dependencies in biological evolution (Sillén-Tullberg 1993), and is particularly well suited to use on sequence data outside biology. It also has similarities with qualitative comparative analysis (QCA) (Ragin 1987; Rihoux & Ragin 2009) – in fact, in some sense it is a stripped down, repeated version of this method. Note that even though Variable A and Variable B may co-vary, the proposed method checks for absolute dependencies in the data, not statistical correlations. This is an important distinction since if and when one can establish such absolute dependencies – and again assuming that the data is more or less complete in coverage – this is evidence of revealing necessary conditions displayed by the data as such that are not contingent on analytical inferences from regression statistics.

To conduct this type of dependency analyses, for each value of one variable, we scan the dataset for the lowest value in the other variables. If higher values in Variable A always correspond to higher “lowest values” in Variable B, then it can be inferred that certain values of Variable A are likely to be conditional on certain values of Variable B. If, simultaneously, for each value of Variable B, the corresponding “lowest value” in Variable A is its minimum, then this shows that Variable B is not restricted by Variable A. These two observations in combination indicate that potential dependencies between the two variables exist only in one direction. Dependency should not be taken here as a causal relation between A and B, only that certain values for one variable are conditional on certain values for the other in the available observations. To allow some margin of errors, a percentile of observations can be specified and treated as the “lowest values,” which will slightly relax the criterion of absolute dependencies. We here report both absolute dependencies and dependencies allowing a 95% “wiggle room,” following the convention in QCA.

Table 2 shows an example of such a procedure. The left table (a) indicates that higher states (2 and 3) in Variable A occur *only* together with higher values in Variable B (2 and 3, respectively). This means, for example, that Variable A is *never* observed to reach value 1, before Variable B has reached value 2. Thus, Variable B must *necessarily* reach value 2, before Variable A

can “start moving”. At least, this is how it has *always been* so far according to the data. The right table (b) indicates no such dependency since Variable A can be 0 at any level of Variable B. Thus, variable A is likely dependent on changes of Variable B having taken place at several stages, while in the opposite direction there is no such dependency. In this case, we would conclude that improvements in Variable B are a *necessary* condition for improvements in Variable A (without implying that there is a direct causal relationship).

Table 2: Example of dependency tables

(a)	Lowest value of		(b)	Lowest value of	
	Variable A	Variable B		Variable B	Variable A
	0	0		0	0
	1	2		1	0
	2	2		2	0
	3	3		3	0

For an analysis of sequential relationships between a larger number of variables, dependency tables can be constructed for all possible combinations of variables, and be summarized in various ways. There are various possible summation measures, each providing different information depending on which parts in the transition process to focus on. We here present three possible measures, depending on whether the interest is on early, late or continuous dependencies in the transition process.

One measure is to sum all the lowest values of Variable B, giving a sum of 7. High values on this summary indicator suggests that for Variable A to improve and reach higher states, high(er) values on Variable B are required early on in the transition process.

Another measure is to sum the number of increments in B. There are two thresholds where B needs to increase for A to increase: B needs to be at least 2 for A to grow beyond 0 and B needs to be at least 3 for A to achieve its maximum value. This measure then indicates when there is a strong correlations and gradual dependence of A on B throughout the transition process.

A third measure is to look at the lowest value in Variable B when A attains its maximum value, here 3. In this example thus, A *never* reaches it highest state without B being at its highest value 3. The reverse is not true: B can reach its highest value 3 and A being zero. Thus one can conclude that the full development of B *always* precedes or comes in tandem with A. This number is informative of the end of the process, and of interest when investigating what is required for full implementation of an institution

A low number of dependencies for all states of Variable B indicates that there are very few necessary conditions for it to assume higher (in our case, more democratic) states. If Variable

A has a high number of contingencies, then this indicates that it never reaches higher (more democratic) states before a number of other variables have reached high levels.

Thus, for the first time we can get a good sense of which variables come first, middle, and last in processes of democratization. In the example above, for example, the sum of dependencies for Variable A is 7 (and the other two measures are 2 and 3), while the sum of dependencies for Variable B is 0 (as for the other two measures), on the whole. This indicates that Variable A needs particular states in Variable B more than Variable B needs particular states in Variable A, at each stage of the process.

An example of how several such bivariate dependency tables can be summarized is found below in Table 3. In this illustration, we use our the third measure: looking at the threshold value for each of the other variables for a focal variable to attain its highest value. For each focal variable, we summarize these threshold values and report them as “#Necessary conditions”.

Table 3. Example of combining dependency by reporting, for each variable, the number of conditions (sum of the lowest values for all other variables) required to reach highest state.

Variable	# Necessary conditions	% of max
B	0	0%
E	4	20%
D	5	25%
C	6	30%
F	14	70%
A	16	80%

In this illustration, the maximum sum of thresholds, or necessary conditions for a variable reaching its highest state, is 20 (five other variables, and each variable’s maximum level is four, for the highest state). The illustrative results would indicate that Variable B comes first in attaining its maximum value in a sequence. It can reach its highest state completely unconditional on any other variables. Variables E, D, and C constitute a middle group with some conditions required for them to reach their highest states. The low number of dependencies indicates that the variables upon which their highest states are conditional, are relatively low states on other variables. (In order to know which variables these are, one would then have to go back to look at the summary table for each of these three variables.) Finally, variables F and A are the “late-comers” that have only been observed at their highest states *after* a greater number of other variables reaching their highest, or close to highest, states. Together, this indicates a rough sequence that can be instructive for analysis of direct policy relevance.

This example of looking at the highest states dependencies, is of particular interest when one is analyzing for example what these conditional relationships look like for achieving democratization, understood as becoming fully democratic. Then it is natural to focus on the

highest states of variables. If one were, for example, interested rather in the onset of transitions, one should probably look at the number of dependencies for different variables reaching the first, or perhaps the second level on the ordinal scales, which would indicate “early moves” rather than “final push”.

4. Bayesian Nonlinear Dynamical Systems

In order to study nonlinear dynamics in the interaction between variables, we also suggest employing a newly developed Bayesian dynamical systems approach that models the probable reform direction of countries depending on state combinations. This method identifies the best nonlinear functions that capture the interactions between two or more variables. Bayes factors are employed to decide how many interaction terms should be included in the model, with a punishment for overly complex models. The method gives a pair of differential equations, modeling how the values in each of the two variables involved affect the direction of each. From this, we can infer which is the most likely trajectory a country will follow, given any starting point. The resulting dynamical system can be illustrated by a phase portrait, where the modeled trajectories are depicted with arrows. The method is described in detail in two papers by Spaiser et al. (2014) and Ranganathan et al. (2014b).

Since differential equations deal with continuous variables, we use continuous versions of the V-Dem variables. It is important to note that the method provides a system for the entire set of possible values for the two variables. That is, it uses all the data points and provides a general description for the entire system, including combinations of the two variables that do not occur in the data. For illustrative purposes, we do not plot the arrows for these points in the phase portraits.

Application: The Development of Difference Aspects of Democratization

Scholars have pointed out various defining features of democracy, such as the protection of civil rights, accountability to citizens' preferences, regular and competitive elections for the chief executive and national legislature, freedom of expression and association, and constraints on executive's behavior. In the post-Cold War era, we have seen regimes with different mixes of the democratic and authoritarian features. For example, in some countries, citizens enjoy liberal rights to a certain extent, but fail to establish competitive elections, while in others, elections take place regularly, but media is largely controlled by the incumbent party. Do some of the democratic features constitute the preconditions of other aspects and have to develop first for a successful democratic transition? What are those features? Is the improvement of some aspects more likely to lead to the enhancement of others? The sequence analysis methods discussed above are well suited to the task of examining the temporal relationships between different aspects of democracy in the process of regime transition. In what follows, we use the example of different components of "polyarchy" to demonstrate how the proposed approach can be utilized for such a task.

Based on Dahl's (1971) conceptualization, polyarchy refers to a political system in which rulers are responsive to the preferences of its citizens. To achieve this, the minimum requirements for polyarchy include: 1) freedom to form and join organizations, 2) freedom of expression, 3) the right to vote, 4) eligibility for public office, 5) the right to compete for support, 6) alternative sources of information, 7) free and fair elections, and 8) institutions for making government policies depend on votes and other expressions of preference. Among these requirements, items 1, 3, 4, 5, 7, and 8 focus on the mechanism of competitive elections in which all citizens should have equal opportunities to participate. These features construct citizens' ability to make a choice that reflects their preferences in an explicit political decision making process. The other two items, freedom of expression and alternative sources of information, concern citizens' ability to formulate and express their political opinions. That is, these requirements enable citizens to define their goals in the public sphere.

We expect that in the process of regime change, the freedom of expression and access to alternative sources of information should develop before the establishment of competitive elections. In the literature on democratization, scholars have pointed out that the challenge to the authoritarianism results from a preference for changing the redistributive equilibrium through democratization (Acemoglu and Robinson 2001, Boix 2003). For preferences about the

redistribution issue to be formulated and for the idea of a democratic system as an alternative solution to be recognized and accepted by more people, citizens have to enjoy at least a certain level of freedom to discuss their government and have access to viewpoints alternative to government propaganda. That is, as different aspects of democratic governance, we expect that the improvement of citizens' expressive freedom precede the establishment of a substantially free and fair electoral regime.

1. Data

To explore the temporal relationship between various aspects of democracy utilizing the proposed sequence analysis approach, we use the V-Dem dataset v4 for the purposes of presenting. V-Dem aims to achieve transparency, precision, and realistic estimates of uncertainty with respect to each data point. The v4 dataset includes 173 sovereign or semi-sovereign states from 1900 to 2012, and data covering 2013–2014 for 60 of these countries.¹

The indicators in the “V-Dem Codebook” fall into three main categories: 1) factual data gathered from other datasets or original sources; 2) evaluative indicators coded by multiple country experts; and 3) aggregated indices constructed by combining several indicators that load on the same dimension based on factor analysis results. The evaluative indicators are produced according to a complex and demanding protocol. Typically, five or more independent country experts code each country-year for each indicator and almost three thousand experts have been involved in the coding to date.² To arrive at the best possible estimates, V-Dem has a team of measurement experts and methodologists who have developed an advanced Bayesian ordinal IRT-model for aggregating and weighting expert ratings and for calculating confidence intervals alongside a series of validity and reliability tests, including tests of intercoder reliability (see Pemstein et al. 2015 and Coppedge et al. 2015c). This model takes into account the possibilities that experts may make mistakes and have different scales in mind when providing judgments.

All V-Dem aggregated indices are interval variables scaled from zero to one. As a first step then, we have transformed these to a set of ordinal variables. The detailed justifications and

¹ Version 5 of the V-Dem dataset is released in January 5, 2016. But for the purposes of this paper laying of the methods as such, it is inconsequential which version is used. A detailed explanation of the V-Dem approach can be found on V-Dem's website, <https://v-dem.net>, along with the other V-Dem documents cited in this paper.

² The coders' considerable knowledge derives from a combination of experience and education: Most have lived in their countries of expertise for nearly thirty years, and 60 percent are nationals of that country. In addition, 90 percent have postgraduate degrees. Ratings accorded to a country are therefore largely the product of in-country expert judgments. In addition to providing a rating on each indicator, country experts also assign a “confidence score” (0 to 100), which measures how certain we can be about the rating. In addition, roughly a fifth of the coders undertake cross-country coding, making it possible for us to calibrate measurements between countries.

the STATA code for these transformations are found in Lindberg (2015) and we restrict ourselves to describe the general logic here. First of all, since the V-Dem indices are the normalized probabilities, we can assume that, as a general rule, equidistant thresholds for a categorical version make sense. At the same time, for the roughest categorization with only three levels, we assume that the breakpoint between being closer to the endpoint (0.5) is critical. Thus for categorization, the transformations of the interval indices to ordinal variables start with the 0.5 threshold and then further subdivide the lower category at 0.25. Visual face validity checks comparing the results of such an approach with the original interval values have also corroborated this intuition (based on inspecting some 10,000 graphs). Thus for most indices, the following generic rule is followed for the transformation into three, and four-level versions of the ordinal indices, where I denotes “index”:

3 CATEGORIES

0.0: $0 \leq I \leq 0.25$
 0.5: $0.25 < I \leq 0.5$
 1.0: $0.5 < I \leq 1$

4 CATEGORIES

0.00: $0 \leq I \leq 0.25$
 0.33: $0.25 < I \leq 0.5$
 0.67: $0.5 < I \leq 0.75$
 1.00: $0.75 < I \leq 1$

In principle, the methods that we have developed and are described in this paper, can be applied to any level of measurement. For the purposes of illustrations below, however, we have chosen the ordinal versions of the indices with five levels. For this ordinal version, Lindberg (2015) have chosen to divide the 0–1 interval scale by equidistant thresholds, thus:

5 CATEGORIES

0.00: $0 \leq I \leq 0.2$
 0.25: $0.2 < I \leq 0.4$
 0.50: $0.4 < I \leq 0.6$
 0.75: $0.6 < I \leq 0.8$
 1.00: $0.8 < I \leq 1$

The resulting ordinal indices typically correlate with the original interval versions at .94 or higher. To construct an ordinal version of the electoral democracy index ($v2x_polyarchy$), Lindberg’s (2015) analyses provide evidence that the categorization of this index should be especially conditional on the values of two critical variables: to what extent elections were multiparty ($2velmulpar$) and to what extent the elections were overall free and fair ($v2elfrfair$). Generally, it was face validity analyses that provided the empirical basis for this conclusion. With the “cruder” categorizations leading to only three or four levels, he found that countries just slightly above a threshold could be of very varying quality. This is the result of the aggregation rule for the original V-Dem indices that are in part multiplicative and in part additive. In the end, the following “ordinalization rules” provide high correspondence between the original values of

the indices, and high face validity (again based on inspection of thousands of graphs, each covering some 100 years of political history for a particular country, last condition in the list applies):

3 CATEGORIES

0.0: $0 \leq v2x_EDcomp_thick \leq 0.25$

0.0: $0.25 < v2x_EDcomp_thick \leq 0.5 \ \& \ 0 \leq v2elmulpar_dos \leq 2.5$

0.0: $0.25 < v2x_EDcomp_thick \leq 0.5 \ \& \ 0 \leq v2elrfair_dos \leq 2$

0.5: $0.25 < v2x_EDcomp_thick \leq 0.5 \ \& \ 2.5 < v2elmulpar_dos \leq 4$

0.5: $0.25 < v2x_EDcomp_thick \leq 0.5 \ \& \ 2 \leq v2elrfair_dos \leq 4$

0.5: $0.5 < v2x_EDcomp_thick \leq 1 \ \& \ 0 \leq v2elrfair_dos < 3$

1.0: $0.5 < v2x_EDcomp_thick \leq 1 \ \& \ 3 \leq v2elrfair_dos \leq 4$

4 CATEGORIES

0.00: $0 \leq v2x_polyarchy \leq 0.25$

0.00: $0.25 < v2x_polyarchy \leq 0.5 \ \& \ 0 \leq v2elmulpar_osp < 2$

0.33: $0.25 < v2x_polyarchy \leq 0.5 \ \& \ 2 \leq v2elmulpar_osp \leq 4$

0.67: $0.5 < v2x_polyarchy \leq 1 \ \& \ 2 \leq v2elrfair_osp \leq 4$

1.00: $0.5 < v2x_polyarchy \leq 1 \ \& \ 3 \leq v2elrfair_osp \leq 4$

The same logic and conditions then apply to transformation to ordinal versions of the electoral component index ($v2x_EDcomp_thick$) and the clean elections index ($v2x_el_refair$). The correlations with the original interval indices are very high, typically above .96. Further details and justifications of the creation of the ordinal versions of the indices are found in Lindberg (2015).

2. Measures

To measure citizens' freedom to discuss political issues and formulate opinions, we rely on the V-Dem "Freedom of expression" index, which combines the freedom of discussion and academic expression, no print, broadcast, and internet censorship, no media self-censorship, and no harassment of journalists. Another democratic feature that we are interested in is whether citizens have the access to alternative sources of information. We rely on the V-Dem "Alternative source information" index, which measures the extent to which there is a media bias against the opposition, whether media criticize the government, and whether media represent a wide range of political perspectives. The construction of the ordinal version of these two indices is based on the procedure described above.

To capture the quality of competitive elections, we rely on the V-Dem Electoral component index. Consistent with the features of polyarchy that focus on the electoral mechanism, this index combines the following elements: whether suffrage is extensive; political and civil society organizations can operate freely; elections are clean and not marred by fraud; and the chief executive is selected through elections. Based on the categorical version of the V-Dem indices, we could then proceed with the following analyses that together constitute the new approach to sequence analysis that this paper introduces.

3. Illustrative Results

Spearman rank correlations revealed variables to be positively correlated with each other. What we aim to investigate is how movement happens between high and low values in polyarchy – are there sequences of change between the different component variables? Specifically, does the improvement of citizens’ expressive freedom and information sources precede the establishment of substantially competitive elections?

1. Frequency Analysis

As can be calculated from the frequency tables (Table 1a-c), Freedom of expression ends up being higher than the Electoral component index in 57.1 % of the cases any variable ends up higher than the other (862/1509), while the two variables end up equal in 30.8 % of all cases (464/1509), and Freedom of expression ends up lower in 12.1 % of the cases (183/1509) (Table 1a; Fig. 2). Similarly, Alternative source information ends up higher than the Electoral component index in 61.8 % of the cases any variable ends up higher than the other (920/1488), while they end up equal in 27.9 % of all cases (415/1488), and Alternative source information ends up lower in 10.3 % of the cases (153/1488) (Table 1b, Fig. 3). Finally, Freedom of expression ends up higher than Alternative source information 20.2 % of the cases any variable is higher than the other (270/1334), while they end up equal in 45.6 % of all cases (609/1334), and Freedom of expression ends up lower in 34.1 % of the cases (455/1334) (Table 1c, Fig. 4). Note that these observations are the same as asking which variable most often has a nominal value larger than the other variable, with the added information of knowing (through visual inspection of the graphs) how this came to be so. From these results and the graphs below, it can be argued that the Electoral component index lags or changes at the same time as both Freedom of expression and Alternative source information, while the two latter variables most often change together. Freedom of expression and Alternative source information thus take on high values before the Electoral component index.

Table 3. Frequency tables of the end results of changes in combinations of the traits Freedom of expression, Alternative source information and Electoral component index.

a)		Electoral component index				
		0	1	2	3	4
Freedom of expression	4	4	34	47	85	57
	3	23	92	150	81	20
	2	116	177	94	24	3
	1	134	148	38	4	0
	0	84	84	10	0	0

b)		Electoral component index				
		0	1	2	3	4
Alternative source information	4	17	46	70	105	54
	3	52	134	130	80	14
	2	100	151	76	18	0
	1	115	128	36	5	0
	0	77	71	9	0	0

c)		Alternative source information				
		0	1	2	3	4
Freedom of expression	4	0	1	6	49	102
	3	0	12	61	153	110
	2	10	70	132	118	23
	1	61	150	93	32	2
	0	72	67	10	0	0

2. Graphical Investigation of Changes

In figures 2–4, the left panel indicates empirical observations of all occurrences and changes in the data, while the right panel depicts the difference between observed occurrences and the expected number of occurrences given the distribution of the variables. In the left hand panels, the size of each arrow and circle is proportional to the number of observations. In the right hand panels the legend indicates the difference between observed and expected frequencies, indicating a ‘preferred reform path’ as designated by the yellow and gray paths.

Figure 2. Graphs showing (left) the frequencies of occurrences and changes and (right) the observed - expected frequencies in the Electoral component index and Freedom of expression. As can be seen, there are more occurrences where the value of Freedom of expression is scored larger than or equal to the value of the Electoral component index. The right figure indicates a ‘preferred reform path’ where Freedom of expression mainly remains larger than the Electoral component index. From these results it can be inferred that the Electoral component index lags or changes simultaneously as Freedom of expression.

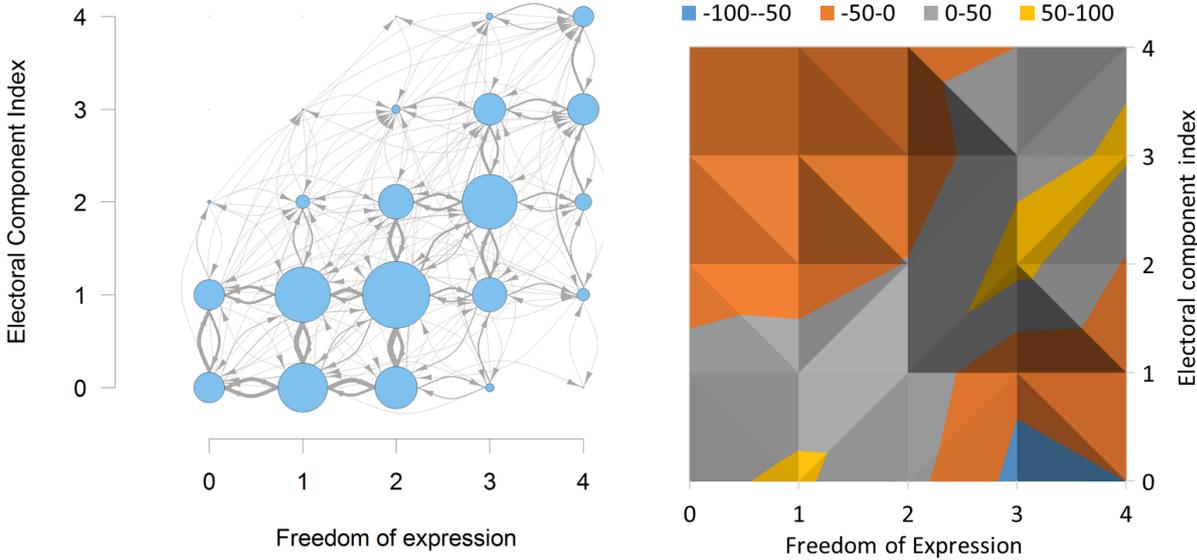


Figure 3. Graphs showing (left) the frequencies of occurrences and changes and (right) the observed - expected frequencies in the Electoral component index and Alternative source information. As can be seen, there are more occurrences where the value of Alternative source information is scored larger than or equal to the value of the Electoral component index. The right figure indicates a ‘preferred reform path’ where Alternative source information mainly remains larger than the Electoral component index. From these results it can be inferred that the Electoral component index lags or changes simultaneously as Alternative source information.

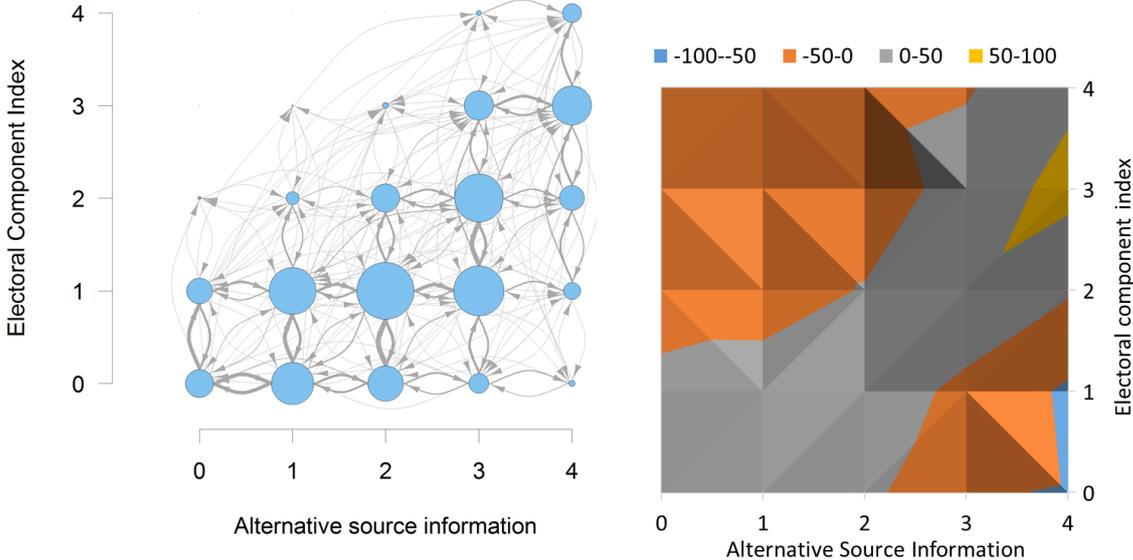
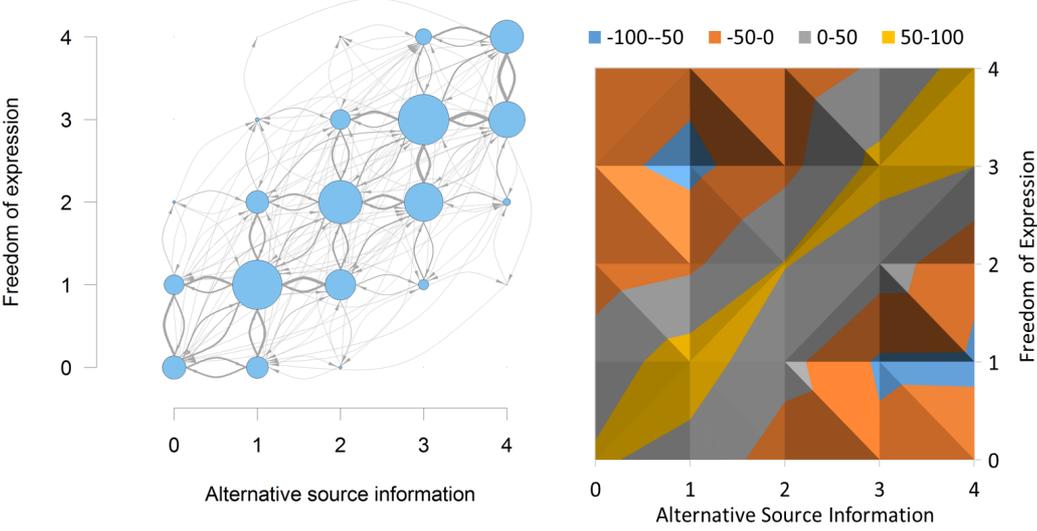


Figure 4. Graphs showing (left) the frequencies of changes and (right) the observed - expected frequencies in the Alternative source information and Freedom of expression. As can be seen, there are more occurrences where the value of Alternative source information is scored as being equal or similar to the value of Freedom of expression. The right figure indicates a ‘preferred reform path’ where Freedom of expression mainly remains similar to Alternative source information. From these results it can be inferred that the Electoral component index changes simultaneously as Alternative source information.



3. Dependency Analysis

To investigate if scoring high on Electoral component index depends on the development of both Freedom of expression and Alternative source information, as suggested by the earlier results, we conduct dependency analyses. Table 4a documents, across all observed combinations, countries’ minimal scores for Freedom of expression and Alternative source information when the country scores 0, 1, 2, 3, or 4 on the Electoral component index. Numbers within parentheses are the absolute minimal values, while numbers outside parentheses are the fifth percentiles, which allow 5% margins of error. For example, there is no country scoring 3 on the Electoral component index when its level of Freedom of expression and Alternative Source information is not at least 1, or 2 if one allows for a 5% wiggle-room.

To rule out the possibility that the development of Freedom of expression and Alternative source information may depend on the Electoral component index, that is, that the minimal values presented in Table 4a are due to correlations and not temporal dependencies, Tables 4b and 4c shows the reversed descriptives to those in 4a. The numbers in Table 4b and 4c are countries’ minimal scores on the Electoral component index when the country scores 0, 1, 2, 3, or 4 for Freedom of expression and Alternative source information. The tables show that only for the highest score of Freedom of expression does the Electoral component index need to score at least 1. Note also that there are interdependencies between Freedom of expression and Alternative source information, indicating that the relationship between these two variables is more akin to a correlation, something that is also indicated earlier in Figure 4.

Table 4. The Electoral component index is dependent on Freedom of expression and Alternative source information, but not the other way around. Freedom of expression and Alternative source information are interdependent.

(a) Electoral component index	0	1	2	3	4
Freedom of expression			1	2 (1)	3 (2)
Alternative source information			1	2 (1)	4 (3)

(b) Freedom of expression	0	1	2	3	4
Electoral component index					1
Alternative source information			1	2 (1)	3 (1)

(c) Alternative source information	0	1	2	3	4
Electoral component index					
Freedom of expression			1	1 (1)	2 (1)

For analyses of sequential relationships between a larger number of variables, dependency tables can be constructed for all possible combinations of variables, and be systematized in various ways, the most informative perhaps being in terms of increasing number of dependencies. A low number of dependencies for variable A indicates that there are very few necessary conditions for it to assume higher (in our case, more democratic) states. If variable B has a high number of contingencies, then this indicates that it never reaches higher (more democratic) states before a number of other variables have reached high levels. Thus, we can for the first time get a good sense of which variables come first, middle, and last, respectively, in processes of democratization.

Table 5 below exemplifies the resulting type of aggregate summary of some 2,000 individual analyses following the dependency analysis approach outlined above, over 22 variables included in the V-Dem indices for electoral and liberal democracy. Here we present the number of necessary conditions for each of these 22 variables reaching their highest state (the top category).

One should naturally not draw any strong conclusions from small differences in the number of dependencies, or necessary conditions, found in such a table. But we can draw pretty strong inferences about sequences from large differences. For example, several indicators of civil liberties have very few dependencies. If we look at which these dependencies are (source data available upon request), they consist exclusively of necessary conditions requiring only one of the lowest levels on other variables. They are thus “weak” dependencies.

Table 5. Number of conditions required to the reach highest state (Category 5)
Necessary conditions % of max
(max = 188)*

Share with suffrage	28	15%
Property rights for men	33	18%

Freedom of movement for men	46	24%
Property rights for women	47	25%
Election voter registry	57	30%
Government intimidation	60	32%
Freedom of movement for women	63	34%
Party ban	62	33%
Barriers to parties	79	42%
Opposition parties autonomy	104	55%
Media self-censorship	109	58%
CSO repression	121	64%
CSO entry and exit	121	64%
Media bias	125	66%
Legislature opposition parties	124	66%
EMB autonomy	134	71%
Compliance with judiciary	142	76%
Access to justice for men	148	79%
Access to justice for women	152	81%
Harassment of journalists	152	81%
High court independence	158	84%
Executive oversight	160	85%

*For this example, we have analyzed 49 variables, summarizing to a total of 192 if all variables were observed in all instances being 4 when the target variable was 4. The sum is 188 as some of the variables had less than 4 states. We here present only a subsection of our variable choice for illustrative purposes.

On the other hand, we note that several variables related to rule of law and oversight of the executive, as well as government influence on the media, have very high numbers in terms of dependencies. This indicates that not only are they dependent on many other variables before they can reach high (=more democratic) levels, but also that for many such variables, they must reach high levels first, before the executive oversight, high court independence, or government media censorship can become really democratic.

The table above is illustrative only. It builds on an actual analysis of V-Dem data but should not be taken as a substantive result. We will use the latest version of the data (v5) to be produced with a new and more reliable measurement model, and we also need to run several tests of robustness and consistency before we can speak of results. Nonetheless, the table above illustrates the kind of nuanced and detailed sequences that can be established empirically with the method proposed here.

4. Bayesian Dynamical Systems

The dynamical system with the highest Bayes' factor when allowing for up to four terms for the variables Freedom of expression (x), Alternative source information (y) and Electoral component index (z) is:

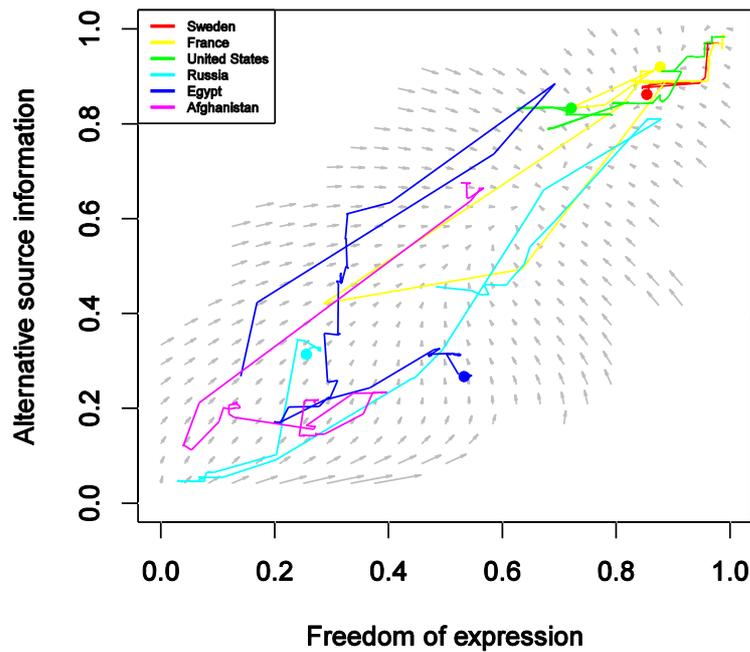
$$dx = \left(-0.069x + 0.024 \frac{y}{x} + 0.015 \frac{x}{y} + 0.032z^2 \right) z \quad (1)$$

$$dy = (0.06 - 0.12y + 0.055xy)z$$

$$dz = (0.099 - 0.44z + 0.31z^2)z + 0.05xy$$

There is an interaction between all the variables. The rate of change of y is linear to the value of z , and x is close to it, with an added cubic term. The rate that is amplified is positive for y unless y is much larger than x , that is, alternative source information cannot remain at high values when there is little freedom of expression. The equation for dx is a bit more complicated, but is positive as long as not both x and y are large. This means that in order for freedom of expression to increase when already high, the electoral component index also needs to be high. Finally, z can grow on its own merits up to 0.28, after which the variable becomes self-limiting. Thus, for the electoral component index to grow beyond a third, both freedom of expression and alternative source information are needed.

Figure 5. Phase portrait showing the expected direction of change given different values of Freedom of expression and Alternative source information. Longer arrows indicate more rapid change. The figure also includes the actual trajectories of six arbitrarily chosen countries.

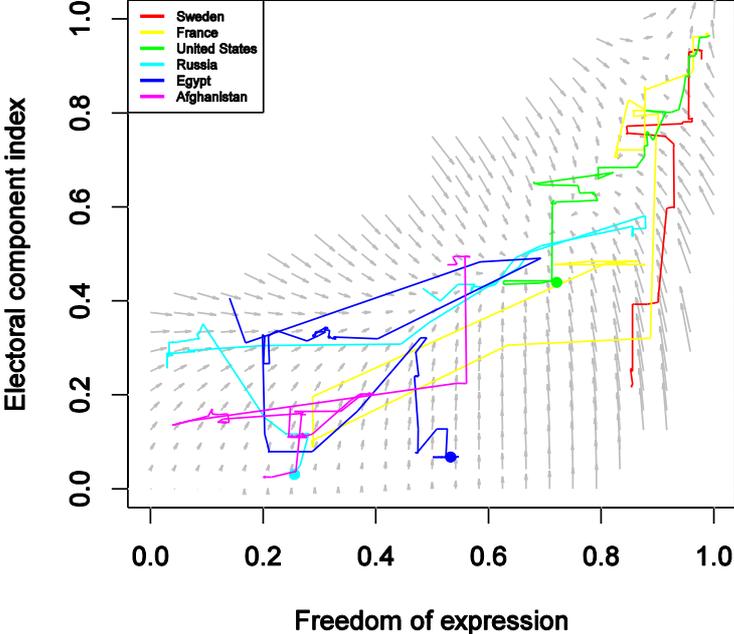


Looking at systems of only two variables at a time enables us to inspect changes graphically in phase portraits, as in Figure 5. Freedom of expression and alternative source information grow largely together. However, while the latter can grow also when the former is small, the former will rarely grow as long as it is larger than the latter.

Figure 6 shows the phase portrait for Freedom of expression and Electoral component index. The system including Alternative source information instead of Freedom of expression is similar. The figure shows that freedom of expression is generally larger than the electoral component index and predicts a trajectory in which the former will attain values larger than the

latter. However, the sequence of events is also here that the variables grow together, and that the Electoral component index grows fastest in the beginning.

Figure 6. Phase portrait showing the expected direction of change given different values of Freedom of expression and Electoral component index. Longer arrows indicate more rapid change. The figure also includes the actual trajectories of six arbitrarily chosen countries.



Concluding Discussion

This paper details a new approach to the study of sequences in social science that includes four supplementary analyses inspired by evolutionary biology, QCA, and Bayesian statistics: frequency analyses, graphical descriptives, dependency tables, and Bayesian dynamic systems analysis. The usefulness of these analytic tools and their supplementary nature is illustrated with a simple sequence analysis of three components of the V-Dem electoral Democracy index (as per the v4 of the V-Dem dataset) and an expanded analysis to illustrate the stringing together of several pairwise comparisons for the dependency analysis.

We believe the approach has great potential for the analysis of many types of pressing issues that social science confronts, not the least the sequences of democratization. We hope to use the tools of this method in the near future to be able to answer critical questions of successful and unsuccessful sequences, as well as what democracy support should focus on at various stages of democratization in ways that can directly inform policy and practitioners' priorities.

References

- Abbott, A. (1995). "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology*, 21: 93-113.
- Abbott, A., & Tsay, A. (2000). "Sequence analysis and optimal matching methods in sociology review and prospect". *Sociological Methods & Research*, 29(1), 3-33.
- Acemoglu, Daron and James A. Robinson. 2001. "A Theory of Political Transitions." *American Economic Review* 91(4): 938-963.
- Agresti, Alan, and Maria Kateri. *Categorical Data Analysis*. Springer Berlin Heidelberg, 2011.
- Beck, N. (2008). *Time-Series-Cross-Section Methods*. Oxford Handbook of Political Methodology, 475-93.
- Boix, Carles. 2003. *Democracy and Redistribution*. New York: Cambridge University Press.
- Buton, François, Claire Lemerrier, and Nicolas Mariot. 2012. "The Household Effect on Electoral Participation. A Contextual Analysis of Voter Signatures from a French Polling Station (1982–2007)." *Electoral Studies* 31(2): 434–47.
- Casper, Gretchen, and Matthew Wilson. 2015. "Using Sequences to Model Crises." *Political Science Research and Methods* 3(02): 381–97.
- Coppedge Michael, Gerring John, Lindberg Staffan I., Skaaning Svend-Erik, and Teorell Jan. 2015a. "Varieties of Democracy: Comparisons and Contrasts". Varieties of Democracy (V-Dem) Project.
- Coppedge, Michael, John Gerring, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kelly McMann, Daniel Pemstein, Svend-Erik Skaaning, Jeffrey Staton, Eitan Tzelgov, Yi-ting Wang, Brigitte Zimmerman, Kyle Marquardt and Farhad Miri. 2015a. "Varieties of Democracy: Dataset v5." Varieties of Democracy (V-Dem) Project.
- Coppedge, Michael, John Gerring, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kelly McMann, Daniel Pemstein, Svend-Erik Skaaning, Jeffrey Staton, Valeriya Mechkova, and Frida Andersson. 2015b. "Varieties of Democracy: Codebook v5." Varieties of Democracy (V-Dem) Project.
- Coppedge, Michael, John Gerring, Staffan I. Lindberg, Daniel Pemstein, Svend-Erik Skaaning, Jan Teorell, Eitan Tzelgov, Kyle Marquardt, Yi-ting Wang, Josefine Pernes, Natalia Stepanova, Farhad Miri, Valeriya Mechkova, and Frida Andersson. 2015c. "Varieties of Democracy: Methodology v5." Varieties of Democracy (V-Dem) Project.
- Dahl, Robert Alan. 1971. *Polyarchy*. Yale University Press.

- Fillieule, Olivier, and Philippe Blanchard. 2012. "Fighting Together. Assessing Continuity and Change in Social Movement Organizations Through the Study of Constituencies' Heterogeneity." In *A Political Sociology of Transnational Europe*, ed. Kauppi Niilo. Colchester, UK: European Consortium for Political Research Press.
- Gauthier, Jacques-Antoine, Eric D. Widmer, Philip Bucher, Cédric Notredame. 2010. "Multichannel sequence analysis applied to social science data." *Sociological Methodology* 40(1): 1-38.
- Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica* 37 (3): 424–438.
- Lindberg, Staffan I. 2015. "Ordinal Versions of the 27 V-Dem Indices: For Classification, Description, Sequencing Analysis and Other Purposes". University of Gothenburg, Varieties of Democracy Institute: V- Working Paper Series, No. 20
- Pemstein, Daniel, Eitan Tzelgov, Yi-ting Wang. 2015. "Evaluating and Improving Item Response Theory Models for Cross-National Expert Surveys." University of Gothenburg, Varieties of Democracy Institute: V-Dem Working Paper 2015:1.
- Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.
- Ranganathan, S., Spaiser, V., Mann, R. P., & Sumpter, D. J. (2014). "Bayesian Dynamical Systems Modelling in the Social Sciences". *PloS one*, 9(1), e86468.
- Rihoux, B., & Ragin, C. C. (Eds.). (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Sage.
- Sillen-Tullberg, Birgitta. "The effect of biased inclusion of taxa on the correlation between discrete characters in phylogenetic trees." *Evolution* (1993): 1182-1191.
- Spaiser, V., Ranganathan, S., Mann, R. P., & Sumpter, D. J. (2014). "The Dynamics of Democracy, Development and Cultural Values". *PloS one*, 9(6), e97856
- Wilson, Matthew Charles. 2014. "Governance Built Step-by-Step: Analysing Sequences to Explain Democratization." In *Advances in Sequence Analysis: Theory, Method, Applications*. eds. Philippe Blanchard, Felix Bühlmann, and Jacques-Antoine Gauthier. New York: Springer International Publishing, 213–27.