Building Sankey Diagrams for Elections

Nuwan I. Senaratna

February 19, 2023

Abstract

This article describes how we might build Sankey Diagrams (a type of flow diagram) for visualizing election results, especially how vote allegiances change from election to election.

We assume that the only data available are the results for each election, and that no data about changes in voter allegiances are available. Hence, the principal problem is to derive changes in voter allegiances, from election results. That is the problem we address in this article.

Contents

List of Figures

List of Tables

List of Algorithms

1 The Problem

1.1 An electoral scenario

Let us consider the following electoral scenario.

Suppose two political parties A and B contest an election in a particular year (E_x) and then contest another election some years later (E_y).

For example, we could consider the 2005 and the 2015 Sri Lankan Presidential Elections, and the UNP and the UPFA (two political parties) [1].

We will also use the following notation.

- Let $v_{A.}$ and $v_{.A}$ denote the number of votes Party A wins in the initial and subsequent elections respectively; similarly, $v_{B.}$ and $v_{.B}$ for B.
- Let $v_{\emptyset.}$ and $v_{.\emptyset}$ denote the number of voters who did not vote (or had their votes rejected) in E_x and E_y respectively.

In this way $v_{A.}$, $v_{B.}$ and $v_{\emptyset.}$ completely describe the results of E_x.

$$E_x = \begin{pmatrix} v_{A.} \\ v_{B.} \\ v_{\emptyset.} \end{pmatrix} \tag{1}$$

[1] For the sake of simplicity, let's ignore other parties. Also, technically, the UNP contested 2015, as the NDF, but we will also ignore this name change for now.

Similarly, $v_{.A}$, $v_{.B}$ and $v_{.\emptyset}$ completely describe the results of E_y.

$$E_y = \begin{pmatrix} v_{.A} \\ v_{.B} \\ v_{.\emptyset} \end{pmatrix} \tag{2}$$

For example, in the 2005 Presidential Election, the UNP won 4,706,366 votes, the UPFA won 4,887,152 votes, while 6,027,393 votes were rejected or not cast [2] [3].

$$E_{2005} = \begin{pmatrix} v_{UNP.} \\ v_{UPFA.} \\ v_{\emptyset.} \end{pmatrix} = \begin{pmatrix} 4{,}706{,}366 \\ 4{,}887{,}152 \\ 6{,}027{,}393 \end{pmatrix} \tag{3}$$

and similarly,

$$E_{2015} = \begin{pmatrix} v_{.UNP} \\ v_{.UPFA} \\ v_{.\emptyset} \end{pmatrix} = \begin{pmatrix} 6{,}217{,}162 \\ 5{,}768{,}090 \\ 2{,}921{,}038 \end{pmatrix} \tag{4}$$

1.2 Sankey Diagrams

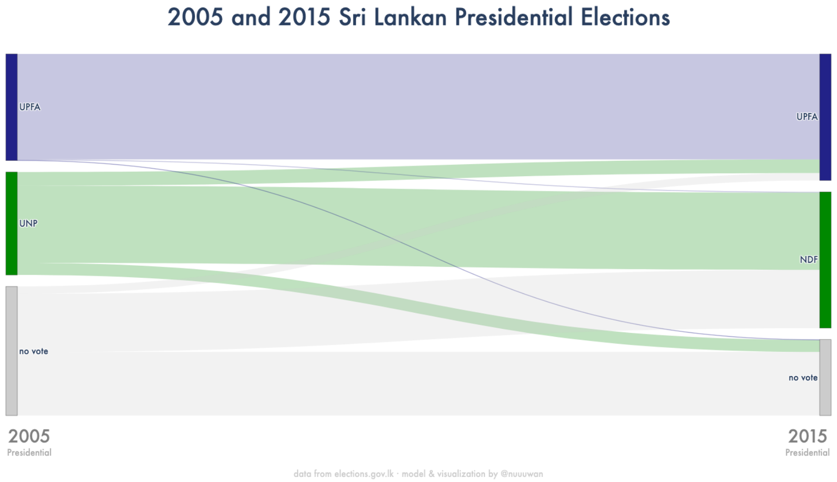
Flow diagram is a collective term for a diagram representing a flow or set of dynamic relationships in a system (Wikipedia contributors (2023a)). For example, we could represent how voter allegiances flow between different political groups in an election.

A Sankey Diagram is a type of flow diagram, in which the width of connections is shown proportionally to the flow quantity (Wikipedia contributors (2023b)).

For example, Figure 1.2 shows a Sankey Diagram of the flow of votes between the 2005 and 2015 Sri Lankan Presidential Elections.

[2] There were several *other* parties contesting in 2005, but as mentioned earlier, we will ignore these, and count them as part of *did not vote*s.

[3] To make the *flows* consistent, count the total number of eligible voters in 2005, to be that of 2015, which is somewhat higher



2005 and 2015 Sri Lankan Presidential Elections

data from elections.gov.lk · model & visualization by @nuuuwan

Figure 1: Sankey Diagram of the flow of votes between the 2005 and 2015 Sri Lankan Presidential Elections(Election Commission of Sri lanka (2005), Election Commission of Sri lanka (2015))

.

The width of the connection between the UNP on the left, and the UPFA on the right is proportional to the number of voters who voted for the UNP in 2005, and went on to vote for the UPFA in 2015.

1.3 What we need to know

In order to draw a Sankey Diagram, we need to know how votes *flowed* from one party in one election to another party in another election; for example, from the UNP in 2005 to the UPFA in 2015.

Or in other words, we need to answer the questions in Table 1.3.

$v_{A \to A}$	Of the voters who voted for party A in E_x, how many of them continued to vote for party in E_y?
$v_{B \to B}$	Of the voters who voted for party B in E_x, how many of them continued to vote for party in E_y?
$v_{A \to B}$	How many voters switched from party A to party B?
$v_{B \to A}$	How many voters switched from party B to party A?
$v_{\emptyset \to A}$	How many voters did not vote in E_x but did vote in E_y? Of these new voters, how many voted for party A?
$v_{\emptyset \to B}$	And how many voted for party B?
$v_{A \to \emptyset}$	How many voters did vote in E_x, but did not vote in E_y? Of these voters, how many voted for party A?
$v_{B \to \emptyset}$	And how many voted for party B?
$v_{\emptyset \to \emptyset}$	Finally, how many voters voted in neither election?

Table 1: What we need to know - a list of questions

Or, in other worlds, how might we find Ω?

$$\Omega = \begin{pmatrix} v_{A \to A} & v_{A \to B} & v_{A \to \emptyset} \\ v_{B \to A} & v_{B \to B} & v_{B \to \emptyset} \\ v_{\emptyset \to A} & v_{\emptyset \to B} & v_{\emptyset \to \emptyset} \end{pmatrix} \tag{5}$$

This information cannot be found in the election results, unless the voters themselves are asked the additional question “For whom did you vote in the previous election?”. It would also require polling those who didn’t vote in both the previous and the current election. This is a lot of work, and it is not always possible to do so.

Instead, we must find ways to estimate these numbers.

2 The Solution

2.1 What we already know

We already know the results of the initial and subsequent election, E_x and E_y.

By definition,

$$\begin{pmatrix} v_{A.} \\ v_{B.} \\ v_{\emptyset.} \end{pmatrix} = \begin{pmatrix} v_{A\to A} & v_{A\to B} & v_{A\to\emptyset} \\ v_{B\to A} & v_{B\to B} & v_{B\to\emptyset} \\ v_{\emptyset\to A} & v_{\emptyset\to B} & v_{\emptyset\to\emptyset} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{6}$$

or

$$E_x = \Omega I \tag{7}$$

and, conversely,

$$\begin{pmatrix} v_{A\to A} & v_{B\to A} & v_{\emptyset\to A} \\ v_{A\to B} & v_{B\to B} & v_{\emptyset\to B} \\ v_{A\to\emptyset} & v_{B\to\emptyset} & v_{\emptyset\to\emptyset} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} v_{.A} \\ v_{.B} \\ v_{.\emptyset} \end{pmatrix}$$

(8)

or

$$\Omega^T I = E_y \tag{9}$$

2.2 What a Linear Model might tell us

Now, suppose we train a linear model, $M_{x\to y}$, that predicts the number of votes that each party wins in E_y, given the number of votes they win in E_x, using data from different polling divisions.

Our linear model represents the following equations:

$$\begin{pmatrix} w_{AA} & w_{AB} & w_{A\emptyset} \\ w_{BA} & w_{BB} & w_{B\emptyset} \\ w_{\emptyset A} & w_{\emptyset B} & w_{\emptyset\emptyset} \end{pmatrix} \begin{pmatrix} v_{A.} \\ v_{B.} \\ v_{\emptyset.} \end{pmatrix} = \begin{pmatrix} v_{.A} \\ v_{.B} \\ v_{.\emptyset} \end{pmatrix} \tag{10}$$

or

$$M_{x\to y} E_x = E_y \tag{11}$$

Now, consider the following matrix $\Omega_{x\to y}$ derived from the component weights of $M_{x\to y}$.

$$\Omega_{x\to y} = \begin{pmatrix} w_{AA}.v_{A} & w_{AB}.v_{A} & w_{A\emptyset}.v_{A} \\ w_{BA}.v_{B} & w_{BB}.v_{B} & w_{B\emptyset}.v_{B} \\ w_{\emptyset A}.v_{\emptyset} & w_{\emptyset B}.v_{\emptyset} & w_{\emptyset\emptyset}.v_{\emptyset} \end{pmatrix} \tag{12}$$

While $\Omega_{x\to y}$ and Ω are not equal, they do share some common properties. For example, the columns of both sum to E_y, that is:

$$\left(\begin{pmatrix} v_{A\to A} & v_{A\to B} & v_{A\to\emptyset} \\ v_{B\to A} & v_{B\to B} & v_{B\to\emptyset} \\ v_{\emptyset\to A} & v_{\emptyset\to B} & v_{\emptyset\to\emptyset} \end{pmatrix} \right)^{T} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} v_{.A} \\ v_{.B} \\ v_{.\emptyset} \end{pmatrix} \tag{13}$$

and

$$\left(\begin{pmatrix} w_{AA}.v_{A} & w_{AB}.v_{A} & w_{A\emptyset}.v_{A} \\ w_{BA}.v_{B} & w_{BB}.v_{B} & w_{B\emptyset}.v_{B} \\ w_{\emptyset A}.v_{\emptyset} & w_{\emptyset B}.v_{\emptyset} & w_{\emptyset\emptyset}.v_{\emptyset} \end{pmatrix} \right)^{T} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} v_{.A} \\ v_{.B} \\ v_{.\emptyset} \end{pmatrix} \tag{14}$$

Now, we can train a *reverse* linear model, $M_{y\to x}$, that predicts the number of votes that each party wins in E_x, given the number of votes they win in E_y. Similar, to before we can derive an $\Omega_{y\to x}$, which shares the same row sums with Ω.

While $\Omega_{x\to y}$ and $\Omega_{y\to x}$ are *similar* to Ω, they are not *equal*. More specifically, while $\Omega_{x\to y}$ shares the same column sum with Ω, it doesn't share the same row sum. And while $\Omega_{y\to x}$ shares the same row sum with Ω, it doesn't share the same column sum.

So, we normalize $\Omega_{x\to y}$ and $\Omega_{y\to x}$ to make them satisfy these conditions.

The Algorithm 2.2 is one way to perform this normalization.

Algorithm 1 Normalize $\Omega_{x\to y}$ and $\Omega_{y\to x}$

```
while $\Omega_{x\to y}$ - $\Omega_{y\to x} \geq \epsilon$ do
    $\Omega_{x\to y_{rows}} \propto \Omega_{rows}$
    $\Omega_{y\to x_{columns}} \propto \Omega_{columns}$
```

ϵ is a small number, e.g. 10^{-2}.

For example, for our 2005 and 2015 scenario, the final estimate for Ω derived is:

$$\Omega = \begin{pmatrix} v_{UNP\to UNP} & v_{UNP\to UPFA} & v_{UNP\to\emptyset} \\ v_{UPFA\to UNP} & v_{UPFA\to UPFA} & v_{UPFA\to\emptyset} \\ v_{\emptyset\to UNP} & v_{\emptyset\to UPFA} & v_{\emptyset\to\emptyset} \end{pmatrix} = \begin{pmatrix} 3{,}526{,}675 & 618{,}536 & 549{,}113 \\ 33{,}894 & 4{,}813{,}969 & 16{,}808 \\ 2{,}656{,}594 & 335{,}585 & 2{,}896{,}569 \end{pmatrix} \tag{15}$$

2.3 Drawing the Sankey Diagram

We can now proceed to draw the Sankey Diagram. Figure 2.3 is such a diagram rendered with the Python library Plotly Plotly (2023), corresponding to the 2005 and 2015 Presidential Election

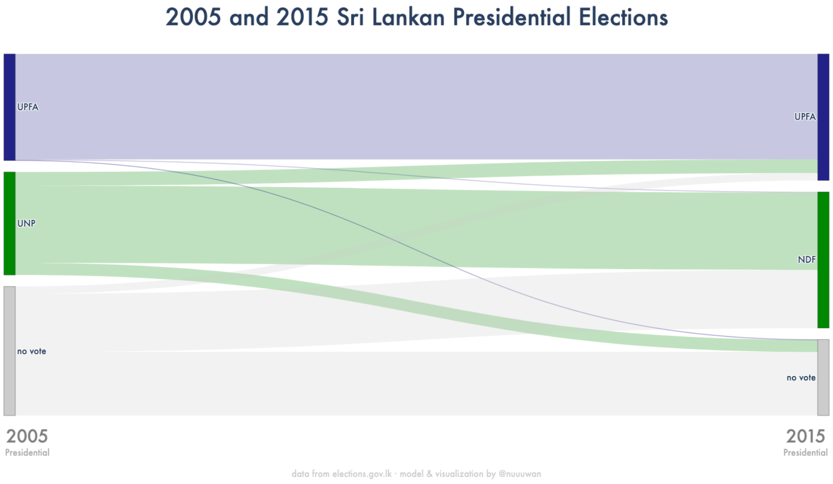


Figure 2: Sankey Diagram of the flow of votes between the 2005 and 2015 Sri Lankan Presidential Elections(Election Commission of Sri lanka (2005), Election Commission of Sri lanka (2015))

.

We can observe the following:

- The *thickest* connections correspond to $v_{UPFA\to UPFA}$ (4,818,474 votes), $v_{UNP\to UNP}$ (3,658,080), $v_{\emptyset\to UNP}$ (2,656,594) and $v_{\emptyset\to\emptyset}$ (2,523,223).
- The narrower connections correspond to $v_{UNP\to UPFA}$ (618,536), $v_{UNP\to\emptyset}$ (549,113) and $v_{\emptyset\to UPFA}$ (335,585).

- The barely-visible connections correspond to $v_{UPFA \to UNP}$ (33,894) and $v_{UPFA \to \emptyset}$ (16,808)

While this article describes transitions between just two elections E_x and E_y, with just two competing parties A and B, the same approach can be applied to transitions between multiple elections each with many and varying numbers of parties.

For example, Figure 2.3 is a multi-step Sankey diagram derived for every Sri Lankan Presidential and Parliamentary including and since 1982.

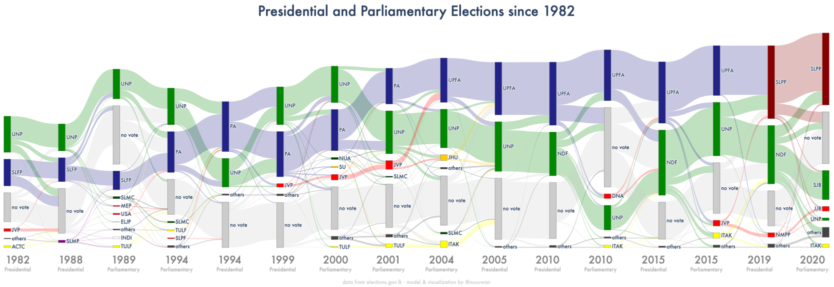


Figure 3: Sankey Diagram for every Sri Lankan Presidential and Parliamentary including and since 1982 Elections(Election Commission of Sri lanka (2019), Election Commission of Sri lanka (2020))

.

2.4 Validating the Solution

The ideal way to validate the results of our method would be to collect actual data on changes in voter allegiances. Sadly, this is likely to be an expensive and time consuming task.

An alternative, but less powerful validation is to inspect the accuracy of the linear regression models $M_{x \to y}$ and $M_{y \to x}$ used in the estimates. A high enough accuracy would imply that the results would be reasonable. This proved, indeed, to be the case for Sri Lankan elections.

References

Election Commission of Sri lanka (2005). 2005 sri lankan presidential election results. [Online; accessed 2023-02-12].

Election Commission of Sri lanka (2015). 2015 sri lankan presidential election results. [Online; accessed 2023-02-12].

Election Commission of Sri lanka (2019). Sri lankan presidential election results. [Online; accessed 2023-02-12].

Election Commission of Sri lanka (2020). Sri lankan parliamentary election results. [Online; accessed 2023-02-12].

Plotly (2023). Sankey diagram in python. [Online; accessed 2023-02-12].

Wikipedia contributors (2023a). Flow diagram. [Online; accessed 2023-02-12].

Wikipedia contributors (2023b). Sankey diagram. [Online; accessed 2023-02-12].