Collecte et visualisation de données législatives avec R

François Briatte

Thématique

Données législatives

- **Nature des données :** amendements, lois, votes… et parlementaires
- **Attributs des documents législatifs :** date d'introduction, vote, mots-clés…
- **Attributs des parlementaires :** âge, sexe, groupe politique, carrière…

Angles d'analyse

- **Unités de référence :** chambres parlementaires, parlementaires individuels
- **Mesures classiques :** productivité, efficacité (probabilité d'adoption)…
- **Mesures relationnelles :** liens entre chambres et entre parlementaires

THE JOURNAL OF LEGISLATIVE STUDIES, 2016
http://dx.doi.org/10.1080/13572334.2016.1163881

Routledge
Taylor & Francis Group

The executive on the battlefield: government amendments and cartel theory in the Chilean Congress

Sergio Toro-Maureira and Nicolás Hurtado

ABSTRACT

This article argues that cartelised coordination inside Chilean congressional committees is important for understanding the success rates of presidential initiatives. By way of an analysis of the amendment process undertaken both in the Chamber and Senate committees in the Chilean Congress during 2006–10, the authors review the approval patterns of legislative amendments. The analysis suggests two chief findings: coordination between government parties and the executive is crucial for the success of amendment; and the opposition's success in generating legal transformations depends on the construction of inter-coalition alliances. It is hoped that the perspective offered here will contribute to the current literature on cartel party theory in Latin America.

KEYWORDS Congress; committees; amendments; Chile; cartel theory

Problème

1. **Disponibilité des données** non systématique

   … avec **plusieurs formats et/ou modalités de collecte** selon les sources

   - Sites Internet officiels des parlements
   - Portails ‘open data’ (officiels ou pas)

2. **Peu de littérature** sur la collecte des données

   … alors que les **incitations à publier des données brutes** augmentent

   - dans les appels à financement
   - dans les processus d’évaluation

Forum

European Union Politics
14(1) 161–178

DOI: 10.1177/1465116512461189
eup.sagepub.com

The collective action of data collection: A data infrastructure on parties, elections and cabinets

Holger Döring
Universität Bremen, Germany

Abstract

Here, I introduce a novel approach towards data collection for comparative research and present a new data infrastructure on parties, elections and governments, the Parliament and Government Composition Database (ParlGov). This data infrastructure combines a database, data presentation in webpages and software scripts in order to generate more dynamic datasets and to facilitate cooperation. So far, it includes information about more than 1000 parties, around 600 elections (national and European Parliament) and almost 1000 governments with their party composition. These observations are linked to a wide set of information about party positions and make it possible to derive various datasets for studies in political science. To provide a first glance into the potential of this new data infrastructure, I map the political space of the European Union (EU) by drawing on this source.

Contraintes

1. **Cadrage comparatif** : collecte de données dans plusieurs chambres parlementaires, sur plusieurs pays

   Échantillon : **33 pays européens**

2. **Méthode programmatique** : aussi peu de récupération manuelle que possible, 100% sources publiques

   Langage privilégié : **R**

   - **Interfaces** avec C, Python, SQL, etc.
   - **Modélisation** de données réseaux

Cadrage

1. Parlements nationaux

- **Chambres** basses (assemblées) et hautes (sénats)
- **Parlementaires** nationaux ou fédérés

  Ex. **Suisse** : Conseil national, Conseil des États

2. Propositions de loi

- **Contrainte légale** en cas d'adoption
- **Signatures nominatives** individuelles

  Ex. **France** : signatures de groupes parlementaires

Network patterns of legislative collaboration in twenty parliaments

Region	Country	Chamber	Period	Years	Legislatures
East	Bulgaria	Unicameral	2005–2015	11	4
	Czech Republic	Lower	1996–2015	20	6
		Upper	1996–2015	20	6
	Estonia	Unicameral	2007–2015	9	3
	Hungary	Unicameral	1998–2015	18	5
	Lithuania	Unicameral	1992–2015	24	6
	Romania	Lower	1996–2015	20	5
		Upper	1996–2015	20	5
	Slovakia	Unicameral	1998–2015	18	5
West	Austria	Lower chamber only	1994–2015	22	7
	Belgium	Lower	1991–2015	25	7
		Upper	1995–2014	20	5
	France	Lower	1986–2015	25*	6
		Upper	1986–2015	30	7
	Ireland	Lower	1997–2015	19	4
		Upper	1997–2015	19	4
	Italy	Lower	1983–2015	33	9
		Upper	1996–2015	20	5
	Portugal	Unicameral	1991–2015	25	7
	Switzerland	Lower	1995–2015	21	5
		Upper	1995–2015	21	5
North	Denmark	Unicameral	2001–2015	15	5
	Finland	Unicameral	1999–2014	16	4
	Iceland	Unicameral	1995–2015	21	6
	Norway	Unicameral	1985–2015	31	8
	Sweden	Unicameral	1988–2015	28	8
Asia	Israel	Unicameral	2009–2015	7	3

*Missing legislature 10 (1993–1997) of the French lower chamber.

Méthodes

1. Scraping

1.1. Données sur les **propositions de loi**

1.2. Données sur les **parlementaires**

Note : **dumps non automatisés** – pas de mise à jour automatique des bases comme avec Python ou Ruby via Morph.io

2. Réseaux

2.1. **Extraction de variables** standardisées

2.2. **Construction et visualisation** des réseaux

Note : **standardisation limitée** – certains attributs des parlementaires et/ou des propositions de loi sont difficilement comparables

althing @ b84d007	Iceland: scraper 2015 update
assembleia @ 4196d58	Portugal: bugfixes
belparl @ cb9d7a1	Belgium: avoid missing file failure
bgparl @ 1f99775	Bulgaria: bugfixes
eduskunta @ a8e666c	fix %e% indexation issue
folketinget @ dabac85	Denmark: minor bug
knesset @ 1b56ac7	Israel: bugfix
nationalrat @ a729e0e	Austria: more data, cleaner code
nrsr @ 76c84b2	compatibility with rvest 0.3.0 (further updates)
oireachtas @ 691f811	Ireland: bugfix
orszaggyules @ 2026aab	Hungary: December 2015 update
parlament @ e8dac0c	Poland: bugfix
parlamento @ 9adf9ac	Italy: final party details
parlamentul @ ecc97a5	Romania: minor fixes
parlement @ 8b27eb4	France: bugfixes
riigikogu @ 85fe2a7	Estonia: fixes to sponsors data
riksdag @ 83e3762	Sweden: bugfixes
seimas @ c026249	Lithuania: final details
stortinget @ e970986	Norway: scraper 2015 update
swparl @ cadc9ff	Switzerland: bugfixes
.gitmodules	add Estonia: submodule and repo init
HOWTO.md	HOWTO: additional package dependency
README.md	README: typo in links to Korea
parlnet.csv	bugfix in Italy 1996 measures

github.com/briatte/parlnet

1. one submodule per country

- replication code
- network data
- codebook and notes

2. additional documentation

- network measures
- links to article preprints
- link to raw data dump

Scraping

- **Download**
  - `utils::download.file`
  - `httr::GET`
  - `psql # interfaçable avec dplyr via RPostgreSQL`
- **Parsing**
  - `rvest::read_html # xml2::read_html`
  - `rvest::read_xml  # xml2::read_xml`
  - `jsonlite::fromJSON`
  - `XML # supplanté par xml2 + rvest`

```r
root = "http://www.althingi.is"
bills = "data/bills.csv"
sponsors = "data/sponsors.csv"

if (!file.exists(bills)) {

  b = data_frame()
  for (i in 145:119) { # accepts down to 20 (1907)

    cat(sprintf("%3.0f", i))

    f = paste0("raw/bill-lists/bills-", i, ".html")

    if (!file.exists(f))
      download.file(paste0(root, "/thingstorf/thingmalalistar-eftir-thingum/lagafrumvorp/?lthing=", i), f,
                    quiet = TRUE, mode = "wb")

    h = read_html(f) %>% html_nodes("#t_malalisti")

    n = html_nodes(h, "td:nth-child(1)") %>% html_text

    if (!length(n)) {

      cat(": no bills\n")

    } else {

      b = rbind(b, data_frame(
        session = i,
        ref = n,
        date = html_nodes(h, "td:nth-child(2)") %>% html_text,
        title = html_nodes(h, "td:nth-child(3)") %>% html_text,
        url = html_nodes(h, "td:nth-child(3) a") %>% html_attr("href"),
        author = html_nodes(h, "td:nth-child(4)") %>% html_text,
        authors = html_nodes(h, "td:nth-child(4) a") %>% html_attr("href")
      ))

      cat(":", sprintf("%5.0f", nrow(b)), "total bills\n")

    }

  }

  b$author = str_clean(b$author)
  b$date = as.Date(strptime(b$date, "%d.%m.%Y"))
  b$n_au = NA

  write.csv(b, bills, row.names = FALSE)

}
```

1. download raw HTML files

to back them up later

2. parse HTML and get variables

using CSSSelect or XPath

3. save to intermediary CSV file

to save time on re-run

Réseaux

- **Assemblage**
  - `base::expand.grid # créer les liens`
  - `base::aggregate   # pondérer les liens`
  - `network::network  # résultat : objet de classe network`
- **Attributs**
  - `sna::degree         # centralité`
  - `tnet::degree_w      # centralité pondérée`
  - `igraph::modularity # modularité`

```r
edges = lapply(unique(data$authors), function(d) {

  w = au$name[ au$authors == d ] # sponsor list is ordered

  d = expand.grid(i = w, j = w[1], stringsAsFactors = FALSE)

  return(data.frame(d, w = length(w) - 1, stringsAsFactors = FALSE)) # number
      of cosponsors

}) %>% bind_rows

# ============================================================================
# EDGE WEIGHTS
# ============================================================================

# first author self-loops, with counts of cosponsors
self = subset(edges, i == j)

# count number of bills per first author
n_au = table(self$j)

# remove self-loops from directed edge list
edges = subset(edges, i != j)

# count number of bills cosponsored per sponsor
n_co = table(edges$i)

# identify directed ties
edges$ij = apply(edges[, 1:2 ], 1, paste0, collapse = "///")

# raw edge counts
raw = table(edges$ij)

# Newman-Fowler weights (weighted quantity of bills cosponsored)
edges = aggregate(w ~ ij, function(x) sum(1 / x), data = edges)

# expand to edge list
edges = data_frame(i = gsub("(.*)///(.*)", "\\1", edges$ij),
                   j = gsub("(.*)///(.*)", "\\2", edges$ij),
                   raw = as.vector(raw[ edges$ij ]), # raw edge counts
                   nfw = edges$w)

# Gross-Shalizi weights (weighted propensity to cosponsor)
edges = merge(edges, aggregate(w ~ j, function(x) sum(1 / x), data = self))
edges$gsw = edges$nfw / edges$w

# sanity check
stopifnot(edges$gsw <= 1)
```

1. buid directed edge list

(possibly redundant)

2. compute edge weights

(see paper appendix)

3. verify edge weights

(non-redundant)

Visualisation

- **Statique**
  - **Syntaxes idiosyncrasiques :** `igraph`, `network`, `sna`
  - **Syntaxes ggplot2** : `geomnet`, `ggraph`, `ggnet`, `ggnetwork`

    `# voir github.com/sctyner/ggnet-paper`
- **Interactive**
  - **Avec R + d3.js** : `networkD3`, `ndtv`
  - **Avec Gephi** : format `GEXF` `# exportable avec rgexf`
  - **Avec JavaScript + PHP** : `Sigma` `# lit le GEXF`

Each network shows bill cosponsorships in the Hungarian Parliament.

All networks are directed graphs drawn with the **Fruchterman-Reingold** force-directed algorithm. The ties connect the first author of each bill to the cosponsors of that bill. The nodes are sized by **unweighted total degree**. When two nodes belong to the same group, any existing tie between them is given the color of that group. There is a **guide to party codes and colors** at the end of this page.

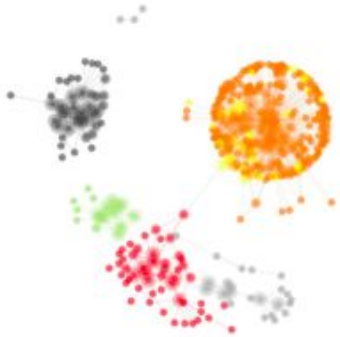
See the **interactive visualization**, or view **other countries**.

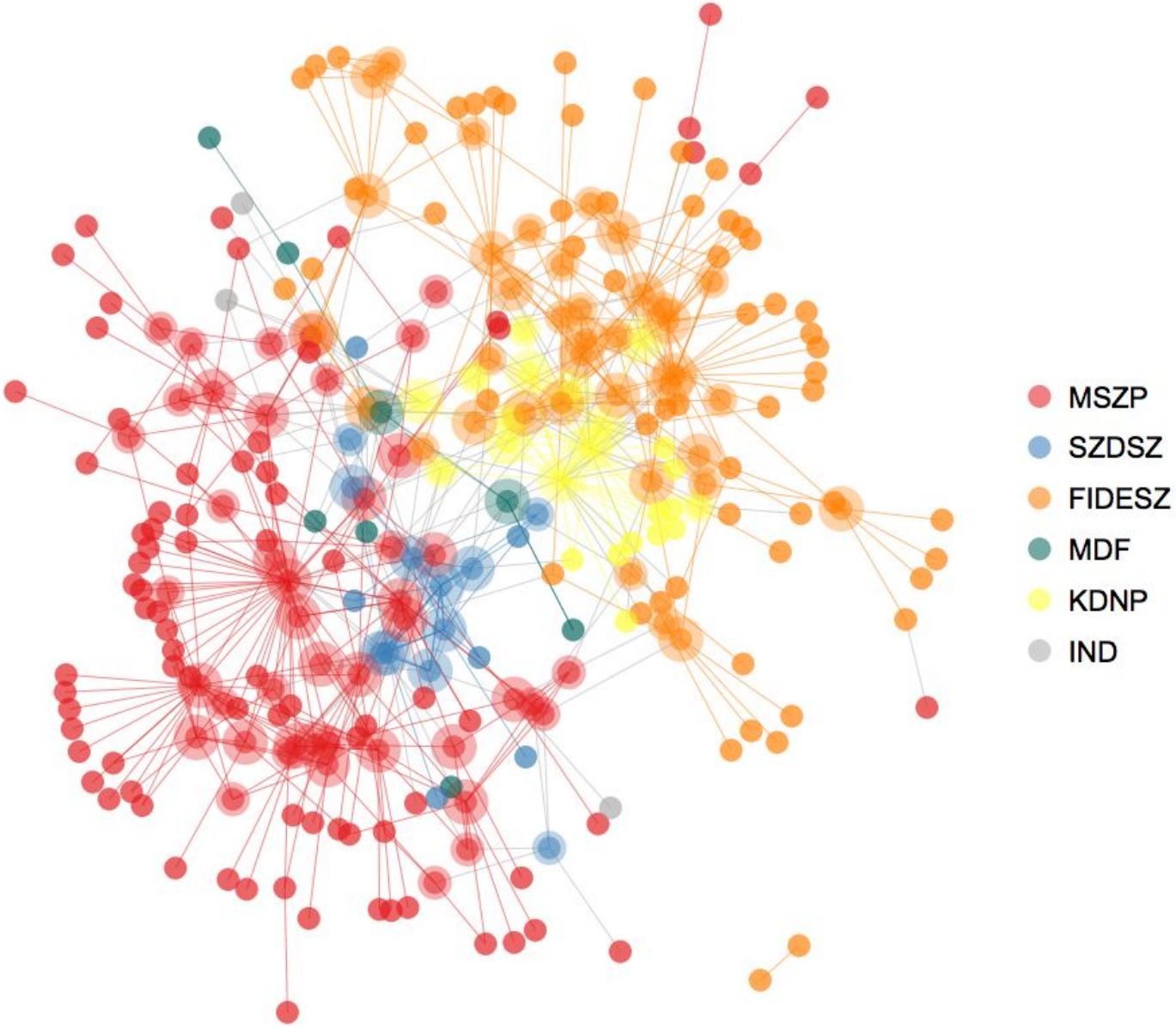
1998–2002

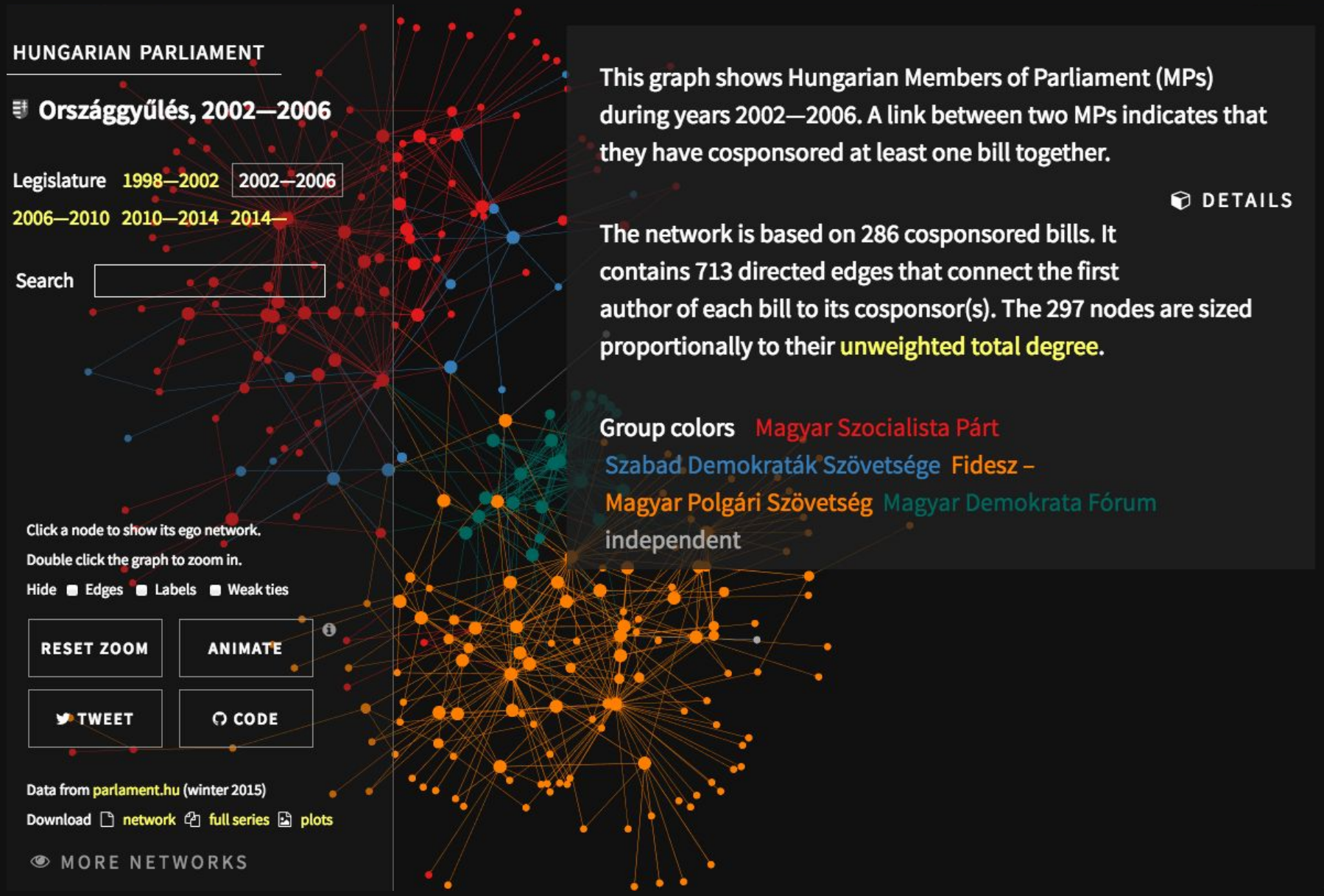
2002–2006

2006–2010

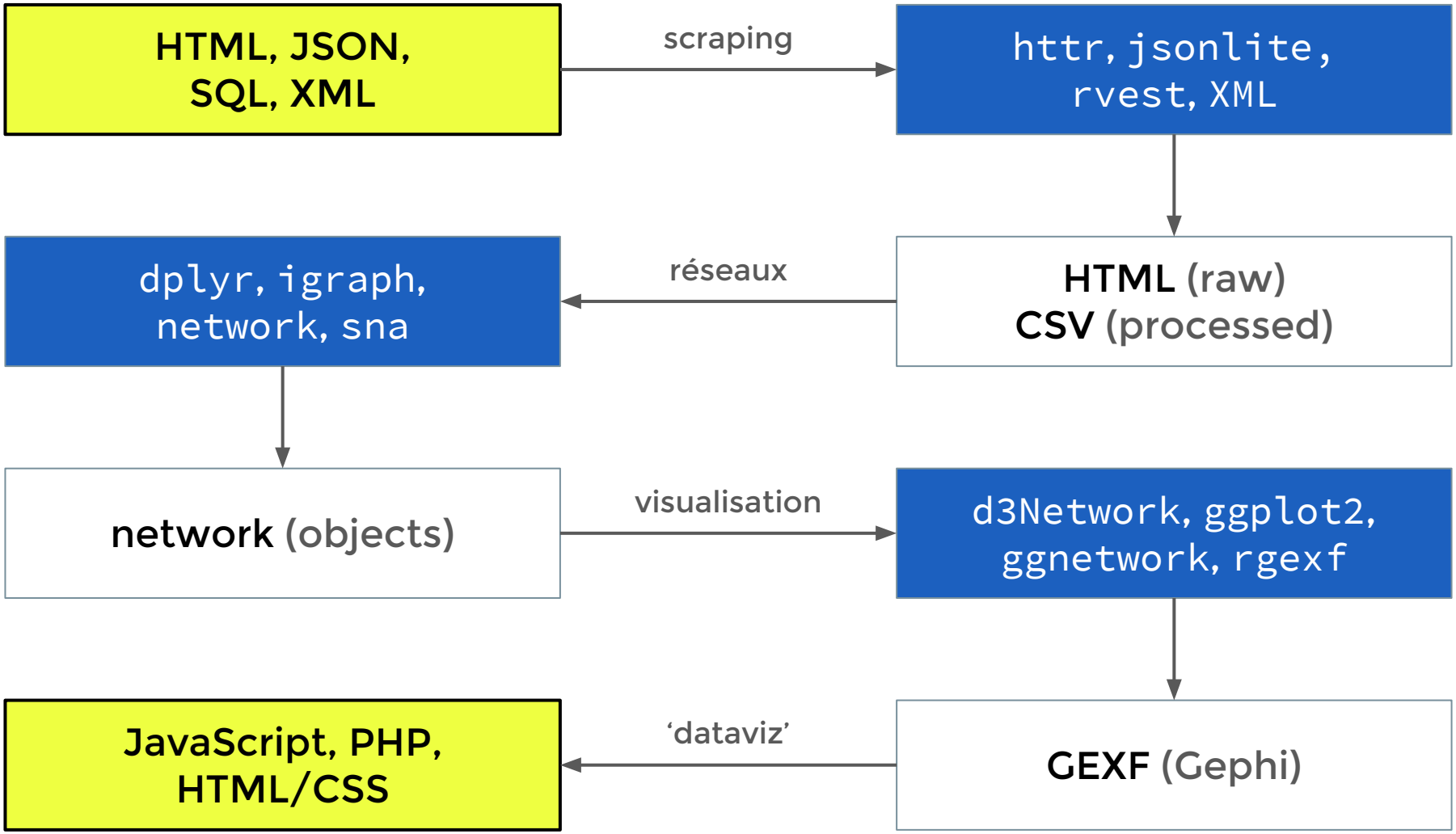
2010–2014




MSZP
SZDSZ
FIDESZ
MDF
KDNP
IND


HUNGARIAN PARLIAMENT
Országgyűlés, 2002—2006
Legislature 1998—2002 2002—2006
2006—2010 2010—2014 2014—
Search
Click a node to show its ego network.
Double click the graph to zoom in.
Hide Edges Labels Weak ties
RESET ZOOM
ANIMATE
TWEET
CODE
Data from parlament.hu (winter 2015)
Download network full series plots
MORE NETWORKS
This graph shows Hungarian Members of Parliament (MPs) during years 2002—2006. A link between two MPs indicates that they have cosponsored at least one bill together.
DETAILS
The network is based on 286 cosponsored bills. It contains 713 directed edges that connect the first author of each bill to its cosponsor(s). The 297 nodes are sized proportionally to their unweighted total degree.
Group colors Magyar Szocialista Párt
Szabad Demokraták Szövetsége Fidesz –
Magyar Polgári Szövetség Magyar Demokrata Fórum
independent

Résumé


HTML, JSON,
SQL, XML
scraping
httr, jsonlite,
rvest, XML
dplyr, igraph,
network, sna
réseaux
HTML (raw)
CSV (processed)
network (objects)
visualisation
d3Network, ggplot2,
ggnetwork, rgexf
JavaScript, PHP,
HTML/CSS
'dataviz'
GEXF (Gephi)

Merci pour votre attention

goo.gl/E96b9M

phnk