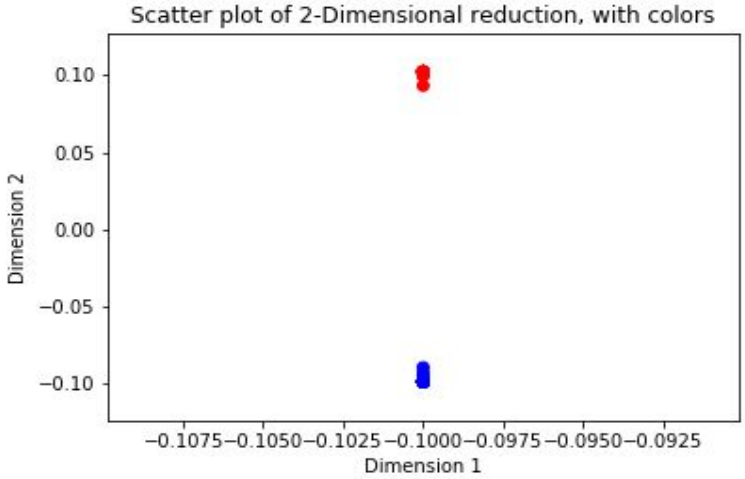
Vote prediction of US Senators from graph properties

Group 7
Mathieu Lamiot
Julien Heitmann
Louis Landelle
Mathias Gonçalves

Co-voting graph

- Milestones showed that there is a clear division
- Adjacency matrix: measure similarity of voting records, apply Gaussian function
- Projection of the Fiedler vector is almost lossless, only Independent Senators are misclassified



US Senate

- Upper chamber of the United States congress, legislature of the US
- 2 senators per state, total of 100 senators (2019: 53 Rep., 45 Dem., 2 Ind.)
- Special powers: hold Impeachment trial, approve presidential appointees

Hearing of Brett Kavanaugh

Project goal

- Learning Task: given a pair (senator, bill), predict the senators vote position

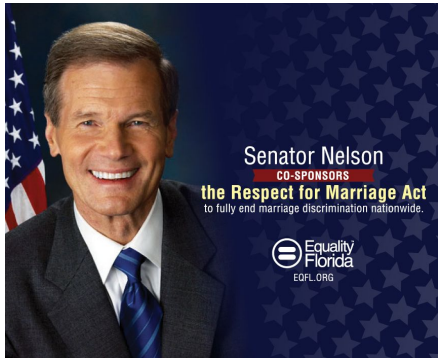
- Data: came from the ProPublica Congress API, used to access a public database (voting records, senator information, committees, etc.)

- A network tour of the senate: using the data we have, build interesting graphs and extract properties such as edge weights, shortest path lengths and other distances as features for a classification task (‘Yes’ or ‘No’)

Co-sponsorship

Bill cosponsors : a "cosponsor" is a senator or representative who adds his or her name as a supporter to the sponsor's bill.

Idea : relate senators more when they appear on the same cosponsoring set of some bill.

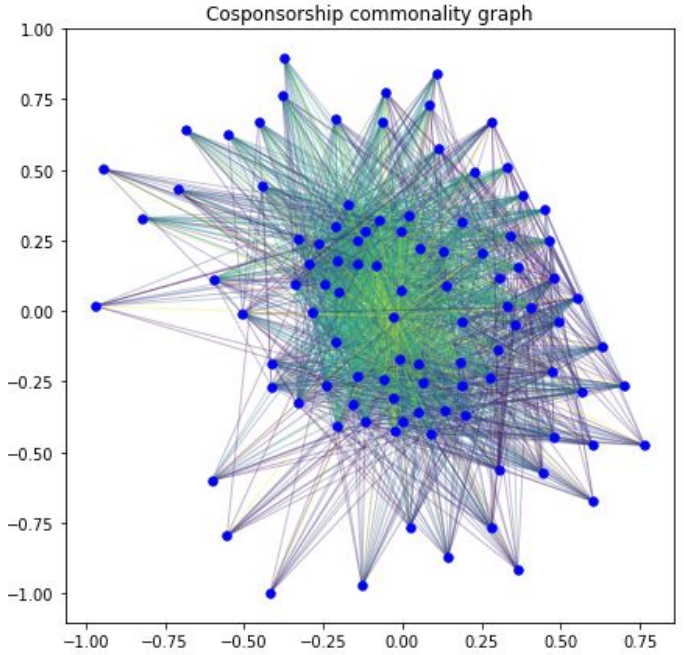


Co-sponsorship graph

Nodes : senators

Edges : $weight_{i,j} = \exp\left(\frac{-jaccard\ index^2}{M^2}\right)$

Feature : sum of distances between senator and a bill cosponsors



Committee co-membership

Senate committee : sub-organization in the senate that handles a specific duty (rather than the general duties of Congress).

- House or senate committees
- Standing or special committees
- Joint committees (bicameral)

Idea : use the distance between the committees of a senator and a voted bill (in our case : standing and special senate committees, we exclude house and joint)

Armed Services Committee during a hearing

Committee co-membership graph

Nodes : committees (not senators)

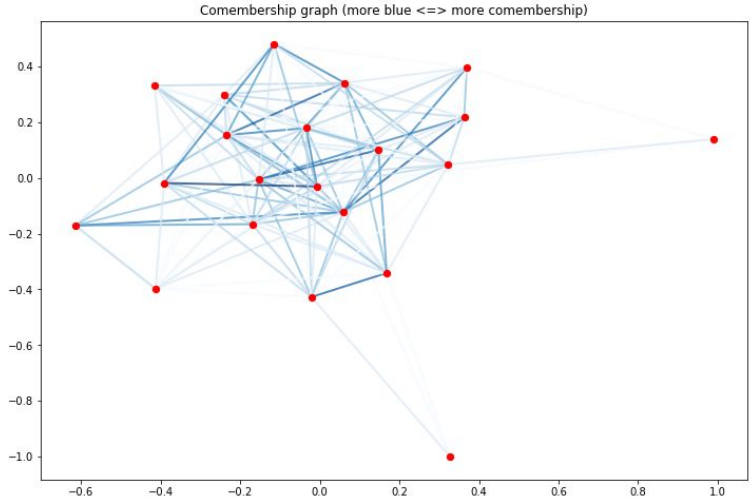
Edges : $C(c1,c2) = |M_{c1} \cap M_{c2}|$

$$\hat{C}(c1,c2) = \frac{C(c1,c2)}{\max_{c1,c2} C(c1,c2)}$$

$$adj_{i,j} = \mathbb{1}\{\hat{C}(c1,c2) > k\}\hat{C}(c1,c2)$$

Feature :

$$min\_dist(s,v) = \min\{shortest\_path(bc,sc), \forall sc \in C_s, \forall bc \in C_b\}$$



Distances as features (1)

Two main ideas:

- Senators have influence on each others.
- Senators can not know all the bills/topics, but they can rely on their peers.

-> **Distance** as a way of measuring influence

-> Available for **all graphs**

-> Need to associate a bill to a node

Distances as features (2)

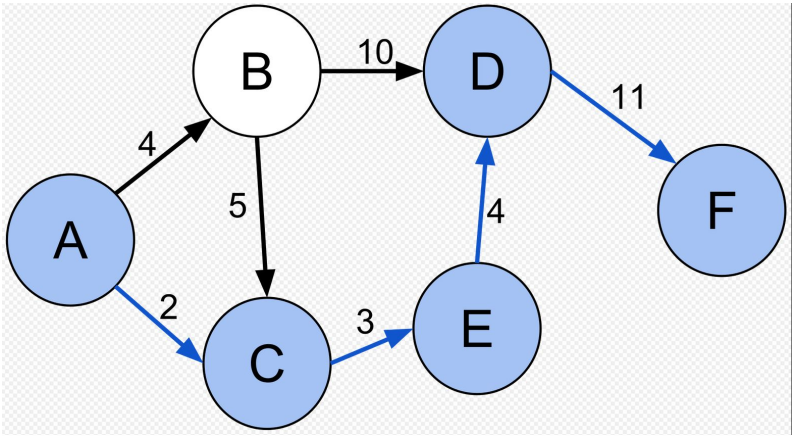
Basic distance: **Shortest path**

-> Our **first feature**

But, it doesn't capture all informations!

Many other possibilities: resistance distance,...

-> We chose "**Heat Kernel Distance**"



Distances as features (3)

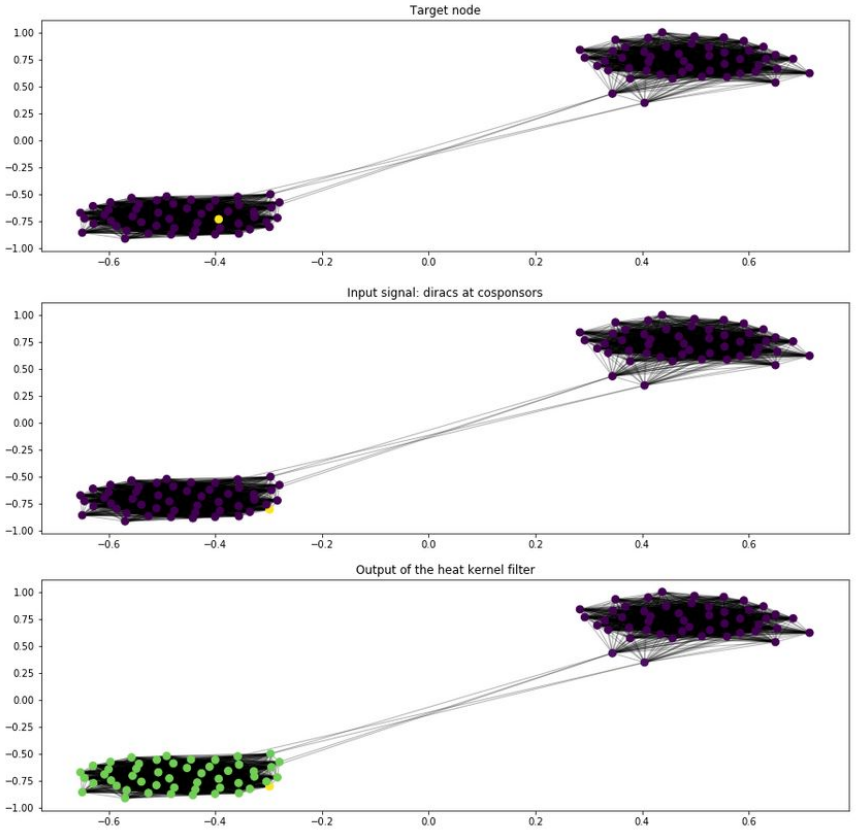
Heat Kernel Distance: How long does it take for the heat to diffuse from the input nodes to the target node ?

Based on the **idea of diffusion.**

f(t,node) = iGFT[exp(-t*e) *GFT(f(0,node)]

Smallest **t** such that:

f(t,target node) > threshold



Dataset

Senator-bill pair

Labels (y):

- Positive vote → 1
- Negative vote → 0

Features (x):

- Laplacian’s 2nd eigenvector (co-voting)
- Shortest path (co-voting)
- Heat kernel distance (co-voting)
- Co-sponsorship features
- Shortest path (co-sponsorship)
- Heat kernel distance (co-sponsorship)
- Minimum distance (co-membership)

Dataset

Senator-bill pair

Requirement for validity:

- One party must sponsor the bill
- The bill must have at least one senator as a co-sponsor
- The senator must have voted for the specific bill

Classifiers

Logistic regression classifier:

73%

Random forest classifier:

82%

Most important feature:

Laplacian’s 2nd eigenvector

How well did we do?

Good prediction accuracy!

LOADS of (senator-bill) pairs didn’t satisfy the requirements...

Thanks for your attention!