

Relatório do projeto Observatório do Congresso

Guilherme Guimarães
February 5, 2020

Abstract

Este relatório trata dos avanços feitos no dia 4 de fevereiro de 2020. Destacando-se a criação de uma hipótese que justifica a baixa modularidade apresentada pela partição resultante do algoritmo Louvain e comprovação dessa hipótese para um caso específico.

1 Hipótese

A partir da leitura do artigo que comunica a respeito do Louvain, percebeu-se, a partir dos exemplos ali apresentados, que a modularidade da partição deveria ser maior do que a que estava sendo obtida. Tendo em vista isso começou-se a investigar os motivos para esse inesperado resultado. Chegou-se a hipótese de que a partição gerada pelo algoritmo era boa, entretanto, sua modularidade era mascarada por uma grande quantidade de arestas de menor importância. Para confirmar essa hipótese foram feitas duas análises, uma quantitativa e outra qualitativa.

2 Análise Quantitativa

Para que a hipótese faça sentido o teste mais básico seria que modularidade da partição gerada no grafo original fosse maior quando calculada no grafo com as arestas mais relevantes (nesse caso o grafo com arestas maiores ou iguais a mediana), e isso se confirma. Dado esse teste inicial, partimos para outros testes.

2.1 Teste de Impacto

A modularidade é calculada, no caso de grafos com pesos, com base nos pesos das arestas. Para que a modularidade seja alta, as arestas que ligam membros de um mesmo grupo, têm que ser mais pesadas que as arestas que ligam membros de grupos diferentes. Baseado nisso, se a hipótese é verdadeira o impacto da retirada das arestas menos relevantes será sentido em grau muito maior pelas arestas entre os grupos. Para aferir esse impacto foi somado o peso das arestas internas a cada grupo antes e depois da retirada, o mesmo foi feito para as arestas que incidem sobre cada grupo. Em seguida comparamos os valores, sendo perceptível uma queda muito mais acentuada nos valores relativos as arestas que incidem sobre os grupos, comprovando que a grande massa de arestas foi retirada das ligações entre grupos.

2.2 Teste de Similaridade

Considerando, segundo a hipótese, que as arestas fracas não interferem expressivamente na clusterização, mas apenas na aferição da modularidade, é natural pensar que os grupos formados a partir do grafo cortado são similares aos grupos formados pelo grafo original. Para testarmos essa ideia quantitativamente usamos a seguinte formula

$$S(A, B) = \frac{|A \triangle B|}{|A \cup B|}$$

onde A e B são grupos formados pela clusterização dos grafos original e cortado, respectivamente.

Quanto mais próximo de 1 for o valor maior a diferença entre os grupos e quanto mais próximo de 0 mais similares os grupos são.

Vale ressaltar que ao eliminar as arestas fracas alguns nós ficaram isolados, gerando assim comunidades unitárias, que foram a maior causa de divergências entre os grupos dos dois grafos. Entretanto, os grupos significativos tiveram seu maior índice de similaridade S igual a 0.43, mostrando que a clusterização foi pouco afetada pela retirada das arestas fracas.

3 Análise Qualitativa

Esta análise se propõe a usar conhecimentos prévios como a coesão dos partidos e suas posições no espectro político para dizer se a clusterização faz sentido frente a realidade.

3.1 Teste da Coesão Partidária

A partir de dados gerados anteriormente eram conhecidos os partidos mais coesos e menos coesos. Com essa informação em mãos verificamos se parlamentares de partidos coesos não foram divididos alocados em grupos distintos, o que se confirmou de forma quase que absoluta, tendo apenas poucas exceções. Também analisamos a alocação de parlamentares pertencentes a partidos menos coesos, e como o esperado, foram divididos e alocados em grupos distintos.

3.2 Teste do Espectro Político

Este teste tem por objetivo identificar se grupos coesos que possuem uma posição bem definida no espectro político estão alocados em grupos diferentes de partidos coesos com características opostas no espectro político. Para tal usamos os 3 partidos não-unitários mais coesos (Novo, PSL, PT). Sabe-se que o PSL e o Novo são de direita e o PT de esquerda, logo o PT não poderia estar alocado junto com nenhum dos dois, o que se confirma nas duas clusterizações realizadas.

4 Conclusão

Dado o resultado dos testes descritos acima, podemos concluir que de fato a clusterização é coerente, entretanto a modularidade tem sido mascarada por arestas menos relevantes. Dado que a retirada da metade menos significativa das arestas não resulta em alterações significativas na clusterização, mas produz alterações significativas na modularidade.