1. Detecção de Comunidades em Redes Complexas (Artigo 9)

Um grafo simples não orientado G consiste de dois conjuntos V e E, sendo V um conjunto finito e não  
vazio cujos elementos são denominados vértices e E um conjunto de subconjuntos de dois elementos de V cujos elementos são denominados arestas. O número de vértices determina a ordem do grafo e é denotado por n =|V | e o número de arestas determina o tamanho do grafo, sendo  
denotado por m =| E |, onde 0 ≤ m ≤ n(n-1)/2. O grau de um vértice vi, d(vi ) é o número de arestas  
incidentes a ele e dois vértices são denominados adjacentes se existe uma aresta entre eles.

Existem inúmeras maneiras de se representar a estrutura de um grafo. A representação matricial é frequentemente utilizada quando há a necessidade de realização de cálculos envolvendo dados estruturais. Dentre as matrizes mais conhecidas, estão as matrizes de adjacência, incidência, laplaciana e outras. Neste trabalho, é utilizada a matriz de adjacência, denotada por *A* (*G*) , sendo a entrada *aij* dessa  
matriz igual a se *vi* e *v j* são adjacentes, ou iguais a ()*,* caso contrário, ∀*i*, *j* =1,..., *n* .

* 1. Análise de agrupamentos

Análise de agrupamentos é o nome dado a um conjunto de métodos que têm por objetivo reunir os objetos em grupos homogêneos, sendo essa homogeneidade avaliada através de uma medida de distância. Dada uma amostra de n objetos, cada um deles medindo p variáveis, procura-se um esquema de agrupamento em k grupos.

Um primeiro passo à aplicação de qualquer algoritmo de agrupamento consiste da definição da matriz de dados. Mais especificamente, selecionada a amostra de *n* objetos, devem-se definir quais atributos (variáveis) serão considerados para os objetos. Esses atributos são considerados no cálculo da  
medida de homogeneidade dos grupos. Dessa forma, os algoritmos de agrupamento trabalham com uma matriz *A* constituída por *n* linhas (objetos) e *p* colunas (variáveis). A similaridade entre objetos é obtida através de coeficientes específicos para cada tipo de variável.

Conforme comentado anteriormente, para avaliar a homogeneidade entre os objetos, utilizamos medidas de similaridade. Consideram-se dois tipos de similaridade: similaridade propriamente dita, que mede quão semelhantes são os objetos, e dissimilaridade, que mede quão diferentes são, sendo uma a complementar da outra. Dessa forma, quanto maior a similaridade, mais semelhante será, e quanto maior a dissimilaridade, menor a semelhança. Assim, definimos a matriz de similaridade, denotada por *B*, como sendo uma matriz de ordem *n*× *p* em que as *n* linhas correspondem aos objetos e as *p* colunas às variáveis.

Supondo atributos quantitativos, as medidas de dissimilaridade mais utilizadas para dois objetos *i* e *j* com essa escala são:

* Distância euclidiana
* Distância de *Manhatan*
* Outras medidas podem ser construídas com base na distância euclidiana

Os métodos de agrupamento são divididos em hierárquicos e não-hierárquicos. No caso dos métodos hierárquicos, o número de grupos não é um parâmetro de entrada. Esses métodos são divididos em duas categorias, quais sejam: os aglomerativos e os divisivos. Nos aglomerativos, inicialmente há *n* grupos de 1 objeto cada, sendo efetuada uma série de uniões até obter *k* grupos; e  
nos divisivos, inicialmente há um único grupo formado por *n* objetos, sendo efetuadas sucessivas divisões dos grupos até que sejam atingidos *k* grupos.

Nos métodos não hierárquicos, o número de grupos (*k*) é especificado previamente. Os grupos formados devem ter: Coesão interna (similaridade interna) e Isolamento (separação). Em geral, os algoritmos computacionais associados a estes métodos são iterativos. Além disso, esses métodos propiciam ter maior capacidade de análise do conjunto de dados de maior porte (com número maior de objetos). Um dos métodos mais conhecido e utilizado em relação aos métodos não hierárquicos é o método *k-means*.

1.2 Medida de qualidade de agrupamento

Existem medidas, denominadas medidas de qualidade, que são aplicadas para avaliar *a posteriori* os *clusters* formados. Em particular, a medida utilizada neste trabalho é o coeficiente de silhueta. Esse coeficiente permite avaliar, para cada objeto, a sua pertinência em relação ao *cluster*, bem como se os *clusters* formados têm uma estrutura natural de agrupamento. Ou seja, permite identificar se cada um dos objetos está bem posicionado em relação ao *cluster*.

No caso da silhueta média, temos que um valor de s acima de 0.7 indica que há uma boa estrutura de agrupamento, já um valor entre 0.51 e 0.7 indica que há uma estrutura razoável de agrupamento e um valor entre 0.26 e 0.5 indica uma estrutura fraca, ou seja, é aconselhável a aplicação de outro método de agrupamento para os dados. E finalmente, os valores abaixo de 0.26 indicam que não há nenhuma estrutura de agrupamento nos dados, ou seja, não existem grupos naturais.

* 1. Problemas de redes e algoritmos de detecção

A detecção de comunidades tem por objetivo a formação de grupos que tenham similaridade entre seus nós, ou seja, que eles desempenhem funções similares dentro do seu grupo. Além disso, tanto o número de comunidades em uma rede quanto o seu tamanho não são conhecidos *a priori*, sendo esses parâmetros determinados mediante a aplicação de algoritmos de detecção de comunidades. Portanto, dada uma rede associada a um grafo *G*(*V* , *E*) com *n* nós e *m* arestas, é possível encontrar subgrupos de nós a partir de qualquer algoritmo de detecção de comunidades. Se considerarmos que *C*1, *C*2, ..., *Ck* são as comunidades encontradas, elas devem satisfazer as seguintes propriedades: *Ci* (interseção)*C j* = 0, *para i* ≠ *j*.

Um dos principais problemas que se apresenta à detecção de comunidades é que, em redes reais, normalmente são um conhecimento prévio no que concerne ao número e ao tamanho das comunidades existentes. Consequentemente, não há nenhuma regra que permita definir a melhor divisão da rede em suas comunidades. Para resolver esse problema, Newman *et* Girvan (2004) propuseram uma medida, denominada Modularidade, definida na próxima subseção.

* 1. Modularidade

A modularidade Q é uma medida originalmente criada para definir um critério de parada para o algoritmo de Newman et Girvan(2004). Ela representa uma das primeiras tentativas de alcançar uma compreensão do princípio do problema de agrupamento em grafos, baseada no conceito de comunidade. Devido a sua eficácia no que concerne ao tempo de processamento, essa medida foi rapidamente difundida em teoria dos grafos e tornou-se um elemento fundamental de muitos métodos de agrupamento.

Teoricamente, esta quantidade mede a proporção de arestas na rede que conectam vértices da mesma comunidade, menos o valor esperado da mesma quantidade na rede com as mesmas divisões de comunidades, mas com conexões aleatórias entre os vértices. Se o número de arestas em uma comunidade não é maior que o valor aleatório, então o valor da modularidade é *Q* = 0 . Valores que se aproximam a *Q* =1, o qual é máximo, indicam uma estrutura forte de comunidade. Em muitas aplicações reais, os valores de *Q* (para as redes associadas) não ultrapassam 0.7 e valores mais altos não são frequentes.

* 1. FastGreedy

Considerando a primeira iteração inicial desse algoritmo, cada vértice da rede é associado a um único objeto de uma comunidade. E, em iterações posteriores, os objetos vão sendo agregados (unidos), definindo, dessa forma, novas comunidades com número maior de objetos. Essa agregação (junção) produz um maior incremento no valor da modularidade *Q*.

O método pode ser representado como uma árvore, na qual as folhas são os vértices da rede original e os nós internos as correspondentes uniões. A árvore é chamada de dendrograma, e representa uma decomposição hierárquica da rede em comunidades, como pode ser visto na Figura2, em que os círculos na parte inferior da figura representam os vértices individuais da rede. Os vértices se  
unem para formar comunidades cada vez maiores como indicado pelas linhas, até chegar ao topo, onde todos estão unidos em uma única comunidade.

1. Homofilia*, group size* e a difusão de informação política

2.1 Introdução

Ao contrário dos meios de comunicação de massa, as informações sobre sites de redes sociais circulam por meio de interações não comerciais entre indivíduos. A crescente importância das plataformas de mídia social, como o Facebook e o Twitter, no fornecimento de informações aos eleitores representa um novo desafio para os formuladores de políticas.1 Enquanto as mídias tradicionais tendem a ser de natureza hierárquica, com um pequeno número de fornecedores fornecendo informações a um grande número dos consumidores, usuários típicos de sites de redes sociais podem consumir e produzir informações nessas plataformas. Além disso, os usuários da mesma plataforma de mídia social podem ser expostos a conteúdos significativamente diferentes, dependendo do conjunto de contas que eles seguem. Por outro lado, leitores de jornais e telespectadores de uma estação de televisão são expostos à mesma informação, pelo menos para uma primeira aproximação. Dadas essas diferenças entre os meios de comunicação tradicionais e as plataformas de mídia social, as práticas na regulação dos meios de comunicação tradicionais podem não se traduzir bem nas plataformas de mídia social.

Para entender melhor a exposição dos eleitores à informação política nas plataformas de mídia social, conceituamos as mídias sociais como uma rede social. Se diferentes tipos de indivíduos têm crenças diferentes e tendem a desenvolver relacionamentos com indivíduos que pensam da mesma forma, um fenômeno conhecido como homofilia, então os usuários podem estar desproporcionalmente expostos a informações políticas com a mesma opinião. Conforme observado acima, a falta de independência entre as fontes de informação pode reduzir a qualidade das informações, ameaçando aumentar a seleção adversa e o risco moral no governo. Além disso, devido à homofilia, os grupos minoritários podem ser menos informados do que os grupos majoritários, o que pode minar a eficácia das instituições democráticas através da redução da competição eleitoral.