

Aplicação de Medidas de Centralidade e Análise da Estrutura da Rede Brasileira de Financiamento de Campanha Eleitoral de 2014

André Manhães Machado

Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514, Goiabeiras - 29075-910 - Vitória - ES
andre.manhaes@gmail.com

Maria Claudia Silva Boeres

Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514, Goiabeiras - 29075-910 - Vitória - ES
boeres@inf.ufes.br

RESUMO

A Análise de Redes Complexas (ARC) é um conjunto de técnicas matemáticas utilizadas para estudar sistemas representados como grafos. Duas abordagens tem se desenvolvido dentro desta área: (i) a análise da topologia da rede, com a proposição de novos modelos e (ii) a análise da importância dos integrantes da rede, com a aplicação de medidas de centralidade. Uma rede complexa inédita é o grafo que representa as contribuições financeiras das eleições brasileiras de 2014. Nestas eleições, os mais de 18.000 candidatos a cargos eleitorais disponibilizaram um documento detalhando as contribuições eleitorais recebidas. A partir desses dados, este trabalho propõe aplicar as técnicas de ARC e modelar como um grafo as contribuições eleitorais das eleições de 2014. Os resultados mostram que a maior componente do grafo é uma rede de pequeno mundo, livre de escala e que há uma correlação forte entre centralidade dos candidatos e número de votos recebidos.

PALAVRAS CHAVE. Redes Complexas. Topologia de Redes. Medidas de Centralidade.

Teoria dos Grafos, Redes Complexas

ABSTRACT

Complex Networks Analysis (CNA) is a set of mathematical techniques for studying complex systems defined as graph. Two different approaches have been developed in CNA: (i) analysis of the network topology, where new models are proposed to capture unique characteristics of networks and (ii) analysis of influence of members within the network, where the application of centrality measures designates a real value to each member. A new and unexplored complex network is the graph representing the financial contributions of Brazilian elections in 2014. In these elections, more than 18,000 running candidates provided a record detailing the election contributions received. Using that dataset, we propose to apply complex networks techniques and to model as a graph the electoral contributions of 2014 elections. The results show that the major component is a small-world network, a scale-free network and there is a strong correlation between candidate centrality and his number of votes.

KEYWORDS. Complex Networks. Network Topology. Centrality Measures.

Graph Theory, Complex Networks

Introdução

Em muitas disciplinas científicas como engenharia, astrofísica, ciências biológicas, ciências sociais e economia, as técnicas de análise de dados modernas transformaram-se em meios de entender as bases de sistema dinâmicos representados como redes complexas [Klemm e Eguíluz, 2002; Junker e Schreiber, 2011]. Durante a última década, as redes complexas tornaram-se populares na análise da heterogeneidade de sistemas dinâmicos com propriedades não lineares [Klemm e Eguíluz, 2002]. Uma rede complexa é um grafo cuja estrutura é irregular, complexa e evolui dinamicamente no tempo, possuindo características topológicas não triviais, derivadas das propriedades emergentes oriundas dos relacionamentos não lineares entres os seus componentes. Normalmente, estas redes possuem milhares ou milhões de nós e arestas [Klemm e Eguíluz, 2002].

Medidas estatísticas locais e globais aplicadas a redes complexas são úteis para entender as inter-relações e o fluxo de informação entre os diferentes componentes dentro destes sistemas. Importante nesta análise é o estudo de três características comumente usadas para determinar a estrutura topológica de uma rede complexa: (a) o menor caminho médio L_G , definido como a média de todos os caminhos médios na rede; (b) o coeficiente de agrupamento global C_G , o qual indica a tendência dos nós para formarem grupos no grafo e (c) a distribuição de grau, a qual informa a probabilidade de um dado nó ter k ligações na rede [Klemm e Eguíluz, 2002]. Estas métricas são utilizadas para classificar as redes em três modelos que são encontrados em diversos sistemas complexos reais: (i) redes aleatórias, nas quais um nó é ligado a outro nó com probabilidade r ; (ii) redes de pequeno mundo, na qual a distância ligando dois nós na rede é pequena quando comparado ao tamanho da rede e (iii) redes livre de escala, nas quais a distribuição de grau de nós segue a função potência.

As redes complexas possuem tipicamente uma distribuição heterogênea dos papéis ou das relações que seus membros tomam dentro da rede [Newman, 2010; Freeman, 1977]. Esta característica de relação ou posição distinta para elementos diferentes da rede levanta a questão de verificar quais participantes são mais centrais ou mais periféricos [Page et al., 1999]. Em análise de redes complexas, as medidas de centralidades são funções que designam um valor numérico aos nós da rede conforme uma noção de prestígio pré-determinada [Newman, 2010]. Existem diferentes tipos de centralidade conforme o conceito de importância ou influência que é utilizado. As principais métricas de centralidade utilizadas na literatura são a centralidade de grau, centralidade de intermediação, centralidade de proximidade e centralidade de autovetor.

A análise de redes complexas tem aplicações em diversos campos científicos como a modelagem de interações de proteínas [Jeong et al., 2001], a difusão de doenças em redes de interações humanas [Wasserman e Faust, 1994], mapeamento de atividades neurológicas [Rubinov e Sporns, 2010], transporte e roteamento de veículos [Sienkiewicz e Hołyst, 2005], evolução de redes ecológicas [Fath et al., 2007] e redes sociais [Pastor-Satorras et al., 2001]. Esta área de pesquisa teve seu início em aplicações no campo social, com o objetivo de identificar os atores mais centrais e de compreender a estrutura da rede. Um dos campos sociais com intenso interesse na última década no Brasil é a análise do sistema de financiamento de campanhas eleitorais [Boas et al., 2014].

As últimas eleições brasileiras de 2014 disponibilizaram mais de 1600 cargos políticos para serem preenchidos por mais de 18.000 candidatos inscritos. A fim de participar neste processo, os candidatos recebem contribuições financeiras para fazer campanha entre o eleitorado. Todo o registro das contribuições dos candidatos é disponibilizado publicamente na Internet pelo Tribunal Superior Eleitoral (TSE). Até onde sabemos, uma rede complexa modelando o financiamento do processo eleitoral brasileiro ainda não foi investigada na literatura científica.

Neste artigo, nós propomos modelar a Rede Brasileira de Financiamento de Campanhas Eleitorais (de agora em diante designado como RBFC) das eleições de 2014 como uma rede complexa. Usando os dados fornecidos pelo Tribunal Superior Eleitoral (TSE), o grafo *RBFC* é definido para representar a existência de relações financeiras (aresta) entre participantes (nós) envolvidos no processo de financiamento eleitoral. O grafo *RBFC* é desconexo, mas a maior componente

representa mais de 96% da rede inteira. A análise da estrutura topológica e as medidas de centralidade são utilizadas para estudar a maior componente da rede *RBFC*. Nosso objetivo neste estudo é mostrar que esta componente é uma rede de pequeno mundo, livre de escala e que há uma correlação forte entre centralidade dos candidatos e o número de votos.

Este artigo é organizado como se segue. A próxima seção estuda as propriedades topológicas e medidas de centralidade em redes complexas. A seção 3 descreve os principais aspectos do processo eleitoral brasileiro e propõe um modelo para a rede *RBFC*. Os resultados e análises são apresentados na Seção 4. A Seção 5 apresenta a conclusão e trabalhos futuros.

Redes Complexas

Noções Gerais

As definições e conceitos de teoria de grafos necessários a esse trabalho estão aqui resumidos e foram retirados de [Junker e Schreiber, 2011; Bondy e Murty, 2007]. Um grafo é definido como o par ordenado $G = (V, E)$, em que o conjunto V designa os vértices e o conjunto E representa as arestas $e = \{v, w\} \in E, v, w \in V$. Adicionalmente, define-se um grafo ponderado como a tupla $G = (V, E, p)$, em que p denota a função $p : E \mapsto \mathbb{R}$ que designa um peso real w_e para a aresta $e = \{v_i, v_j\} \in E$. O peso w_e normalmente representa um custo, fluxo ou capacidade do grafo. Designa-se por $N(v)$ o conjunto de vértices vizinhos de $v \in V$, ou seja $N(v) = \{w : \{v, w\} \in E\}$.

Para a representação matricial de um grafo ponderado $G = (V, E, p)$ é utilizada a matriz de adjacência ponderada. A matriz de adjacência ponderada $W_{n \times n}$, $n = |V|$, é definida como $W_{v,w} = p(\{v, w\})$ se $\{v, w\} \in E$ ou $W_{v,w} = 0$, caso contrário.

Um caminho de $s \in V$ a $t \in V$ no grafo G é definido como um sequência alternada de vértices e arestas, iniciando em s e terminando em t , de tal forma que cada aresta no caminho conecta o vértice anterior com o vértice posterior. O tamanho do caminho de s a t , $d(s, t)$, é definido como a soma dos pesos das arestas que estão no caminho num grafo ponderado ou a quantidade de arestas num grafo não ponderado. Designa-se por *caminho mais curto* ou *geodésica* de s a t , como o caminho que possui o menor valor de $d(s, t)$. Um ciclo é um caminho alternado de vértices e arestas em que o primeiro e último vértices são iguais. Se designa por k – *ciclo* um ciclo que contenha k arestas. Em especial, 3 – *ciclo* é denominado triângulo ou tríade.

Um grafo $G = (V, E)$ é conexo se, para cada par de vértices $v \in V$ e $w \in V$, existe pelo menos um caminho em G ligando v a w . Define-se um subgrafo $H = (V_H, E_H)$ de G como o grafo H em que $V_H \subset V$ e $E_H \subset E$. Uma componente conexa ou simplesmente componente $L = (V_L, E_L)$ de um grafo $G = (V, E)$ é um subgrafo conexo maximal de G , ou seja, a adição de qualquer nó $r \in V - V_H$ em L desconecta o subgrafo L .

Sistemas complexos são caracterizados por representar um grande número de interações, implicando no surgimento de propriedades emergentes que transcendem o princípio da linearidade. Redes complexas são representações formais destes sistemas complexos e, devido às interações não lineares, as redes complexas possuem propriedades topológicas não triviais, propriedades estas não encontradas em grafos aleatórios ou grafos regulares [Junker e Schreiber, 2011]. Uma rede complexa é definida como um grafo $G = (V, E)$ em que o conjunto de nós V e o conjunto de arestas E possuem milhares ou milhões de elementos. A pesquisa em redes complexas se dedica a definir novos conceitos e medidas para caracterizar a topologia de redes reais. Na última década, o principal resultado foi a identificação de princípios e propriedades comuns a maioria das redes analisadas, a partir de resultados obtidos pelo estudo do coeficiente de agrupamento global, da distribuição de grau e do menor caminho médio [Kang et al., 2004]. Além disso, a análise das relações de atores em contextos sociais fez surgir a necessidade de definições de métricas que medissem o grau de relevância destes atores em função das interconexões dentro da rede, dando origem a um conjunto de métricas chamadas de medidas de centralidade [Rubinov e Sporns, 2010]. As próximas subseções apresentam as propriedades topológicas básicas de uma rede complexa e as medidas de centralidades mais utilizadas na literatura.

Propriedades Topológicas

Definido inicialmente por Watts e Strogatz [1998], o coeficiente de agrupamento C mede a probabilidade que dois nós conectados $v_1 \in V$ e $v_2 \in V$ estejam conectados a um terceiro nó $w \in V$ no grafo $G = (V, E)$. Seja h_v o número de vizinhos de v , t_v o número de triângulos observados para v e $h_v(h_v - 1)/2$ o número total de triângulos possíveis para v . Então, o coeficiente de agrupamento $C(v)$ para o nó $v \in V$ é definido com a razão entre o número de triângulos de v e o número total de possíveis triângulos para v :

$$C(v) = \frac{2t_v}{h_v(h_v - 1)} \quad (1)$$

O coeficiente de agrupamento global C_G para o grafo $G = (V, E)$ é definido como a média da soma de todos os coeficientes de agrupamento $C(v)$, $\forall v \in V$:

$$C_G = \frac{1}{|V|} \sum_{v \in V} C(v) \quad (2)$$

onde $|V|$ é o número de nós de G .

O menor caminho médio L_G do grafo $G = (V, E)$ é definido como a média dos comprimentos das geodésicas do grafo $G = (V, E)$. A métrica L_G representa quão bem um grafo está conectado e é um dos parâmetros mais importantes para medir eficiência de transmissão de informações e integração em redes complexas [Rubinov e Sporns, 2010]. A métrica L_G é definida como:

$$L_G = \frac{1}{|V|(|V| - 1)} \sum_{v, w \in V} d(v, w) \quad (3)$$

em que $d(v, w)$ é a distância entre v e w , i.e., o menor caminho entre $v \in V$ e $w \in V$.

A distribuição de grau $P(k)$ é a probabilidade que um nó, escolhido aleatoriamente, tenha k vizinhos. Formalmente, a distribuição de grau $P(k)$ no grafo $G = (V, E)$ é definida como:

$$P(k) = \frac{n_k}{|V|} \quad (4)$$

em que n_k é o número de nós que tem k conexões em G . Redes complexas reais normalmente possuem uma distribuição de grau que segue uma função potência, isto é, a distribuição segue a função $P(k) \propto k^{-\gamma}$, em que k é o grau e γ é o expoente da função potência. Redes complexas com distribuição de grau potência são designadas como redes livre de escala e muitas redes na natureza, ecologia, economia e tecnologia são livre de escala [Ravasz e Barabási, 2003].

Estas três propriedades são usadas para classificar a topologia da rede em três categorias: redes de pequeno mundo, redes livre de escala e redes aleatórias [Motter et al., 2002]. Redes aleatórias são redes que possuem o menor caminho médio L_G e coeficiente de agrupamento global C_G pequenos. Redes de pequeno mundo são redes com um coeficiente de agrupamento global C_G alto e um menor caminho médio L_G pequeno. Redes livre de escala aparecem em contextos de redes dinâmicas, nos quais os nós se conectam preferencialmente àqueles fortemente conectados, produzindo uma distribuição de grau que segue uma função potência [Barabási, 1999].

Medidas de Centralidade

O objetivo das medidas de centralidade é responder a seguinte questão : “Qual é o vértice/elemento mais importante do grafo/rede?” [Junker e Schreiber, 2011]. Em geral, busca-se determinar a importância de cada vértice ou aresta em relação a sua posição na rede [Newman, 2010] e, dado que há diversas concepções de importância, existem distintas medidas de centralidade. As principais medidas de centralidade utilizadas na literatura são: (i) centralidade de grau (*degree centrality* em inglês); (ii) centralidade de autovetor (*eigenvector centrality* em inglês); (iii)

centralidade de intermediação (*betweenness centrality* em inglês) e (iv) centralidade de proximidade (*closeness centrality* em inglês). Nos próximos parágrafos, nós apresentamos as principais medidas de centralidades que podem ser aplicadas num grafo ponderado $G = (V, E, p)$ com matriz de adjacência ponderada W .

Inicialmente proposta por Nieminen [1974], a centralidade de grau (M_g) determina que nós importantes são aqueles que possuem o maior número de conexões diretas com outros nós no grafo [Scott e Carrington, 2011]. Define-se a centralidade de grau $M_g(v)$ de um vértice $v \in V$ como [Newman, 2004]:

$$M_g(v) = \sum_{\{v,w\} \in E} W_{v,w} \quad (5)$$

A centralidade de intermediação (M_b) é uma métrica que calcula a importância de um vértice $v \in V$ em função da razão entre o número de geodésicas que passam por v em relação ao total de geodésicas no grafo $G = (V, E)$. Vértices com valores altos de centralidade de intermediação tem geralmente influência maior no fluxo de informação dentro da rede [Newman, 2010]. Define-se a centralidade de intermediação $M_b(v)$ de um vértice $v \in V$ como:

$$M_b(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (6)$$

em que $\sigma_{st}(v)$ é o número de geodésicas de $s \in V$ a $t \in V$ que passam por $v \in V$ e σ_{st} é o número total de geodésicas de s a t no grafo $G = (V, E, p)$.

A centralidade de proximidade (M_c), proposta inicialmente por Sabidussi [1966], para um nó $v \in V$ é baseada no tamanho das geodésicas entre v e todos os vértices do grafo G . Diferentemente, da centralidade de grau (M_g), a qual mede o envolvimento de um nó v em termos de seus vizinhos imediatos, a centralidade de proximidade calcula a centralidade de um nó em termos da capacidade para acessar, via geodésicas, outros nós na rede. A centralidade de proximidade $M_c(v)$ para o vértice $v \in V$ é definido como [Newman, 2010]:

$$M_c(v) = \frac{1}{\sum_{u \in V} d(v, u)} \quad (7)$$

Bonacich [1972] foi um dos primeiros autores a sugerir que o autovetor da matriz de adjacência de um grafo pode ser uma boa medida de centralidade. A centralidade de autovetor M_e mede a centralidade dos vértices segundo a soma ponderada das conexões diretas e indiretas de cada vértice. A centralidade de autovetor expande a noção de grau de um vértice $v \in V$, incorporando a informação dos graus dos nós vizinhos $N(v)$ de $v \in V$. Esta centralidade supõe que é útil estar conectado a pessoas importantes, na hipótese que contatos poderosos ajudarão quando solicitados. A centralidade de autovetor $M_e(v)$ para o vértice $v \in V$ é definido como:

$$M_e(v) = \frac{1}{\lambda} \sum_{w \in N(v)} W_{v,w} M_e(w) \quad (8)$$

em que λ é o maior autovalor (índice) da matriz de adjacência ponderada W .

Definição e Modelagem da Rede Brasileira de Financiamento de Campanha Eleitoral (RBFC) de 2014

As eleições brasileiras de 2014 simultaneamente elegeram o Presidente da República (1 cargo), um terço dos senadores (27 cargos), todos os membros da Câmara dos Deputados (513 cargos), todos os governadores (27 cargos) e deputados estaduais (1035 cargos), totalizando 1602 vagas.

No processo eleitoral, os candidatos a um cargo eleitoral são obrigados a submeter um relatório com o registro das contribuições para campanha para o Tribunal Regional Eleitoral (TRE), o qual remete posteriormente o relatório para o Superior Tribunal Eleitoral (STE). Estes relatórios, disponíveis em [Court, 2014], fornecem informações detalhadas de toda contribuição financeira para a execução da campanha, incluindo o nome do contribuidor, o nome do beneficiário, o tipo da contribuição, assim como a quantia total e a data da contribuição. Partidos e candidatos podem criar comitês e diretórios com o objetivo de levantar fundos para um ou mais candidatos.

As contribuições financeiras para candidatos podem vir de várias fontes: (a) pessoas físicas; (b) empresas (pessoas jurídicas); (c) outros candidatos; (d) comitês, diretórios e partidos (doravante designados como PCD). Além disso, candidatos podem realizar doações para comitês e diretórios, os quais por sua vez podem doar para candidatos e entre si. Empresas também podem doar para comitês e diretórios. A Figura 1 mostra um diagrama representando os possíveis tipos de doação dentro da rede de financiamento de campanha eleitoral.

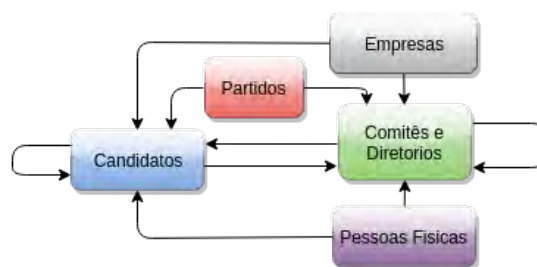


Figura 1: Representação do fluxo de doações eleitorais nas eleições de 2014 no Brasil segundos os diferentes tipos de atores.

Usando os relatórios disponíveis em [Court, 2014], nós propomos neste trabalho modelar a rede *RBFC* como um grafo ponderado $G = (V, E, p)$ em que:

- O conjunto de nós V consiste dos contribuidores e beneficiários no processo de financiamento eleitoral de 2014. Os elementos de V representam os partidos políticos, os comitês, os diretórios, os candidatos, as pessoas jurídicas (empresas) e as pessoas físicas (indivíduos);
- O conjunto de arestas E é tal que $e = \{v, w\} \in E$ se $v \in V$ está recebendo ou efetuando contribuições eleitorais de/ou $w \in V$.
- A função de pesos p designa para cada aresta $e = \{v_1, v_2\}$, um peso $p(w_e) \in \mathbb{R}$ que representa a quantidade de doação realizada entre os nós v_1 e v_2 .

O grafo representando a rede modelada neste trabalho está disponível no repositório de grafos dos autores [Machado e Boeres, 2016].

Resultados e Discussão

A Tabela 1 apresenta as propriedades do grafo *RBFC*. A coluna Propriedades mostra as características e a distribuição dos participantes em cada rede. A coluna *RBFC_M* refere-se a maior componente do grafo *RBFC*. A coluna *RBFC_S* designa o conjunto de componentes do grafo *RBFC*, excluindo a maior componente *RBFC_M*. As colunas R_m e R_s indicam a percentagem dos atributos da coluna Propriedade das redes *RBFC_M* e *RBFC_S* em relação ao mesmo atributo no grafo *RBFC*, respectivamente. A rede *RBFC* consiste de 1232 componentes diferentes. A maior componente *RBFC_M* corresponde a mais de 96% dos nós e mais de 97% das arestas no grafo original *RBFC*. Os nós e arestas restantes estão distribuídos entre as 1231 componentes menores (agrupadas num único grafo designado por *RBFC_S*).

Dado que a componente conexa *RBFC_M* tem a maioria dos nós e arestas do grafo *RBFC* e cada componente em *RBFC_S* tem somente em média 5,8 nós, a análise da topologia

Tabela 1: Comparação entre a componente $RBFC_M$ e o grafo de componentes menores $RBFC_S$.

Propriedade	$RBFC_M$	$RBFC_S$	R_m (%)	R_s (%)
Componentes	1	1231	0.08	99.92
$ V $	180601	7264	96.1	3.9
$ E $	241509	6889	97.2	2.8
Candidatos	17609	1538	92.0	8.0
Empresas	16326	397	97.6	2.4
Pessoas Físicas	146037	5297	96.5	3.5
PCD	628	32	95.1	4.9

da rede e da aplicação das medidas de centralidade será realizada na componente $RBFC_M$. Para a análise da topologia é considerada a versão não ponderada do grafo $RBFC_M$, enquanto as medidas de centralidade tomam em consideração os pesos das arestas em $RBFC_M$.

Para classificar a topologia da rede $RBFC_M$, as métricas L_G e C_G de $RBFC_M$, designadas de agora em diante como L_M e C_M , podem ser comparadas com um grafo aleatório com o mesmo número de nós $|V|$ e o mesmo número de arestas $|E|$. Seja $G_{rand} = (V_{rand}, E_{rand})$ um grafo aleatório com grau médio \bar{k} e $|V_{rand}| \gg \bar{k}$. O menor caminho médio L_{rand} é definido como [de Moura et al., 2003]:

$$L_{rand} \approx \frac{\ln(|V_{rand}|)}{\ln(\bar{k})} \quad (9)$$

em que $\ln(\cdot)$ representa o logaritmo neperiano. Para $|V_{rand}| \gg \bar{k}$, o coeficiente de agrupamento global C_{rand} para um grafo aleatório se aproxima de [de Moura et al., 2003]:

$$C_{rand} \approx \frac{\bar{k}}{|V_{rand}|} \quad (10)$$

em que o grau médio \bar{k} é definido como:

$$\bar{k} = \frac{2|E|}{|V_{rand}|} \quad (11)$$

A Tabela 2 mostra o resultado da comparação entre a maior componente $RBFC_M$ e a rede aleatória G_{rand} de mesmo tamanho. A coluna Propriedade refere-se às propriedades topológicas comparadas e as colunas $RBFC_M$ e $G_{rand}(V_{rand}, E_{rand})$ mostram os respectivos valores para estas propriedades, relativos à maior componente conexa de $RBFC$ e ao grafo aleatório G_{rand} . Quando comparado a um grafo aleatório, a maior componente $RBFC_M$ tem um valor menor para o caminho médio L_M e um valor alto para o coeficiente de agrupamento global C_M . De fato, o coeficiente de agrupamento global C_M é muito maior que C_{rand} , $C_M \gg C_{rand}$. Como o menor caminho médio é pequeno e o coeficiente de agrupamento é alto em relação a uma rede aleatória, pode-se concluir que a componente $RBFC_M$ tem uma estrutura de pequeno mundo [de Moura et al., 2003]. É importante compreender este resultado. O fato da rede ser de pequeno mundo significa que, apesar da rede conter milhares de nós, os participantes representados nela estão muito próximos um dos outros. Dado que a rede $RBFC_M$ representa: (i) quase a totalidade dos participantes no processo de financiamento eleitoral, contendo representantes de todos os estados da federação; (ii) nos quais os candidatos estão concorrendo a distintos cargos de nível federal ou estadual, possuindo propostas políticas que vão da direita a esquerda; (iii) tanto empresas quanto pessoas físicas de todo o território nacional estão presentes; implica que, apesar desta diversidade de qualidade (cargo, território, projeto político, nível da federação), analisando pelo aspecto de financiamento

das eleições, a topologia da rede aproxima os participantes que estão dentro dela. Uma explicação para este comportamento de pequeno mundo é a presença de *hubs* (nós que possuem um grande número ligações) na rede. A presença de *hubs* podem ser inferida da análise da distribuição de grau da rede.

Tabela 2: Comparação entre a componente $RBFC_M$ e a rede aleatória $G_{rand}(V, E)$

Propriedade	$RBFC_M$	$G_{rand}(V_{rand}, E_{rand})$
$ V $	180601	180601
$ E $	241509	241509
\bar{k}	2.67	2.67
C_G	1.749×10^{-2}	1.48×10^{-5}
L_G	5.66	12.30

A Figura 2 mostra a distribuição de grau da componente $RBFC_M$. A linha vermelha corresponde a curva aproximada $P(k) \approx 4.7 \times 10^4 k^{-2.117}$ com $R^2 = 0.9608$, ou seja, a distribuição de grau da componente $RBFC_M$ segue uma função expoente. O expoente γ é 2,117, o qual está dentro do intervalo 2 e 3, tipicamente encontrado em redes complexas reais [Zhou et al., 2005]. Portanto, a componente $RBFC_M$ é uma rede livre de escala. Adicionalmente, uma cauda longa, como mostrado na Figura 2, confirma uma característica da distribuição de grau, indicando a presença de *hubs* (nós) com grau muito maior que a maioria dos outros nós na rede, propriedade também comum em redes de pequeno mundo. Além disso, a componente $RBFC_M$ tem um grande número de nós terminais (nós com grau unitário). Estes dois fatos indicam que a rede $RBFC_M$ tem um grande número de participantes (nós terminais) pouco ou marginalmente ligados ao financiamento da campanha, enquanto outros possuem um envolvimento muito grande (nós *hubs*).

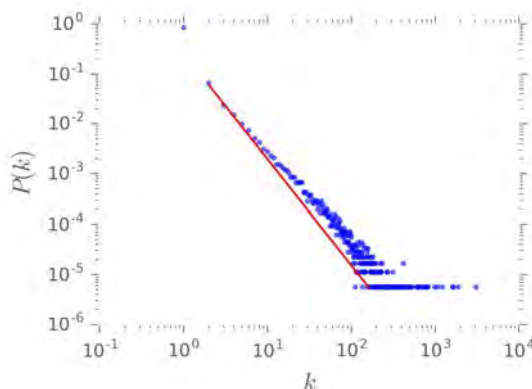


Figura 2: Distribuição de grau $P(k)$ da componente $RBFC_M$.

A análise da topologia da rede $RBFC_M$ mostra que os participantes envolvidos na campanha eleitoral brasileira de 2014 formam uma rede complexa, na qual a maior componente contém mais de 96% dos envolvidos, possui estrutura de pequeno mundo e livre de escala. O próximo passo na análise da estrutura da rede é compreender quais são os participantes mais influentes ou importantes dentro da rede. Dado que a componente $RBFC_M$ contém a maioria dos participantes, as medidas de centralidade são aplicadas nesta componente.

A Tabela 3 apresenta um comparativo dos 15 primeiros resultados da aplicação das medidas de centralidade (equações 5 a 8), em ordem decrescente de importância. A coluna R designa a

posição (entre 1° e 15°) e as colunas M_g , M_b , M_c e M_e , os participantes do processo eleitoral, indicados pelos rótulos do vértices de $RBF C_M$, ordenados pelos valores das medidas de centralidade calculadas. Em negrito estão os participantes que aparecem em pelo menos três colunas da tabela relativas às medidas de centralidade, nas primeiras quinze posições. Os comitês estão representados na tabela pelo nome do partido aos quais estão ligados politicamente seguido do sufixo $-Cn$, onde n é um número natural usado para distinguir comitês distintos do mesmo partido. Os candidatos estão marcados com uma estrela (*), as empresas com uma asterisco (*) e os comitês com um sinal positivo (+).

Pode-se verificar pelos dados apresentados na Tabela 3 que há similaridades entre as distintas métricas, com mais de 50% das ocorrências em cada coluna aparecendo pelo menos em outras duas, com exceção da coluna que indica a centralidade de proximidade (M_c) em que somente sete participantes aparecem em outras duas centralidades. As empresas representam mais de 53% das ocorrências nestas 15 primeiras posições, superando comitês e candidatos, sendo que não há nenhuma pessoa física presente.

Tabela 3: Classificação dos participantes da componente $RBF C_M$ de acordo com o cálculo das medidas de centralidade.

R	M_g	M_b	M_c	M_e
01	JBS *	DILMA V. ROUSSEF *	PSB-C1 +	DILMA V. ROUSSEF *
02	DILMA V. ROUSSEF *	JBS *	PMDB-C1 +	JBS *
03	PSDB-C1 +	ANDRADE GUTIERREZ ENGENHARIA *	PSDB-C1 +	ANDRADE GUTIERREZ ENGENHARIA *
04	AÉCIO N. DA CUNHA *	CONSTRUTORA OAS *	CONSTRUTORA TRIUNFO *	CERVEJARIA PETROPOLIS *
05	ANDRADE GUTIERREZ ENGENHARIA *	PSDB-C1 +	CERVEJARIA PETROPOLIS *	PSDB-C1 +
06	CONSTRUTORA OAS *	CONSTRUTORA NORBERTO ODEBRECHT *	SUCOCITRICO CUTRALE *	CONSTRUTORA OAS *
07	PMDB-C1 +	PSDB-C2 +	COMPANHIA BRASILEIRA DE ALUMINIO *	AÉCIO N. DA CUNHA *
08	CONSTRUTORA QUEIROZ GALVAO *	CERVEJARIA PETROPOLIS *	CRBS *	PMDB-C1 +
09	PMDB-C2 +	PSB-C1 +	UTC ENGENHARIA *	CONSTRUTORA OAS *
10	PSB-C1 +	AÉCIO N. DA CUNHA *	BANCO BTG PACTUAL *	AMIL ASSISTENCIA MEDICA INTERNACIONAL *
11	UTC ENGENHARIA *	PMDB-C1 +	AMIL ASSISTENCIA MEDICA INTERNACIONAL *	CONSTRUTORA TRIUNFO *
12	CONSTRUTORA NORBERTO ODEBRECHT *	CONSTRUTORA QUEIROZ GALVAO *	TELEMONT ENGENHARIA DE TELECOMUNICACOES *	CRBS *
13	PSDB-C2 +	TARSO F. H. GENRO *	ANDRADE GUTIERREZ ENGENHARIA *	PSB-C1 +
14	CERVEJARIA PETROPOLIS *	BANCO BTG PACTUAL *	CONSTRUTORA OAS *	UTC ENGENHARIA *
15	COSAN LUBRIFICANTES E ESPECIALIDADES *	UTC ENGENHARIA	JBS *	GERDAU ACOS ESPECIAIS *

A Tabela 4 apresenta a distribuição percentual dos participantes nas primeiras 1000 posições segundo as classificações das medidas de centralidade. A coluna P designa o tipo de participante e as colunas M_g , M_c , M_b e M_e referem-se às classificações construídas a partir das medidas de centralidade. Analisando os dados, verifica-se que as medidas de centralidades de grau (M_g), autovetor (M_e) e intermediação (M_b) possuem valores similares para a distribuição percentual de candidatos, em torno de 50% a 60%, enquanto que na componente $RBF C_M$ os candidatos representam somente 9,75%. De forma semelhante, as empresas representam de 23% a 33% dos participantes nas primeiras 1000 posições para estas centralidades, enquanto na componente $RBF C_M$ as empresas compõem 9,04% dos membros. As pessoas físicas estão sub-representadas em todas as classificações, exceto a fornecida pela centralidade de autovetor (M_e), com um valor de 26%, mas mesmo neste caso, o valor é inferior ao percentual de pessoas físicas no grafo $RBF C_M$, no qual os indivíduos somam 80% da rede.

A medida de centralidade de proximidade M_c possui valores divergentes em relação às outras métricas para os dados apresentados na Tabela 4. Para verificar se há alguma correspondência entre os resultados das centralidades, a Tabela 5 mostra a aplicação do coeficiente de correlação

Tabela 4: Distribuição percentual das primeiras 1000 posições segundo as medidas de centralidades aplicadas a componente $RBFC_M$.

P	M_g	M_c	M_b	M_e
Candidatos	54,5%	2,6%	59,7%	59,0%
Pessoas Físicas	5,0%	0,0%	7,4 %	26,0%
Empresas	33,3%	78,1%	23,9%	32,8%
PCD	7,2%	21,1%	8,9%	5,60%

de Pearson [Wasserman e Faust, 1994] para as centralidades em estudo. A coluna Métrica refere-se as medidas de centralidade e em negrito estão as correlações fortes entre as distintas métricas. Para o cálculo da correlação, considera-se a classificação completa fornecida pelas centralidades. Verifica-se que há um correlação forte entre as medidas de centralidade de grau (M_g), intermediação (M_b) e autovetor (M_e), característica normalmente encontrada em outras aplicações de medidas de centralidade [Valente et al., 2008]. Além disso, este correlação forte explica o número de ocorrências repetidas para os participantes nas primeiras quinze posições da Tabela 3, dado que há uma uniformidade nas classificações obtidas pelas medidas de centralidade. A medida de centralidade de proximidade (M_c) não possui nenhuma correlação positiva ou negativa com as outras métricas.

Tabela 5: Coeficiente de correlação de Pearson para as medidas de centralidades aplicadas em $RBFC_M$.

Métrica	M_g	M_c	M_b	M_e
M_g	1	≈ 0	0.92	0.96
M_c	≈ 0	1	≈ 0	≈ 0
M_b	0.92	≈ 0	1	0.93
M_e	0.96	≈ 0	0.93	1

A razão de medir a importância ou influência de um ator dentro de uma rede social é poder correlacionar este valor de relevância com dados que sejam externos ao modelo. Com este propósito, a Tabela 6 apresenta o coeficiente de correlação de Spearman entre os valores de centralidade e o número de votos recebidos por cada candidato na componente $RBFC_M$. As linhas da tabela (M_g , M_b , M_c e M_e) representam as medidas de centralidades, enquanto as colunas representam a lista de candidatos para cada métrica (linha). A coluna F_{todos} representa a lista de todos os candidatos. Dado que cada cargo tem sua peculiaridade, também subdividiu-se esta lista em lista de senadores (F_{sn}), deputados estaduais (F_{de}), deputados federais (F_{df}), governadores (F_{gv}) e presidentes (F_{pr}). A correlação é feita da seguinte forma: para cada métrica (linha) obtém-se os valores de centralidade para todos os membros na componente $RBFC_M$ e posteriormente filtra-se esta lista segundo o tipo de candidato (coluna). Posteriormente, correlaciona-se a centralidade dos candidatos filtrados com o seu número de votos usando o coeficiente de Spearman, gerando o valor que está representando no cruzamento entre linha e coluna. Em negrito estão os valores de correlação mais fortes.

Tabela 6: Coeficiente de correlação de Spearman entre as posições dos candidatos e o número de votos.

Métrica	F_{sn}	F_{de}	F_{df}	F_{gv}	F_{pr}	F_{todos}
M_g	0.83	0.77	0.80	0.89	0.92	0.78
M_b	0.61	0.72	0.74	0.76	0.60	0.72
M_c	0.81	0.61	0.65	0.86	0.84	0.62
M_e	0.73	0.34	0.49	0.79	0.80	0.41

Analisando a Tabela 6 verifica-se que a centralidade de grau (M_g) obteve os melhores

resultados para todos os tipos de candidatos (F_{todos}), com a menor correlação para o cargo de deputado estadual (F_{de}) e o maior para o cargo de presidente (F_{pr}). Entretanto, todas as centralidades mostram correlação significativa entre centralidade e número de votos, indicando que há uma correlação entre a posição do candidato dentro da rede $RBFC_M$, denotado pelo valor da sua centralidade, e o número de votos recebidos.

Conclusão e Trabalhos Futuros

Neste trabalho nós propomos uma nova rede complexa real modelada a partir dos registros de financiamento das campanhas eleitorais brasileiras de 2014. Os registros contém os dados financeiros e o envolvimento de candidatos, partidos políticos, comitês, empresas e pessoas físicas no processo de obtenção de recursos para a execução das campanhas políticas de 2014. Aplicando as técnicas da análise de redes complexas, mostramos que a maioria dos participantes que tomam parte no processo de financiamento eleitoral, seja doando ou recebendo recursos, estão dentro da maior componente da rede complexa.

A estrutura topológica da maior componente da rede complexa foi analisada com a aplicação das métricas de coeficiente de agrupamento global, menor caminho médio e o cálculo de distribuição de grau. Os resultados destas propriedades mostram que a rede tem um comportamento de pequeno mundo e uma estrutura de rede livre de escala, ou seja, apesar da rede conter milhares de participantes, em geral os membros da rede estão próximos um dos outros, enquanto que alguns participantes estão fortemente conectados dentro da rede e outros, a maioria dos membros, possuem poucas ou uma conexão na rede.

As medidas de centralidade de grau, intermediação, proximidade e autovetor foram aplicadas à maior componente da rede complexa, destacando os participantes mais importantes ou influentes. Os resultados mostram que, entre os quinze mais importantes em cada centralidade, 7 ocorrem em pelo menos em outras duas centralidades, uma taxa de quase 50% de semelhança para os participantes mais centrais na rede. A análise da classificação dos candidatos mostra que há uma correlação forte entre o valor da centralidade do candidato e o número de votos que ele obteve na eleição para todas as medidas de centralidades aplicadas.

A análise de redes complexas permite diversas outras aplicações na rede de financiamento eleitoral proposta. Como trabalhos futuros destacamos: (i) a análise da evolução da rede usando o conceito de ligação preferencial, o qual analisa o comportamento de novas associações na rede em função da distribuição prévia de conexões; (ii) a análise de comunidades dentro da rede e (iii) a aplicação de outras medidas de centralidade.

Referências

- Barabási, A. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Boas, T. C., Hidalgo, F. D., e Richardson, N. P. (2014). The spoils of victory: Campaign donations and government contracts in Brazil. *The Journal of Politics*, 76(02):415–429.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120.
- Bondy, J.-A. e Murty, U. S. R. (2007). *Graph theory*. Graduate texts in mathematics. Springer, New York, London. ISBN 978-1-8462-8969-9. OHX.
- Court, S. E. (2014). Estatísticas eleitorais - eleições 2014. Disponível em <http://www.tse.jus.br/eleicoes/estatisticas/estatisticas-eleitorais-2014>.
- de Moura, A. P. S., Lai, Y. C., e Motter, A. E. (2003). Signatures of small-world and scale-free properties in large computer programs. *Physical Review E*, 68(1).
- Fath, B. D., Scharler, U. M., Ulanowicz, R. E., e Hannon, B. (2007). Ecological network analysis: network construction. *ecological modelling*, 208(1):49–55.

- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, p. 35–41.
- Jeong, H., Mason, S. P., Barabási, A.-L., e Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833):41–42.
- Junker, B. H. e Schreiber, F. (2011). *Analysis of biological networks*, volume 2. John Wiley & Sons.
- Kang, B., Goh, K., Lee, D., e Kim, D. (2004). Complex networks: structure and dynamics. *Sae Mulli*, 48(2):115–141.
- Klemm, K. e Eguíluz, V. M. (2002). Growing scale-free networks with small-world behavior. *Physical Review E*, 65(5).
- Machado, A. M. e Boeres, M. C. S. (2016). Repositório das redes de financiamento eleitoral das eleições brasileiras de 2014. Disponível em <https://github.com/andremanhaes/grafos-financiamento-2014>.
- Motter, A. E., de Moura, A. P. S., Lai, Y. C., e Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, 65(6).
- Newman, M. (2004). Analysis of weighted networks. *Physical Review E*, 70(5).
- Newman, M. (2010). *Networks: an introduction*. OUP Oxford.
- Nieminen, J. (1974). On the centrality in a graph. *Scand J Psychol*, 15(1):332–336.
- Page, L., Brin, S., Motwani, R., e Winograd, T. (1999). The pagerank citation ranking: bringing order to the web.
- Pastor-Satorras, R., Vázquez, A., e Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical review letters*, 87(25).
- Ravasz, E. e Barabási, A. L. (2003). Hierarchical organization in complex networks. *Physical Review E*, 67(2).
- Rubinov, M. e Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603. ISSN 0033-3123.
- Scott, J. P. e Carrington, P. J. (2011). *The SAGE Handbook of Social Network Analysis*. Sage Publications Ltd. ISBN 1847873952, 9781847873958.
- Sienkiewicz, J. e Hołyst, J. A. (2005). Statistical analysis of 22 public transport networks in poland. *Physical Review E*, 72(4).
- Valente, T. W., Coronges, K., Lakon, C., e Costenbader, E. (2008). How correlated are network centrality measures? *Connections (Toronto, Ont.)*, 28(1):16.
- Wasserman, S. e Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.
- Watts, D. J. e Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *nature*, 393 (6684):440–442.
- Zhou, T., Yan, G., e Wang, B. (2005). Maximal planar networks with large clustering coefficient and power-law degree distribution. *Physical Review E*, 71(4). ISSN 1550-2376.