

Métricas de Centralidade e Conversações em Redes Sociais na Internet: Desvelando Estratégias nos Debates Presidenciais de 2014

Raquel Recuero¹
UCPel

Resumo

O presente trabalho busca explorar os debates presidenciais de 2014 no Twitter, através das métricas de Análise de Redes Sociais. Buscamos analisar e discutir como os candidatos apareceram nas conversações estabelecidas pelos atores utilizando as hashtags oficiais em dois debates presidenciais. Utilizamos, para esta discussão, as métricas de centralidade eigenvector, betweenness e grau.

Palavras-chave: Twitter; análise de redes sociais; centralidade; conversação; redes sociais na internet.

1. Introdução

A Análise de Redes Sociais (ARS) é uma abordagem teórico-metodológica que foca o estudo das estruturas sociais a partir de elementos da Sociometria e da Teoria dos Grafos (Scott, 2000, Wasserman & Faust, 1994). A abordagem, que tem suas raízes na análise estrutural, busca estudar de forma sistemática a estrutura das redes, extraindo daí propriedades amplas e buscando discutir essas propriedades a partir de métricas específicas. A metáfora da rede e as métricas da ARS poderiam, assim, ser aplicadas a produtos das relações sociais. Para o analista, assim, interessam elementos como a posição do nó na rede e as características emergentes desta rede, suas relações e seus atributos (Scott, 2000). Esses elementos auxiliam, assim, a desvelar relações entre esses nós e compreender propriedades estruturais dessas redes.

Neste trabalho, interessa-nos discutir como essa abordagem pode oferecer importantes *insights* para o estudo da participação política online, tomando como objeto as eleições presidenciais de 2014 e de forma específica, os debates dos presidencialistas. Assim, estaremos aplicando a abordagem para aquelas estruturas sociais constituídas no ciberespaço, principalmente através dos sites de rede social e de forma especial, ao Twitter durante os debates dos candidatos à presidência transmitidos pela televisão e pela internet em 2014. Nossa questão central aqui, portanto, está focada em como os candidatos repercutiram suas participações nos debates e o que podemos aprender a partir disso.

2. Sites de Rede Social e Redes Sociais na Internet

Para iniciar essa discussão, assim, é preciso estabelecer as semelhanças e diferenças conceituais entre sites de rede social, redes sociais e mídia social. Os sites de rede social foram definidos por boyd & Ellison (2007) como aquelas ferramentas que permitem o estabelecimento de perfis individuais que

¹ Professora e pesquisadora dos cursos de Jornalismo e Publicidade e Propaganda e do Programa de Pós Graduação em Letras da Universidade Católica de Pelotas e pesquisadora do CNPq. Este trabalho tem o apoio do CNPq, sob o processo número 408650/2013-3.

representam atores sociais, de conexões entre esses atores e da possibilidade de navegar por essas conexões de forma pública ou semi-pública. O site de rede social, assim, é o suporte que proporciona aos indivíduos "traduzir" e complexificar suas redes sociais.

Como dissemos, as redes sociais são metáforas para o estudo dos agrupamentos sociais. Uma rede é, assim, uma forma de representar um grupo, onde os atores podem representar indivíduos e instituições e suas conexões as relações entre esses atores. Redes sociais são parte da sociabilidade humana e assim, seu estudo está focado nos modos sociais dos indivíduos e na estrutura da sociedade. Essas redes online, entretanto, são diferentes das redes offline (Recuero, 2009). Quando um ator constrói uma representação de si no espaço online e estabelece conexões com outros atores, ele não está apenas traduzindo seus grupos sociais, mas igualmente amplificando-os e tornando-os mais densos. Isso porque os sites de rede social mantêm essas conexões permanentemente abertas e permanentemente recebendo e enviando informações. E essas conexões, suportadas pelas ferramentas, proporcionam canais de informação, por onde um ator pode receber e propagar informações, de forma mais ampla e mais permanente. Na esfera da estrutura dos grupos online, assim, temos a representação dos atores, que pode ser compreendida a partir dos perfis nessas redes (no caso específico do Twitter, que abordaremos neste artigo) e das conexões entre os perfis (que neste artigo serão as citações ou retweets que cite um determinado ator).

A apropriação dos sites de rede social com a criação de mais laços entre os atores reduziu as distâncias sociais (vide o estudo de Backstrom et al., 2012), criando maior abrangência para as informações. Os canais de circulação de informação assim se amplificaram e passaram a ser bastante dependentes das ações dos atores, que coletivamente proporcionam que determinadas informações tenham mais visibilidade do que outras, atuando como filtros e câmaras de eco. Nesta intersecção nasce o que chamamos, neste artigo, de mídia social: A ampliação e a reverberação das falas individuais dos atores, refletindo pensamentos e opiniões que acontece a partir das apropriações dos sites de rede social e da tradução das redes sociais no espaço online. Nosso conceito, assim, abarca ao mesmo tempo a dimensão técnica (as ferramentas) e sua apropriação pelas redes sociais e efeitos emergentes (efeitos na difusão de informações). A ideia de "mídia social", aqui, busca salientar o efeito do canal, principalmente através de sua apropriação social emergente como eco, reforço e amplificação de informações. A própria estrutura da internet dá potencial de mídia de massa às informações geradas nos sites de rede social, amplificando os efeitos do dito neste espaço. Hansen, Shneiderman e Smith (2011, p.3-4) destacam que a mídia social compreende uma plataforma de interação social, que torna mais visíveis as conexões sociais e permite que sejam observadas de forma mais direta pelo pesquisador. Para os autores, a mídia social é um conjunto de ferramentas, mais do que uma mídia emergente das ações dos atores sobre essas ferramentas, conforme defendemos. Entretanto, seu conceito aponta para questões relevantes, relacionadas à permanência das conexões, que permite que as mesmas sejam rastreadas e amplificadas pela própria rede.

É nesse âmbito que boyd (2010) aponta características para o que chama de "públicos em rede", conceito bastante semelhante ao de mídia social. Para a autora, o conceito compreende aqueles públicos "reestruturados pela tecnologia", compreendendo, ao mesmo tempo, o espaço constituído por essa

tecnologia e os públicos que são imaginariamente construídos sobre as práticas dessas tecnologias. Para a autora, os públicos em rede têm propriedades estruturais, ou seja, que advém do suporte tecnológico dos sites de rede social e dinâmicas próprias, que vêm da apropriação desse suporte pelos atores. As estruturais dizem respeito à permanência das informações (que uma vez publicadas ali permanecem), a "buscabilidade" (que é uma decorrência direta da permanência, ou seja, que as informações publicadas podem ser buscáveis e recuperáveis), a "escalabilidade" (a propriedade do suporte que permite que uma dada informação seja escalável em termos de espalhamento na rede) e, finalmente, a replicabilidade (onde as informações publicadas por um ator são facilmente reproduzidas pelos outros atores). Além dessas características, temos as dinâmicas, que são: A presença de "audiências invisíveis", que é uma decorrência direta da complexificação das redes sociais pela apropriação dessas ferramentas, e que constitui a presença de audiências mais amplas do que aquelas imaginadas pelo emissor da informação; O colapso dos contextos, que compreende o descolamento da informação de seu contexto original, intencionalmente ou não; E, finalmente, o "borramento" das fronteiras entre o público e o privado. Este conceito muito se assemelha àquele que trouxemos de mídia social, pois traz esses públicos ao mesmo tempo como a tecnologia, sua apropriação e seus efeitos. Boyd aponta, assim, os elementos estruturais e de apropriação que vão constituir o contexto dos efeitos da mídia social conforme defendemos. A mídia social, assim é um fenômeno emergente da conversação entre os diversos atores na rede social (Recuero, 2012).

Usar ARS para compreender fenômenos característicos das redes sociais na internet não é uma novidade. Diversos trabalhos buscaram explorar as propriedades destes grupos a partir das métricas da análise de redes sociais. Garton, Haythorntwaite & Wellmann (2006) apresentaram elementos que permitiriam o estudo das redes sociais online a partir da ARS. Os autores pontuam que, dentro das relações estruturais na rede estão aquelas de comunicação, apontando como a mediação do computador modifica o contexto da análise. Os padrões de comunicação, que nos interessam sobremaneira no estudo da conversação que faremos aqui, são discutidos como uma das aplicações possíveis da abordagem. Mislove et al. (2007) focaram as propriedades emergentes dessas redes. Outros trabalhos, mais direcionados, focaram também o Twitter e suas aplicações. Por exemplo, Cheong & Cheong (2011) analisaram tweets buscando compreender o uso da ferramenta em situações de emergência, focando as enchentes da Austrália de 2010 e 2011. Os autores demonstraram como a rede social complexificava e tornava mais efetiva a disseminação de informações para as redes offline.

3. Mídia Social e Eleições

Muitos trabalhos já focaram as questões centradas nas relações entre a mídia social, os processos eleitorais e a participação política. Kushin e Yamamoto (2010), por exemplo, identificaram uma relação positiva entre atenção à internet e expressão online e participação política em alunos usuários de Facebook. Zúñiga, Jung e Valenzuela (2012), em um trabalho também focando a questão da mídia social e a participação política, indicam que a busca de informações na mídia social é um indicador válido de participação política, apontando o envolvimento do cidadão.

Neste sentido, nossa proposta é explorar como podemos entender redes emergentes (de conversação) (Recuero e Zago, 2009; Recuero 2012) sobre os debates no Twitter e que tipo de indicativo essas redes trazem quando focamos a discussão nos presidenciáveis, através da análise de redes sociais. No próximo capítulo discutiremos esses indicativos e suas métricas.

4. Medidas de Centralidade e sua Aplicabilidade para Redes Emergentes

Neste artigo, focaremos, de modo específico em três medidas de centralidade, que consideramos relevantes para a análise do desempenho dos candidatos nas conversações na internet. A centralidade refere-se à uma medida relativa aos nós da rede. Quando buscamos compreender o quão central é um determinado nó, estamos avaliando sua relevância para a rede. Também observaremos a posição do nó com relação ao grupo onde esse nó está e para separar os grupos, utilizaremos algoritmos de modularidade.

Os grafos, dentro da ARS podem ser direcionados ou não direcionados. O grafo direcionado é aquele onde a conexão entre os nós A e B tem uma direção. Assim, por exemplo, em nosso caso, no Twitter, um nó A que cita um nó B em seu tweet constrói uma relação $A \rightarrow B$. Se B cita A, essa conexão fica, então bidirecional ($A \leftrightarrow B$). Há também grafos não direcionados, ou seja, quando não é possível determinar a direção da conexão. Neste trabalho, utilizaremos grafos do Twitter que são, assim, direcionados.

a) Grau² - O grau de um determinado nó refere-se ao número de conexões recebidas (indegree) e o número de conexões feitas (outdegree). Scott (2000) discute o grau como uma métrica de centralidade local, ou seja, de um nó em relação aos demais. No caso de uma rede de conversação, o indegree refere-se ao número de interações que um determinado nó recebeu como direcionadas a ele. Assim, o número de menções direcionadas e retweets de tweets de uma determinada conta no Twitter atribuem seu indegree no nosso mapa. Já o outdegree refere-se ao número de conexões que um determinado nó faz. Deste modo, o número de menções e retweets realizados pela mesma conta (não aqueles feitos a ela) constituem-se no seu outdegree. Em termos conversacionais, quanto mais citado é um determinado ator, mais ele aparece e é central (como assunto) para a conversa que está acontecendo no momento. Quanto mais ele cita outros atores, mais ele participa e responde. Assim, um equilíbrio entre indegree e outdegree indica um ator ativo da conversa. Já um indegree muito maior do que um outdegree indica um ator que é assunto da conversa, mas não participa. O contrário, um outdegree muito maior do que um indegree indica que o ator participa, mas não necessariamente está sendo respondido pelos demais (pode ser, por exemplo, uma conta de spam).

b) Centralidade Betweenness (Freeman, 1979) - A centralidade betweenness indica o quão "ponte" é uma determinada conta. Ela foca o número de caminhos entre as conexões que precisa passar pelo nó entre os vários grupos. Quanto maior essa métrica, mais "ponte" entre diferentes grupos um mesmo nó é (Scott, 2000). O nó, portanto, tem um alto grau de intermediação na rede. A medida desta centralidade de um nó determinado é dada pelo cálculo da fração de caminhos mais curtos entre dos nós que necessariamente

² Embora "grau" seja também referida como uma métrica de centralidade (vide Newmann, 2003), para este trabalho referimos apenas como "grau" para evitar confusões com as demais métricas.

precisam "passar" pelo nó em questão. Newmann (2003) explica, assim, que a centralidade betweenness pode ser uma métrica de "influência que um determinado nó tem no espalhamento de informação na rede"³. Em termos conversacionais, um nó que tem uma centralidade betweenness alta entre citações e retweets é um nó que foi muito citado ou participou da conversa com diferentes grupos (essa métrica pode ser avaliada em conjunto com a de indegree e outdegree). Portanto, quando dizemos que um determinado ator tem uma centralidade betweenness muito alta no Twitter, estamos observando que ele serve de "ponte" entre diferentes grupos. Como estamos avaliando conversações, o ator foi muito citado/participou da conversa com muitos grupos diferentes.

c) Centralidade Eigenvector (Bonacich 1972)- Esta métrica não apenas leva em conta as conexões de um determinado nó A, mas as conexões também dos nós que se conectam a ele. Assim, ela mede a importância do nó na rede também pela importância de seus vizinhos. Difere-se da métrica de indegree porque, enquanto apenas pelo número de conexões um determinado nó pode ser muito citado por nós sem importância (imaginemos, por exemplo, bots de spam, que não são citados na rede), pois não se dá atenção à rede indireta. Com a métrica da centralidade eigenvector, é possível compreender melhor o quanto relevante é o nó para a rede, pois avaliam-se também as conexões não diretas. Assim, não basta apenas um determinado ator ser citado, mas leva-se em conta, também, se os atores que o citam também são citados por outros no decorrer da conversação. Em termos de Twitter, essa métrica tem problemas e vantagens. Por exemplo, se um determinado tweet realizado pelo ator A que cita ator B recebe muitos retweets, tanto ator A quanto ator B poderão ter sua centralidade eigenvector aumentada, A porque está conectado a B e B porque está conectado a A. A prática de retweets entre contas de apoio, assim, pode artificialmente inflar essa centralidade. Entretanto, isso é possível de ser avaliado pelo pesquisador na medida em que se vê o núcleo onde essas contas estão na rede. Por outro lado, a métrica pode ser influenciada por contas que são muito citadas por sua credibilidade, auxiliando, assim, a delimitar a importância do ator na rede. Essa métrica também é associada a prestígio e influência na rede.

d) Modularidade - Neste trabalho, também utilizaremos o conceito de modularidade. A modularidade é uma métrica de rede, não de nó, e refere-se ao algoritmo utilizado para observar o número de clusters de um determinado grafo. A modularidade refere-se a uma medida de vizinhança, ou seja, do quanto um determinado nó tende a aparecer dentro de um determinado grupo. Não se trata, assim, de uma métrica de centralidade, mas de uma métrica de grupo, que identifica grupos, o que nos auxiliará a observar a centralidade dos nós também dentro dos referidos grupos.

5. Abordagem Metodológica

Para este trabalho, utilizamos uma abordagem metodológica de coleta e outra de análise de dados. Conforme explicamos, nosso objetivo é discutir como a análise de redes sociais permite que se extraiam elementos relevantes para a compreensão das redes de conversação (Recuero e Zago, 2009) que emergem em torno dos debates, enquanto esses acontecem na televisão, no Twitter. Os debates mapeados aconteceram em diferentes emissoras de TV e todos tiveram transmissão ao vivo pela internet. Para cada debate, selecionamos uma hashtag genérica e relacionada às narrativas dos debates em si (havia hashtags

³ the influence a node has over the spread of information through the network

de cada candidato, propostas pela militância). Foram realizadas várias coletas de cada hashtag e, para a análise, utilizaremos dois momentos de cada debate. Outros detalhes sobre os eventos na Tabela 1.

1) Debate da Band - Debate que aconteceu na rede Bandeirantes no dia 26/08/2014, terça-feira, às 22h, com duração de pouco mais de duas horas. Neste trabalho, analisaremos a hashtag "#DebatenaBand".

2) Debate do SBT - Debate que aconteceu no SBT (Sistema Brasileiro de Televisão) no dia 01/09/2014, segunda-feira. Este debate aconteceu às 17h45min e durou, também, pouco mais de duas horas. Para a análise, focaremos nas redes através da hashtag "#DebatenoSBT".

Tabela 1: Dados das Redes coletadas nos debates

Hashtag #DebatenaBand	Rede 1 23:30	Rede 2 00:36
Total de nós do grafo	10288	5298
Total de conexões	14373	7066
Horário de coleta	23:30	00:36
Total de Tweets	18.000	7942

Hashtag #DebatenoSBT	Rede 1 18:45	Rede 2 20:01
Total de nós no grafo	6711	7568
Total de conexões	10.418	10891
Horário da Coleta	18:45	20:01
Total de tweets	20.000	20.000

A coleta de dados foi feita através do NodeXL⁴, um software que permite o *crawling*⁵ dos tweets a partir da REST API. Essa API limita as buscas a até cerca de 18 mil tweets ou sete dias no passado. A coleta foi realizada várias vezes durante os debates, sempre após cada bloco. Para este trabalho, utilizamos duas coletas: Aquela realizada ao meio de cada debate e aquela realizada ao final. O objetivo, ao trabalhar com dois mapas, é tentar capturar um pouco melhor a dinâmica da rede de conversação e não apenas um quadro final, que achate as discussões. A análise de dados foi feita utilizando a parte de análise do próprio NodeXL, que a partir dos parâmetros apontados pelo pesquisador, calcula os graus de centralidade especificados. Para a análise dos dados, aqueles atores com mais de uma conexão entre si (por exemplo, aqueles que enviaram e receberam vários tweets para a mesma conta) tiveram essas conexões "achatadas" e representadas dentro da mesma proposta, mudando-se o "peso" da conexão. Trabalharemos, nesta análise, com grafos direcionados e focaremos as métricas de centralidade discutidas no item 2.2: indegree, betweenness e eigenvector. Essa análise foca o perfil oficial dos candidatos. Assim, falaremos brevemente sobre outros nós que também têm posição proeminente na rede, mas nosso foco será, de modo específico, a posição dos candidatos na rede. A visualização dos dados, que será apresentada neste trabalho através de grafos, foi realizada com o NodeXL ou com o Gephi⁶ (que também calcula métricas de análise de redes sociais).

6. Análise

Como dissemos, essas métricas identificam atributos dos nós e suas relações (Scott, 2001), e por isso, centraremos a análise nos candidatos à presidência, embora destacando outros nós relevantes. A seguir, comentamos cada debate.

⁴ <http://nodexl.codeplex.com>

⁵ Coleta automatizada por meio de um software.

⁶ <https://gephi.github.io/>

6.1 O Primeiro Debate

O primeiro debate aconteceu nas dependências da Rede Bandeirantes e foi transmitido ao vivo no site da emissora, às 22h do dia 26 de agosto. Durou cerca três horas e contou com a participação de Dilma Rousseff (PT, Coligação PMDB, PDT, PC do B, PP, PR, PSD, PROS e PRB), Aécio Neves (PSDB, Coligação PSDB, DEM, PTB, SD, PMN, PTC, PT do B, PEN e PTN), Marina Silva (PSB, Coligação PSB, PRP, PPS, PSL, PPL e PHS), Luciana Genro (PSOL), Pastor Everaldo (PSC) e Levy Fidelix (PRTB). Neste debate, os primeiros blocos mostraram uma disputa maior entre Aécio e Dilma, que até então estavam mais bem colocados nas pesquisas⁷ e uma menor concentração desta disputa com Marina Silva.

Na tabela a seguir (Tabela 2) apresentamos os dados gerais das redes coletadas. Os nós com maior centralidade são apresentados. Optamos por colocar os dez itens com maior valor dentro de cada métrica, para contextualizar e os candidatos, caso estes não estivessem dentro dos mais centrais. Nenhum candidato ou conta de campanha apresentou outdegree relevante.

Tabela 2: Dados Gerais do #DebateDaBand

	Rede 1 (23:35)	Número	Rede 2 (00:36)	Número
Nós com maior indegree	diimabr	851	oglobopolitica	489
	silva_marina	668	kibeloco	408
	hugogloss	407	jeanwyllys_real	381
	marcelotas	363	andrezinhop4	338
	kibeloco	339	aecioneves	213
	aecioneves	323	josiane_pmoraes	176
	jooseanee	316	miltonneves	170
	queremosaecio	227	mallandrosergio	150
	gordinhooficial	221	joseanee	117
	psdbparana	189	silva_marina	109
	dilmabr	170	dilmabr	74
	lucianagenro	101	lucianagenro	11
	eduardojorge43	38	eduardojorge43	9
	levyfidelix	17	levyfidelix	7
	everaldo_20	18	everaldo_20	5
Nós com maior centralidade betweenness	diimabr	13.777.692,23	oglobopolitica	4.993.547,891
	silva_marina	12.402.328,76	aecioneves	3.387.307,653
	hugogloss	5.669.266,76	jeanwyllys_real	3.297.258,298
	aecioneves	5.588.911,419	kibeloco	3.211.095,346
	marcelotas	5.474.887,93	andrezinhop4	1.516.155394
	kibeloco	5.342.207,577	miltonneves	1.347.356,74
	jooseanee	4.967.467,811	silva_marina	1.025.230,204
	queremosaecio	2.984.559,315	jooseanee	989.301,675
	dilmabr	2.443.521,167	comentv	921.954,975
	dilmarouselff	2.195.434,539	dilmabr	643.168,370
	lucianagenro	1.304.898,972	lucianagenro	85.835,981
	eduardojorge	320.056,394	eduardojorge43	75.926,00
	everaldo_20	121.547,404	everaldo_20	28.351,169
	levyfidelix	107.736,552	levyfidelix	14,00
Nós com maior centralidade eigenvector (apenas dos candidatos e contas de apoio)	aecioneves	0,025349	aecioneves	0,011685
	maisaecio	0,023713		
	rede 45	0,002528		

⁷ DataFolha de 18/08 - <http://g1.globo.com/politica/eleicoes/2014/noticia/2014/08/dilma-tem-36-marina-21-e-aecio-20-diz-pesquisa-datafolha.html> e a pesquisa que saiu na véspera do debate, já apontando Marina bem mais colocada na disputa: <http://g1.globo.com/politica/eleicoes/2014/noticia/2014/08/pesquisa-ibope-mostra-dilma-com-34-e-marina-silva-com-29.html>

Na tabela observamos que no início do debate, a candidata Marina Silva é a mais citada e com maior centralidade betweenness, seguida pelo candidato Aécio Neves. Dilma vem a seguir e depois, Luciana Genro. Eduardo Jorge, Pastor Everaldo e Levy Fidélis têm pouquíssimas citações. Entretanto, ao observarmos a centralidade eigenvector, vemos que o candidato Aécio é o único a aparecer, juntamente com contas de apoio, indicando possivelmente uma articulação entre essas contas. Ao observarmos também a centralidade betweenness, vemos que é Marina quem aparece com maior número de pontes, sendo citada junto a grupos diferentes. Logo a seguir, temos Aécio e Dilma, mas ambos com uma centralidade bem inferior.

Neste primeiro grafo (Figura 1) vemos a rede do primeiro momento do debate articulada em torno do indegree e plotada através do algoritmo Force Atlas 2 (Jacomy et al., 2014), que é um algoritmo dirigido por força e que promove a aproximação dos nós por clusters e seu afastamento de acordo com as conexões (onde os clusters funcionam como força gravitacional).

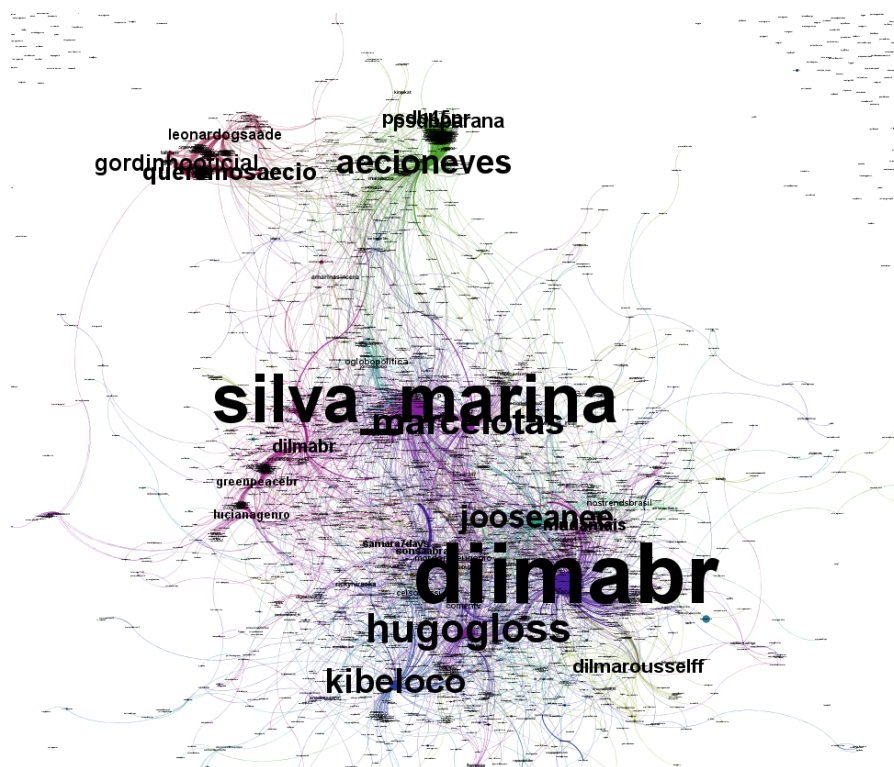


Figura 1: Grafo do primeiro momento no debate da band.

Vemos, nesta primeira imagem, que enquanto Marina Silva situa-se mais ao centro do grafo, próxima de Dilma Rousseff e Luciana Genro, Aécio Neves distancia-se, aproximando-se de outros clusters (acima), indicando de forma mais clara que o candidato é citado em contextos diferentes e por contas diferentes dos demais. Esse afastamento, bem como os dados anteriores, reforçam a hipótese de que o candidato Aécio aparece com centralidades altas pela ação de sua militância principalmente, mais do que pelas citações orgânicas com relação ao debate. Além disso, seu afastamento também implica em clusterização com nós diferentes, enquanto os demais candidatos, com citações mais orgânicas, aparecem

próximos uns dos outros (condizendo com a narrativa do debate). O grafo está colorido por modularidade, indicando vizinhança (tendência de nós com vizinhos iguais ficarem no mesmo grupo).

Ao observarmos o gráfico a seguir (Figura 2), construído por grupos dentro do algoritmo de Clauset-Newman-Moore⁸ e separados os clusters, temos outro desenho que auxilia a compreender melhor o momento. Em verde, temos a candidata Marina, em azul, Aécio e no cluster laranja, os demais candidatos.

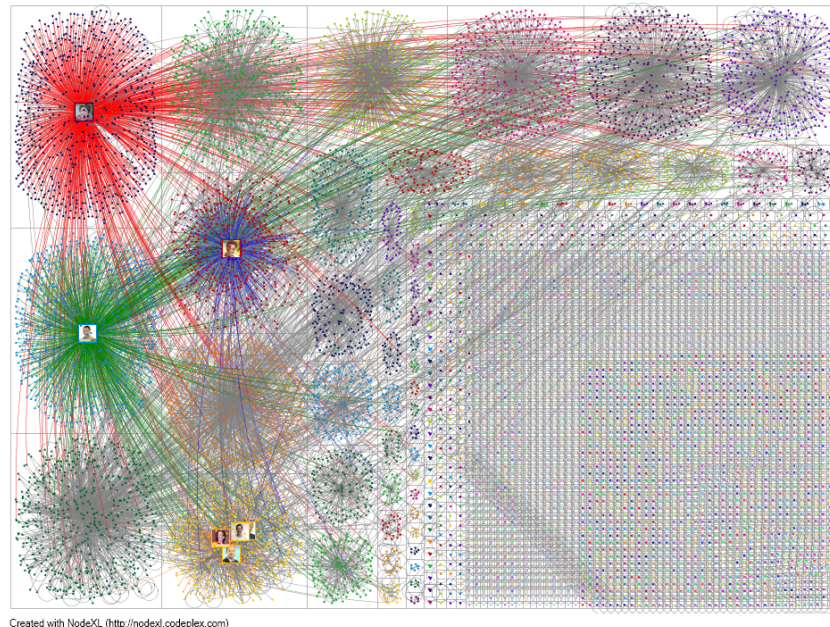


Figura 2: Grupos

É relevante observarmos também a importância das contas fake de apoio à candidata (e presidenta) Dilma, como @diimabr (a "Dilma Bolada") e @dilmarousselff (a "Dilminha"), que também aparecem com boa centralidade betweenness e alto indegree fora do cluster de @dilmabr, atingindo pessoas diferentes e repercutindo em grupos diferentes.

No segundo momento do debate já temos dados diferentes (Tabela 2). O candidato Aécio sobe e todos os demais caem. É relevante observar que a centralidade betweenness dele e de Marina são as mais altas entre os candidatos indicando provavelmente um crescimento de citações dos candidatos em grupos diferentes em relação aos demais (mas uma possível perda de articulação que existia no início do debate, já que esses números diminuíram). Quando vemos os grafos (Figura 3), há também outro desenho. Novamente, o grafo mostra o candidato Aécio Neves mais afastado, com Marina Silva também um pouco mais afastada do centro. Outros atores ganham maior centralidade. Curiosamente, o candidato a deputado do PSOL Jean Wyllys torna-se mais central que a própria candidata ao governo do partido, Luciana Genro.

No segundo grafo, também separado por cluster de acordo com o algoritmo de Clauset-Newman-Moore, vemos novamente que os candidatos aparecem mais em um contexto conjunto (com exceção de Eduardo Jorge, que está separado). Entretanto, ainda é possível ver a articulação em torno do candidato

⁸ Esse algoritmo de visualização é baseado em uma medida de modularidade que leva em conta a quantidade de conexões existentes dentro de um determinado cluster e o número possível dessas conexões para separar os clusters por modularidade (Clauset, Newman & Moore, 2004).

Aécio, com o conjunto de conexões azuis que o liga a contas de campanha que repercutem seu nome (Figura 3).

É importante observar nesses dados, também, a presença de uma única conta jornalística (oglobopolítica) e de uma grande quantidade de contas humorísticas (como as de Dilma, @kibeloco e etc.), indicando a relevância deste tipo de narrativa na conversação. Vemos, assim, que neste debate, a posição central foi inicialmente ocupada por Marina Silva, logo a seguir sendo ultrapassada por Aécio. É importante observar, também, a articulação em torno do candidato Aécio e os perfis falsos que acabam também popularizando a conta de Dilma. Entre os candidatos menos falados, Luciana Genro manteve-se em quarto lugar, com Eduardo Jorge a seguir.

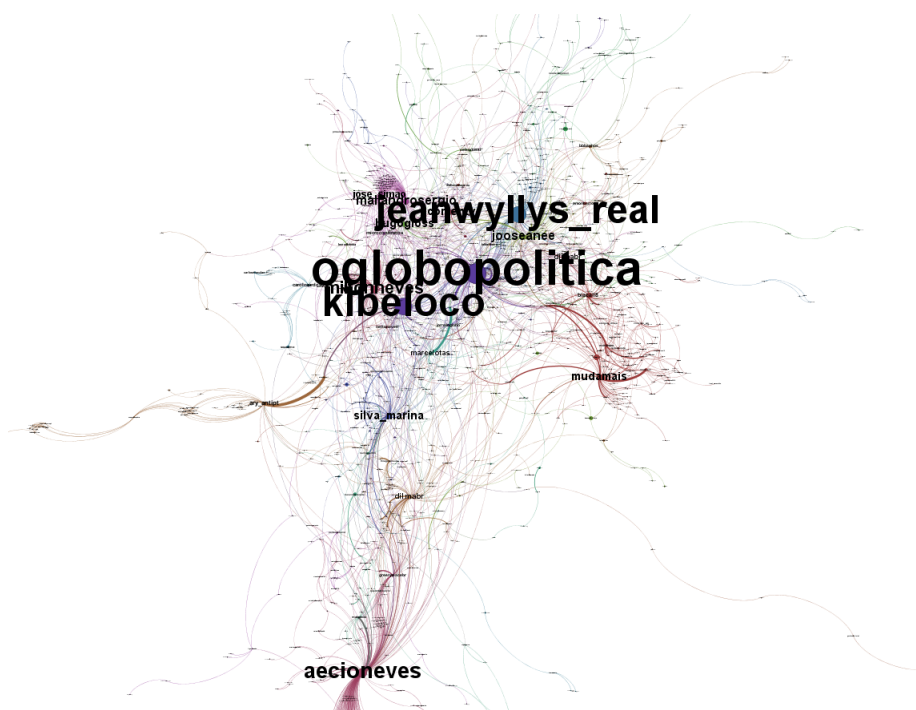


Figura 3: Grafo do segundo momento do #debatedaband.

O segundo debate aconteceu nas dependências do SBT (Sistema Brasileiro de Televisão) e foi transmitido ao vivo pelo SBT, pela UOL e pela Folha de São Paulo online. Iniciou às 17h45min do dia 01 de setembro de 2014, uma terça-feira e terminou por volta de 20h da mesma data. Estiveram no debate os mesmos candidatos do debate anterior. O debate teve seus primeiros blocos concentrados em um duelo mais acirrado entre as candidatas Dilma Rousseff e Marina Silva, provavelmente por conta da pesquisa do DataFolha⁹, divulgada alguns dias antes, que apontava esta última como um risco real à reeleição de Dilma, com um empate entre as duas e um crescimento grande das intenções de voto na última, além de uma queda de Aécio. Os últimos blocos foram caracterizados por críticas mais acirradas à Dilma por parte de Aécio e Marina além de críticas à Marina, por todos os candidatos. O candidato Eduardo Jorge repercutiu muito por seus posicionamentos e atitudes durante o debate, roubando um pouco o palco

⁹ <http://www1.folha.uol.com.br/poder/2014/08/1508270-marina-empata-com-dilma-na-corrida-presidencial-diz-datafolha.shtml>

principal em algumas participações¹⁰, Levy Fidelix e Luciana fizeram uma participação semelhante à do primeiro debate.

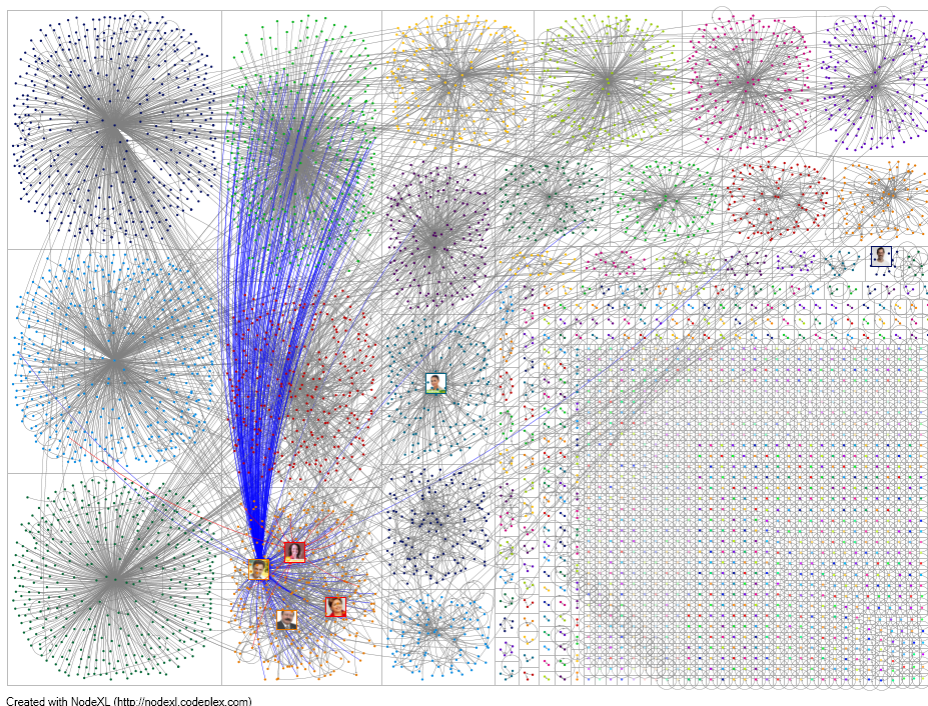


Figure 4: Grupos do segundo momento do #debatedaband

6.2 O Segundo Debate

Como no debate anterior, utilizamos o procedimento de crawling do NodeXL, obtendo os dados das redes coletadas às 18h45min (meio do debate) e às 20h (final do debate). Os dados são sumarizados a seguir (Tabela 3).

Tabela 3: Dados Gerais do #DebatedoSBT

	Rede 1 (18:45)	Número	Rede 2 (20:00)	
Nós com maior indegree	diimarousselff	765	william_castro	1346
	diimabr	762	silva_marina	558
	aecioneves	449	rafinhabastos	466
	maisaecio	358	diimabr	292
	eduardopcardoso	339	diimarousselff	193
	joaquimtdasilva	193	aecioneves	190
	nostrendsbrasil	178	dacostaneusa	177
	walteerfelix	163	ivaniaaziz	170
	william_castro	158	jooseanee	155
	rede45	148	diimabr	144
	diimabr	124	levyfelix	111
	silva_marina	109	lucianagenro	60
	lucianagenro	73	everaldo_20	19
	everaldo_20	29	eduardojorge43	15
	levyfelix	14		
	eduardojorge43	7		
Nós com maior centralidade	diimabr	7.585.658,256	silva_marina	7.276.268,225
	diimarousselff	6.829.852,311	rafinhabastos	5.835.159,29
	aecioneves	2.410.933,143	diimabr	3.608.187,751

¹⁰ A íntegra do debate está disponível aqui - <http://www.sbt.com.br/jornalismo/noticias/44352/Assista-na-integra-ao-debate-entre-os-candidatos-a-Presidencia-.html#.VAhvl2RdU0Y>

betweenness	rede45	1.688.704,421	aecioneves	2.415.509,889
	maisaecio	1.450.405,368	dilmarousselff	2.204.489,026
	dilmabr	1.162.560,183	dilmabr	1.426.666,127
	silva_marina	827.035,2512	aecioblog	1.137.863,483
	aecioblog	681.406,9495	levyfidelix	1.059.050,8
	lucianagenro	616.456,610	lucianagenro	825.496,6513
	everaldo_20	112.467,183	eduardojorge43	125.996,2999
	levyfidelix	79.336,257	everaldo_20	68.423,237
Nós com maior centralidade eigenvector	eduardojorge43	41.959,546		
	aecioneves	0,025349	Nenhum candidato apresentou dados significativos	
	maisaecio	0,023713		
	rede 45	0,002528		

No primeiro mapeamento deste debate, observamos o indegree dos perfis dos candidatos. Vemos, assim, que o candidato com maior indegree é o candidato Aécio Neves. A seguir, situam-se as candidatas Dilma e Marina. Isso indica que o candidato foi o mais citado por outras contas durante o primeiro momento do debate. Ao observarmos, entretanto, as métricas seguintes, de centralidade betweenness e eigenvector, vemos também que Aécio tem uma alta centralidade, bastante superior a dos demais candidatos. As candidatas Dilma e Marina aparecem entre os maiores nós com centralidade betweenness, mas bem abaixo de Aécio. A primeira vista, parece que o candidato Aécio Neves foi aquele que dominou a conversação na rede. Entretanto, ao observarmos com maior cuidado, percebemos que suas contas de campanha também aparecem em todas as métricas. Há, aqui, um forte indício da ação de um único cluster entre si, fazendo com o que o candidato apareça entre os mais citados. Ao observarmos o grafo, essa situação fica mais clara.

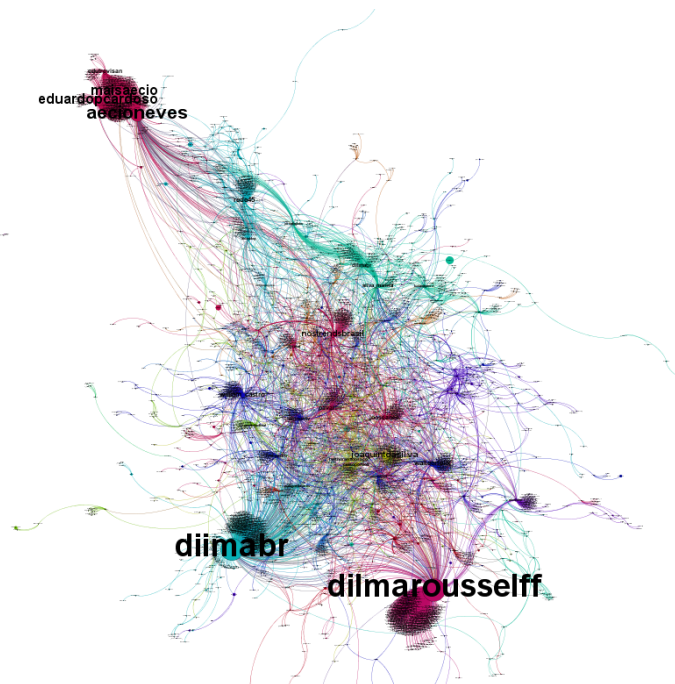
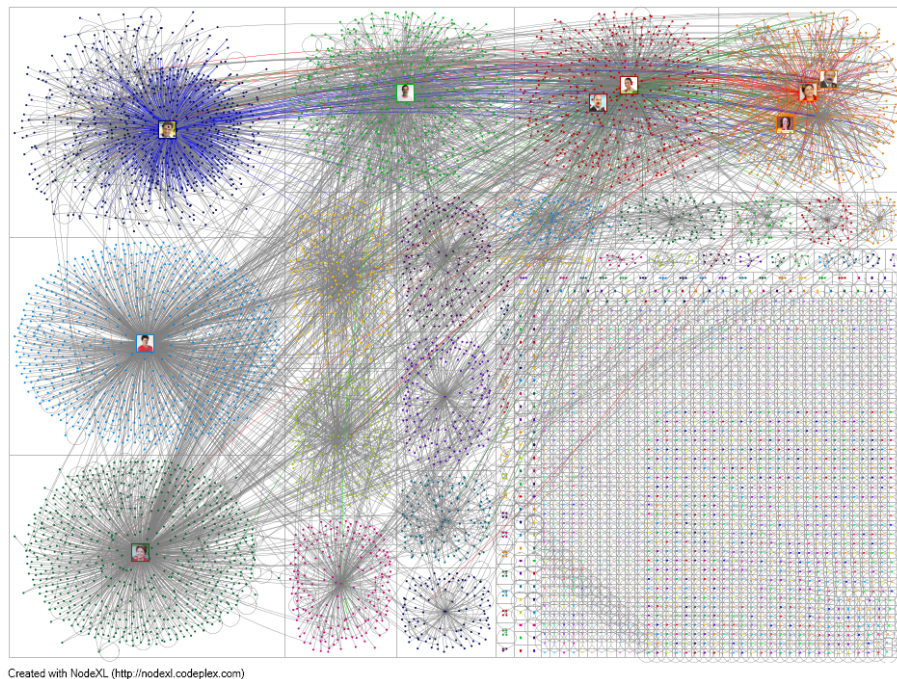


Figure 5: Mapa do segundo momento do debate

Observamos um cluster em torno do nome do candidato Aécio, mostrando com maior clareza que há uma articulação forte em torno de seu nome (em vermelho, acima). Já as candidatas Dilma, Marina aparecem de forma mais orgânica na rede, sendo citadas por atores diferentes, ao centro, dentro do cluster verde claro. Assim, no caso de Aécio, vemos várias contas de apoio que também aparecem com

um indegree e betweenness alto, indicando possivelmente uma estratégia de campanha para elevar a visibilidade do candidato. É importante observar, também e como no caso do debate anterior, a presença forte das contas de humor, e especialmente, das contas que apoiam Dilma, repercutindo em grupos onde o nome oficial não aparece.

No segundo mapa, a seguir (Figura 4), vemos os grupos separados por clusters (Clusset-Newman-Moore). Os grupos indicam quais contas estão mais próximas de quais candidatos, pelas citações e retweets. Vemos as conta de Dilma, Levy e Luciana num mesmo cluster (laranja), indicando comentários à sua participação no bloco feitos por várias contas, de forma semelhante ao grafo anterior. Ou seja, mais tweets/contas que citaram Dilma citaram também os outros dois candidatos. A conta de Marina aparece mais próxima do Pastor Everaldo (cluster vermelho), também indicando a ocorrência de ambos em conjunto nos tweets. Eduardo Jorge (cluster verde) e Aécio (cluster azul) aparecem em outros clusters, indicando que aqueles que citaram esses candidatos pouco citaram os demais. O cluster do candidato Aécio é o mais fechado, com menor número de citações externas para outros clusters.



Created with NodeXL (<http://nodexl.codeplex.com>)

Figure 6: Grupos do primeiro momento do debate no SBT.

No segundo momento, ao final do debate, já vemos outro desenho. Na tabela 3, vemos um crescimento da candidata Marina, em indegree e betweenness, e coloca-se mais ao centro do grafo (figura Aécio Neves cai, apesar de ainda ocupar posição central e fica mais próximo dos candidatos Levy Fidélis e Dilma Rousseff, que crescem). Na figura a seguir, vemos o grafo geral por indegree. Observamos também que, de forma diferente do grafo anterior, há menos centralização e maior densidade. Isso indica uma maior participação das pessoas, bem como outras contas sendo protagonistas na narrativa.

Ao observarmos os clusters por grupos, na Figura 6, vemos contextos de conversa diferentes. Enquanto os clusters maiores referem-se à tweets humorísticos narrando o debate dos usuários @william_castro e @rafinhabastsos, vemos que os três candidatos mais centrais estão no mesmo cluster, porque foram citados pelos mesmos usuários (provavelmente aqueles que narravam o debate) e também

foram citados em conjunto. Já os demais candidatos, Luciana, Levy e Eduardo aparecem separados em outros clusters, sendo citados por grupos diferentes.

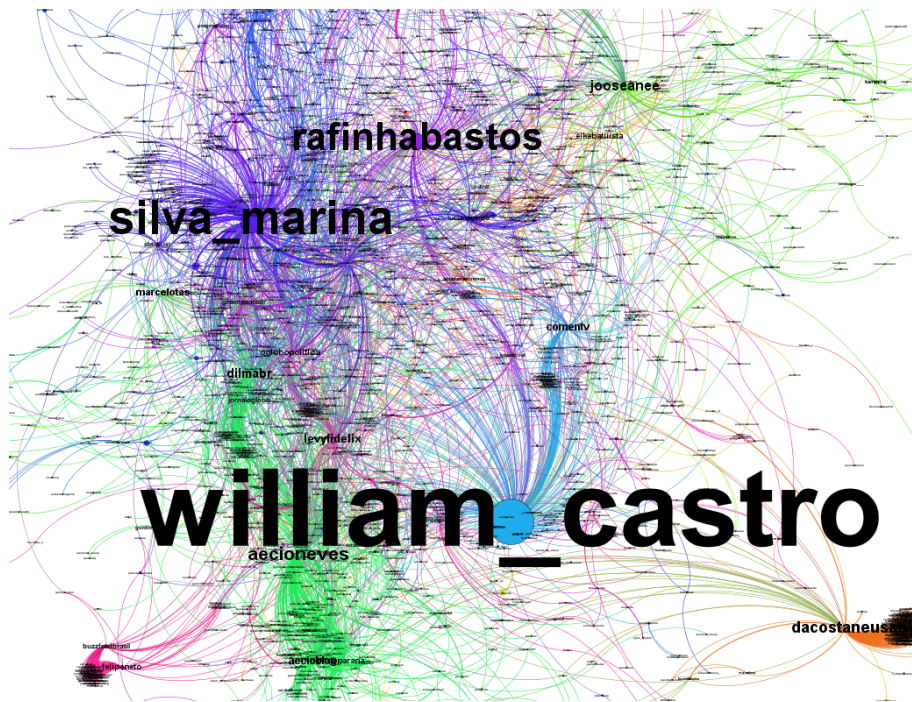
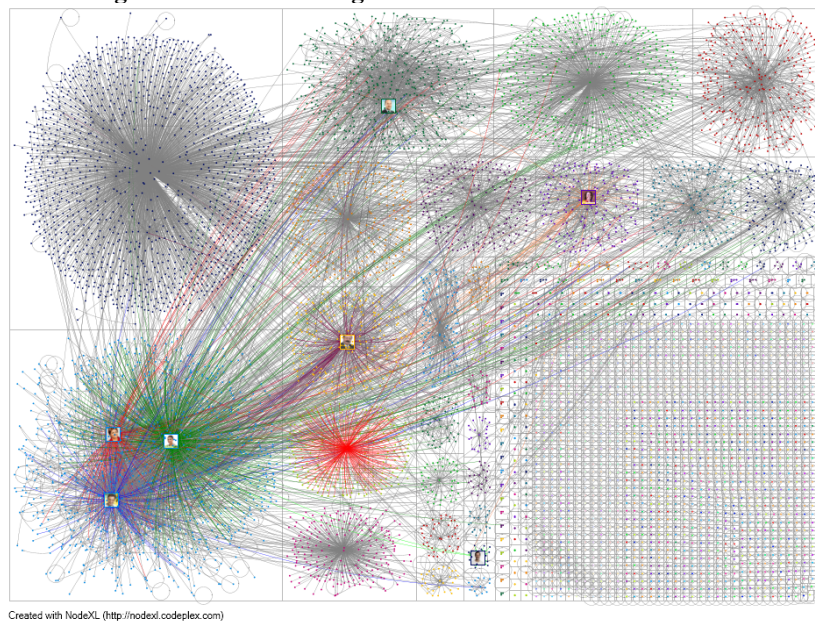


Figura 7: Cluster do segundo momento do debate no SBT



Created with NodeXL (<http://nodexl.codeplex.com>)

Figure 8: Grupos do segundo momento do debate no sbt.

Conforme vimos até aqui, na primeira medida do debate, o candidato Aécio Neves ocupava uma posição mais central na discussão, tanto por ação de militância, quanto por ação de contas de campanha. No segundo momento, essa centralidade passa a ser dominada por Marina Silva, que se torna a candidata mais citada. Também é relevante observar que os três candidatos tornam-se mais próximos no decorrer do debate, provavelmente aparecendo em contextos de narrativas do debate em si. Entretanto, é importante fazer um aparte: Apesar da candidata Dilma não aparecer com tanto foco em nenhuma das análises, os

perfis falsos que a divulgam, @diimabr (a "Dilma Bolada") e @dilmarousselff (a "Dilminha") ocupam posições centrais e com *indegree* alto em ambas as redes. Isso significa que, apesar da conta da presidenta não estar tão central no grafo, há um espalhamento de sua "persona" por outros grupos, notadamente aqueles conectados às contas de humor. Essa visibilidade pode fazer diferença, na medida em que chega a pontos da rede diferentes daqueles da conta oficial da candidata. Luciana Genro perde um pouco de espaço na segunda metade e Eduardo Jorge cresce na segunda coleta.

7. Conclusões: Discutindo Estratégias e os Candidatos na Midia Social

A partir dos dados coletados, é possível traçar alguns apanhados estratégicos a respeito de como estão sendo citados e em quais contextos e suas estratégias de campanha. Observamos, por exemplo, que alguns candidatos têm citações mais orgânicas (o que não quer dizer que não tenham militância, mas que a militância não atua em torno do seu nome especificamente ou não atua com a hashtag do debate), enquanto outros, como Aécio Neves, têm citações consistentes e focadas em um mesmo grupo de atores (bots ou militância) que cita apenas sua conta.

Conforme Newman (2003) e Freeman (1979), a centralidade *betweenness* é capaz de identificar influência entre diferentes esferas de grupos. Vimos que, entre os candidatos esse índice, destacam-se, no primeiro debate, Marina Silva e Aécio, seguidos, com um número inferior, por Dilma. No segundo debate, temos Aécio e Dilma, com Marina em terceiro no início, mudando depois para Aécio, Marina e Dilma. Isso significa que os dois candidatos, Aécio e Marina, tendem a atuar mais como pontes do que Dilma, ou seja, ser citados em tweets pertencentes a atores que não estão próximos entre si, o que indicaria um apoio mais difuso, orgânico, pois os nós são mais eficientes em unir diferentes grupos (Hansen, Shneiderman & Smith, 2011). Entretanto, ao mesmo tempo, também observamos que o candidato que aparece em citações mais orgânicas é Dilma, quando observamos os grupos por modularidade. É ela, também, que aparece mais próxima dos demais, indicando contextos de citação onde os demais candidatos também aparecem (narrativas do debate).

Vimos também que o único candidato a aparecer com centralidade *eigenvector* (Bonacich 1972) relevante é Aécio. A campanha de Aécio Neves, parece utilizar um conjunto de contas que apoiam e atuam narrando a presença do candidato nos debates, citando-se e retuitando-se entre si também. Com isso, as contas de campanha têm um *indegree* alto, contribuindo diretamente para o aumento da centralidade *eigenvector* da conta oficial. Aqui, a métrica pode representar prestígio (Rusinowska et. al, 2011), mas especialmente entre um grupo específico de atores que também são atuantes nas redes - no caso, as contas de campanha e apoiadores. Sendo intencional ou não, a atuação desses atores contribui para elevar as demais métricas de citação, posicionando Aécio melhor durante o debate e tornando-o mais visível para as audiências.

Marina Silva também tende a aparecer separada dos demais candidatos, em um contexto onde as contas não falam das demais. Sua centralidade parecer ser mais orgânica, ou seja, há menos militância e mais conversação em torno de seu nome, ao contrário do caso anterior. No primeiro debate, Marina começa muito citada e com alta centralidade *betweenness* (o que é esperado devido à circunstância da indicação de seu nome), logo a seguir, no segundo momento, caindo bastante. No segundo debate, Marina

começa discreta e cresce, também tanto em citações quanto em centralidade. Ambos os dados indicam uma maior oscilação da candidata, ao contrário dos candidatos anteriores, provavelmente mais relacionado à organicidade de suas relações. Observamos também que modularmente Marina também tende a aparecer próxima de outros atores, indicando também seu contexto de narrativa do debate. Ela desloca-se sempre para grupos próximos a outros candidatos (clusters), indicando aí contextos de participação onde repercute mais.

A exceção a esse tipo de presença extrema é a conta de Dilma, que aparece sempre mais conectada aos demais candidatos, possivelmente aparecendo de forma mais comparativa, já que é a atual presidenta e a candidata a ser batida (que está em primeiro lugar nas pesquisas). Apesar da presença discreta da conta oficial da candidata, entretanto, é essencial observar a atuação central das contas fakes que trazem um perfil positivo da candidata, que obtêm grande visibilidade durante o debate (@diimabr e @dilmaroussellff). Essas contas atingem grupos que a militância e as narrativas dos debates não atingem (aparecendo em clusters separados). Além disso, essas contas estão dentro do discurso humorístico que parece ganhar bastante espaço durante os eventos.

Luciana permanece sempre em quarto lugar em termos de citações e centralidade. Em alguns casos, entretanto, seus apoiadores aparecem muito mais do que ela, como é o caso do deputado Jean Wyllys, do PSOL, que narrou o debate da Band obtendo mais destaque do que a candidata. Entretanto, como no caso de Dilma, tal narrativa é relevante, pois demonstra apoio à candidata, repercutindo em redes onde seu nome não apareceu com tanta força. Também é importante observar a presença de Eduardo Jorge no primeiro debate e a queda de visibilidade do candidato no segundo, sendo suplantado por Levy Fidélis e Pastor Everaldo. Esses candidatos, que tendem a aparecer e repercutir em grupos separados dos candidatos com maior apoio também aparecem, mas também de forma menor. Levy adquire maior centralidade no segundo debate, mas ainda de forma discreta em relação aos candidatos com maior intenção de voto.

Como vimos, as características dos "públicos em rede" (boyd, 2010), dão à mídia social características importantes de reverberação dentro de elementos semelhantes àqueles de esfera pública. Assim, o uso de ferramentas para coleta e análise de dados, bem como de métricas para a compreensão da posição dos sujeitos na rede é essencial para que se estude essas reverberações. Neste trabalho, procuramos discutir métricas de análise de redes sociais nos comentários sobre os debates no Twitter, buscando compreender como aparecem os candidatos e que informações esses elementos podem nos trazer sobre a participação política na internet e seus impactos nas eleições presidenciais. Há, é claro, limites, principalmente com relação ao conjunto de dados coletados (apenas pessoas que utilizaram as hashtags, ou seja, há uma parcialidade quanto a isso), bem como elementos relacionados à coleta de dados (limites da API do Twitter e forma de coleta). Ainda assim, acreditamos que os resultados são importantes para a discussão.

Referências

- BACKSTROM, L. ; BOLDI, P.; ROSA, M.; UGHANDER, J.; VIGNA, S. Four degrees of separation. **Proceedings of WebSci '12 Proceedings of the 4th Annual ACM Web Science Conference**. New York: ACM, 2012. **Pages 33-42**
- BOYD, d.; ELLISON, N. (2007) 'Social network sites: Definition, history, and scholarship', *Journal of Computer-Mediated Communication* Vol. 13, Issue 1, p210-230, Blackwell. <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full>
- BOYD, d. (2010). "Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications." In *Networked Self: Identity, Community, and Culture on Social Network Sites* (ed. Zizi Papacharissi), pp. 39-58.
- BONACICH, P. [Factoring and weighting approaches to status scores and clique identification](#). *Journal of Mathematical Sociology*, V. 2, N. 1, 1972. PP. 113-120. Taylor & Francis Group
- CHEONG, F. e CHEONG, C. "Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods" (2011). *PACIS 2011 Proceedings*. Paper 46. <http://aisel.aisnet.org/pacis2011/46>
- CLAUSET, A. NEWMAN, M. E. J.; MOORE, C. Finding community structure in very large networks. *American Physical Society, Physical Review*, vol 70, 6., 2014. 10.1103/PhysRevE.70.066111, <http://link.aps.org/doi/10.1103/PhysRevE.70.066111>
- DEGENNE, A. e FORSE M., *Introducing Social Networks*. London: SAGE Publications Ltd; 1 edition, 1999.
- FREEMAN, L. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1978/79.
- FREEMAN, L. The development of social network analysis. **Vancouver: Empirical Press, 2004.**
- GARTON, L.; HAYTHORNTWAITE, C.; WELLMAN, B. Studying Online Social Networks. *Journal of Computer-Mediated Communication*, Volume 3, Issue 1, page 0, June 1997. <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1997.tb00062.x/abstract>
- GIL DE ZUÑIGA, H.; JUNG, N.; VALENZUELA, S. Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, 17: 319–336, 2012. doi: 10.1111/j.1083-6101.2012.01574.x
- GOMES, Wilson. Participação política online: questões e hipóteses de trabalho. In: GOMES, Wilson; MAIA, Rousiley C. M.; MARQUES, Francisco J. A. *Internet e Participação Política no Brasil*. Porto Alegre: Sulina, 2011.
- Hansen, Derek L.; Schneiderman, Ben; Smith, Marc (2010), *Analyzing social media networks with NodeXL: Insights from a Connected World*, Morgan Kaufmann. New York: Elsevier, 2011.
- JUNGHER, A.; JÜRGENS, P.; SCHOEN, H. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment". *Social Science Computer Review* May 2012 30: 229-234, first published on April 25, 2011 doi:10.1177/0894439311404119
- KUSHIN, M. J.; YAMAMOTO, M. Did Social Media Really Matter? College Students' Use of Online Media and Political Decision Making in the 2008 Election. *Mass Communication and Society* Vol. 13, Iss. 5, 2010 <http://www.tandfonline.com/doi/full/10.1080/VBbjj5dU0Y>

MARQUES, Francisco J. A. **Participação, Instituições Políticas e Internet: Um exame dos canais participativos presentes nos portais da Câmara e da Presidência do Brasil.** In: GOMES, Wilson; MAIA, Rousiley C. M.; MARQUES, Francisco J. A. Internet e Participação Política no Brasil. Porto Alegre: Sulina, 2011.

MISLOVE, A., MARCON, M.; GUMMADI, K.; DRUSCHEL, P. **Measurement and analysis of online social networks.** Proceedings of IMC '07 Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, New York, NY. (pp. 29-42)

NEWMAN, J. **The structure and function of complex networks.** 2003. Disponível em: <http://arxiv.org/pdf/cond-mat/0303516.pdf>

RECUERO, R. **Redes Sociais na Internet.** Porto Alegre: Sulina, 2009.

RECUERO, R. **A Conversação em Rede.** Porto Alegre: Sulina, 2012.

RECUERO, R. e ZAGO, G. Em busca das "redes que importam": Redes Sociais e Capital Social no Twitter. **Revista Líbero**, v. 12, n. 24, pp. 81-94, 2009. Disponível em: <http://casperlibero.edu.br/wp-content/uploads/2014/05/Em-busca-das-%E2%80%99Credes-que-importam%E2%80%99D.pdf>>. Acesso em 13/10/2014.

SCOTT, J. **Social Network Analysis: A handbook.** New York: SAGE Publications Ltd; 2nd edition, 2000.

VENTURINI, J. M.; VENTURINI, T., HEYMANN, S; BASTIAN, M. **ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.** PLoS ONE 9(6): 2014. e98679. doi:10.1371/journal.pone.0098679

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications.** 1.ed. Cambridge: Cambridge University Press, 1994.