

Combinando Interações de Endosso e Comunicação em Redes Sociais Multipolarizadas

Pedro H. Calais Guerra, Wagner Meira Jr.

¹Departamento de Ciência da Computação - Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, MG, Brazil

{pcalais,meira}@dcc.ufmg.br

Abstract. *Polarized social networks arise from domains where people have two conflicting and opposing viewpoints regarding an issue, such as abortion, gun control and same-sex marriage. However, in many real scenarios, multiple positions can be adopted in relation to a topic, as in multipartisan political systems and sports competitions. In this work, we show that multipolarized social networks unveil inconsistencies and simplistic assumptions on models that deal with bipolarized networks. In particular, we find that the proximity of a pair of users in the social network does not necessarily implies sharing of viewpoints. Moreover, we propose a simple model that capture relationships of support, antagonism and indifference which can be seen on multipolarized networks, by combining endorsement and communication interactions among users.*

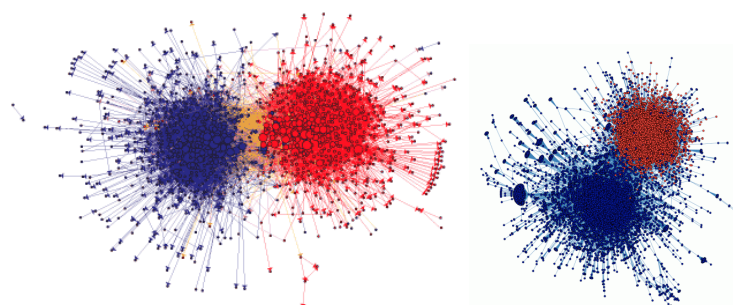
Resumo. *Redes sociais polarizadas, caracterizadas por intermediar embates de ideias conflitantes, vêm sendo extensamente estudadas no contexto de debates de temas polêmicos, como aborto, controle de armas e casamento entre pessoas do mesmo sexo. Classicamente, neste cenário, dois lados com opiniões opostas se contrastam. No entanto, em vários cenários reais, múltiplas posições podem ser assumidas, como no caso de sistemas políticos multipartidários ou de disputas esportivas. Nesta pesquisa, mostramos que redes sociais multipolarizadas revelam sutilezas e inconsistências nos modelos de análise de redes bipolarizadas. Em particular, mostramos que a proximidade de um usuário com outro na rede social não reflete, necessariamente, similaridade de pontos de vista. Além disso propomos um modelo simples que captura as relações e apoio, antagonismo e indiferença que se manifestam em redes multipolarizadas a partir da combinação de interações de endosso e comunicação por meio das quais os usuários interagem.*

1. Introdução

Polarização é o processo social por meio do qual um grupo se divide em dois (ou mais) sub-grupos com ideias, objetivos e pontos de vista conflitantes, e que, ao passar do tempo, gera cada vez menos membros com opiniões neutras [Sunstein 2002, Isenberg 1986]. Observamos debates polarizados em uma gama de tópicos de interesse para a sociedade, principalmente relacionados à Política e Políticas Públicas, como casamento entre pessoas do mesmo sexo, controle de armas, aborto e aquecimento global [McCright and Dunlap 2011, Mouw and Sobel 2001, Hunter 1992]. O entendimento de como esses debates são travados nos ambientes *online* é importante tanto do ponto de vista sociológico, já que polarização induz uma segregação

na sociedade [Paul DiMaggio 1996], como pelas implicações em algoritmos que vissem detectar comunidades e mesmo a polaridade das opiniões expressadas neste contexto [Tan et al. 2011a].

A literatura de sociologia e ciência da computação têm se concentrado em estudos e modelos que interpretem opiniões expressadas no caso clássico de polarização, em que exatamente dois lados se opõem, posicionando-se contra ou a favor a uma ideia ou proposta. De fato, ao se manifestar nas redes sociais, usuários que opinam e concordam com outros usuários com opinião semelhantes em geral se agrupam em duas comunidades segregadas, como, por exemplo, nas comunidades de republicanos e democratas encontradas em blogs políticos e no Twitter, como ilustrado na figura 1.



(a) Rede social de blogs sobre política - (b) Rede Social de Eleições Presidenciais Norte-Americanas perfis do Twitter sobre política nos Estados Unidos em 2004 [Adamic and Glance 2005] [Conover et al. 2011]

Figura 1. Dois grafos sociais que apresentam comunidades bipolarizadas. Na Figura 1(a), os nós representam blogs sobre temas políticos durante as Eleições Presidenciais nos Estados Unidos, em 2004. A Figura 1(b) mostra, no Twitter, democratas e republicanos divididos em dois grupos e conectados por *retweets*.

No entanto, em muitos cenários reais, há mais do que dois lados em conflito. É o caso, por exemplo, de sistemas políticos multipartidários (como no Brasil), competições esportivas (32 países disputam uma Copa do Mundo; mais de 200 países disputam as Olimpíadas) e programas de TV (no Big Brother Brasil, por exemplo, torcidas podem se formar em torno de cada um dos 16 participantes). Nesses cenários, os quais chamamos de **multipolarizados**, começamos a observar relações mais complexas entre os lados do debate, ao invés do dualismo *apoio versus antagonismo*. No caso bipolarizado, o fato de um indivíduo pertencer a um grupo implica, em geral, que ele é antagônico ao outro grupo. Isso ocorre como consequência da segregação gerada por temas polêmicas como aborto (contra ou a favor), casamento homossexual (contra ou a favor), e posição política em sistemas bipartidários (democrata ou republicano). Quando há mais de dois lados de debate, em que um indivíduo pode se posicionar em torno de K “lados” da discussão ($K > 2$), a identificação de um indivíduo com um grupo não implica necessariamente no antagonismo dele com relação a **todos** os $K - 1$ grupos restantes (ou, pelo menos, na mesma intensidade). Ele pode ser, inclusive, indiferente, ou neutro, a um subconjunto desses grupos, e pode apoiar mais de um grupo simultaneamente. Esse cenário pode ser instanciado, por exemplo, nas comunidades formadas ao redor de torcidas de futebol, como ilustrado na figura 2. Mais ainda, em cenários multipolarizados as relações entre as comunidades tende a ser menos estática: por exemplo, uma disputa eleitoral com 3 candidatos pode

fazer com que um candidato migre de uma posição de neutralidade ou antagonismo para uma posição de apoio ao grupo que representa um dos outros candidatos, após o primeiro turno das eleições. No cenário esportivo, essa dinâmica acontece quando, embora haja rivalidades históricas pré-definidas, uma equipe depende do resultado de outras para que mantenha chances de título ou de classificação, por exemplo.

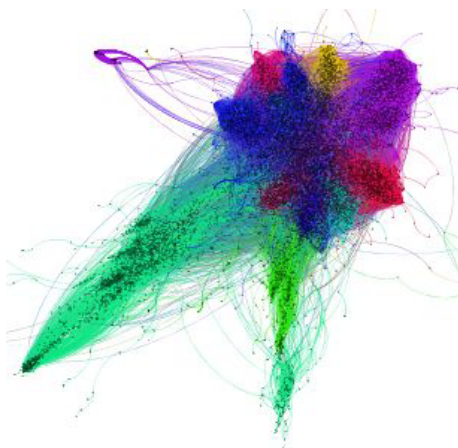


Figura 2. Rede multipolarizada obtida a partir de dados do Twitter. As comunidades representam torcidas das 12 principais equipes do futebol brasileiro.

Neste artigo, buscamos cumprir dois objetivos principais:

1. Demonstramos que contextos de discussão multipolarizados revelam sutilezas e inconsistências que são “escondidas” nas tradicionais redes bipolarizadas em que a literatura atual se concentra. Como consequência dos diferentes tipos de relação entre diferentes grupos que defendem seu lado em um debate, descobrimos até o momento que fãs de clubes de futebol rivais no Brasil são *mais próximos* entre si em uma rede de *retweets* do Twitter do que em relação a outros times. Este paradoxo, o qual chamamos de “*Paradoxo dos Rivais*”, acontece porque os modelos atuais lidam bem com a bipolarização, mas não com a multipolarização.
2. Propomos uma estratégia de mineração de redes sociais capaz de capturar relações de apoio, antagonismo e indiferença em redes multipolarizadas. Nosso modelo combina dois tipos de interação presentes na maior parte das mídias sociais (endosso e comunicação), a partir da intuição de que elas capturam informações complementares sobre a relação entre a origem e o destino da interação. Como será discutido na seção 4, o endosso (como um *retweet* ou “Curtir”) tende a indicar apoio, enquanto a comunicação elimina a possibilidade de que um usuário seja indiferente ao outro. Ao combinar as duas interações, conseguimos separar as relações entre grupos de usuários entre apoio, antagonismo e indiferença.

O trabalho está organizado como se segue. A seção 2 discute trabalhos relacionados; enquanto a seção 3 ilustra o *Paradoxo dos Rivais*. Na seção 4, demonstramos como as interações de endosso e comunicação podem ser combinadas de forma a ressaltar o tipo de relação entre as diversas comunidades encontradas em uma rede multipolarizada. Finalmente, as conclusões e os próximos passos do trabalho são discutidos na seção 5.

2. Trabalhos Relacionados

Polarização, do ponto de vista sociológico, é formalmente entendido como um estado em que as opiniões expressadas acerca de uma tema são opostas quando comparadas a uma máximo teórico [Mouw and Sobel 2001]. Uma forma de medir a polarização na sociedade é fazer pesquisas de opinião e avaliar a presença de opiniões extremas na opinião pública [Abramowitz and Saunders 2005]. Quando a informação sobre relações entre os indivíduos que emitem opiniões está disponível, uma estratégia é verificar se a rede social induzida por essas relações pode ser dividida em duas comunidades coesas, em que cada comunidade representa uma posição ou preferência diferente em relação a um tópico, como liberais *versus* conservadores e apoiadores de controle de armas *versus* uso livre de armas, por exemplo. Grande parte dos trabalhos que analisam oposição e conflito de ideias em redes sociais se concentram no cenário clássico de polarização, principalmente no contexto político [Adamic and Glance 2005, Conover et al. 2011, Livne et al. 2011]. Nosso trabalho se diferencia de trabalhos anteriores por analisar o caso em que há múltiplos lados de embate. Em particular, estamos interessados em estudar redes sociais em que há K grupos de indivíduos agrupados de acordo com seu ponto de vista, e $K > 2$. Desconhecemos trabalhos que analisem redes sociais multipolarizadas do ponto de vista algorítmico; grande parte dos estudos de multipolarização do ponto de vista de relações sociais reside nas análises históricas das relações internacionais pré e pós-Guerra Fria. Em particular, o contexto histórico em que a multipolarização se manifesta entre as relações entre países motiva nosso trabalho à medida em que o mundo bipolarizado, durante a Guerra Fria, dominado por EUA e URSS, “sugeriu estratégias de segurança e desenvolvimento internacional facilmente inteligíveis às bases eleitorais e partidárias das lideranças políticas” [Sato 2000]. A literatura histórica é consistente em reconhecer que o mundo multipolarizado que emergiu após a Guerra Fria tornou mais complexo o entendimento das relações entre os grupos (no caso, países) e a distribuição de responsabilidades entre eles [Runchang 1998, Sato 2000].

De fato, é sabido que laços sociais podem ter interpretações e intensidades diferentes, e considerar todas as relações de forma idêntica é uma abordagem bastante simplista [Gilbert and Karahalios 2009]. Nosso objetivo é classificar as relações entre indivíduos que se relacionam em uma rede multipolarizada em 3 categorias, a saber, $\{apoio, antagonismo, indiferença\}$. Nesse sentido, nosso trabalho se relaciona com pesquisas que tentam classificar as relações entre usuários em grafos com relações positivas e negativas, como em [Leskovec et al. 2010a, Kunegis et al. 2010]. Esse problema surge em cenários como sistemas em que usuários avaliam os outros quanto à qualidade da tarefa que executam (como avaliação das revisões na Wikipedia) e avaliam a qualidade das opiniões emitidas por outras pessoas (como no Epinions [Leskovec et al. 2010b]). Nesses casos, cada aresta no grafo social possui um rótulo – **positivo** ou **negativo**, e diversas características dos nós e da estrutura da rede podem ser empregadas para prever o sinal das arestas.

Uma gama de trabalhos busca classificar as arestas de uma rede social em categoriais mais gerais, como amizade, relacionamento de trabalho e laço familiar [Tang et al. 2012, Tang et al. 2011]. Esses trabalhos, assim como os trabalhos que tentam prever o sinal (polaridade) de um relacionamento social, empregam uma lógica supervisionada, em que a classe de alguns relacionamentos são previamente conhecidos. No cenário em que trabalhamos, não há declaração explícita do relacionamento entre os

usuários; mas interações de diferentes tipos que podem fornecer evidências sobre o relacionamento entre elas.

Nosso trabalho busca determinar o tipo de relacionamento sem utilizar informação prévia já existente sobre alguns relacionamentos (dado que esta informação é indisponível), mas a partir da combinação de tipos de interação diferentes (endosso e comunicação). Nesse sentido, nosso trabalho se relaciona também com pesquisas que buscam detectar comunidades e padrões em redes heterogêneas, em que mais de um tipo de interação é estabelecida entre os nós [Davis et al. 2011, Tang et al. 2009, Cai et al. 2005].

3. O Paradoxo dos Rivais

Para estudar redes sociais multipolarizadas, escolhemos trabalhar com um cenário em que a multipolarização se manifesta claramente – o debate e interação entre torcidas de futebol. Em particular, consideramos usuários que escreveram mensagens sobre o campeonato brasileiro de futebol em 2010, 2011 e 2012 no Twitter, a mais popular plataforma de *microblogging*, que permite a qualquer usuário enviar mensagens curtas que instantaneamente são tornadas disponíveis aos usuários que “seguem” o autor da mensagem [Kwak et al. 2010]. A tabela 1 apresenta um sumário geral dos dados coletados ao longo de 3 anos, utilizando a API do Twitter para coletar tweets que mencionassem o nome de uma das 12 principais equipes do futebol brasileiro. O grafo é criado a partir dos usuários (nós), que são conectados com uma aresta direcionada caso um usuário retuíte algum outro. As menções são uma interação em que um usuário se comunica com o outro e serão utilizadas na seção 4.

Tabela 1. Descrição Geral do *dataset* considerado neste trabalho.

informação	valor
período	05/2010 a 12/2012
# de entidades-chave (polos)	12
# de <i>tweets</i>	35.834.453
# de <i>retweets</i> (RTs)	26.665.821
# de menções	23.901.121
# usuários no grafo	348.752

Como primeiro passo da metodologia, buscamos encontrar as comunidades nas quais os usuários se organizam. A detecção de comunidades em grafos é um problema estudado extensivamente pela comunidade de redes sociais e é um dos problemas clássicos da área [Clauset et al. 2004]. Nossa abordagem para encontrar comunidades é baseada na ideia de trabalhos anteriores de que, em alguns contextos, alguns nós são previamente conhecidos como nós que devem fazer parte de uma comunidade [Lin and Cohen 2010]. Por exemplo, no caso de debates políticos travados em algum meio online, os perfis oficiais dos candidatos ou dos partidos são sementes naturais e que definem comunidades. O algoritmo aplicado executa *random walks* a partir dessas sementes, e os nós são assinalados à semente mais próxima [Calais et al. 2011]. Note que o algoritmo específico empregado para encontrar comunidades é ortogonal à metodologia de análise de multipolarização aqui apresentada. A figura 2 exhibe, visualmente, as comunidades obtidas a partir da escolha dos perfis oficiais dos 12 principais clubes do futebol brasileiro, no Twitter.

A partir das 12 comunidades obtidas, buscamos entender como uma se relaciona com a outra, no que se refere à intensidade e à natureza das relações entre cada grupo.

Tabela 2. Proporção de arestas de endosso (retweet) que migram da comunidade i (linha) para a comunidade j (coluna). A tabela indica que as comunidades que representam torcidas de times rivais tendem a estar mais densamente conectadas no Twitter.

	CRU	CAM	GRE	INT	FLA	VAS	BOT	FLU	SPFC	PAL	COR	SAN
Cruzeiro	–	40%	6%	4%	11%	5%	3%	9%	5%	6%	7%	4%
Atlético-MG	34%	–	6%	5%	10%	7%	5%	8%	5%	6%	9%	5%
Grêmio	6%	5%	–	24%	10%	6%	3%	8%	9%	10%	12%	7%
Inter	5%	5%	37%	–	9%	5%	3%	6%	6%	9%	8%	5%
Flamengo	6%	7%	5%	4%	–	14%	7%	20%	7%	9%	15%	5%
Vasco	4%	5%	5%	3%	21%	–	8%	19%	7%	10%	12%	5%
Botafogo	5%	5%	5%	3%	21%	16%	–	22%	4%	6%	10%	4%
Fluminense	6%	6%	5%	3%	25%	16%	8%	–	7%	8%	11%	5%
São Paulo	3%	3%	6%	4%	8%	8%	7%	2%	–	19%	27%	11%
Palmeiras	3%	4%	6%	4%	9%	8%	2%	8%	16%	–	27%	12%
Corinthians	3%	4%	6%	4%	13%	7%	3%	8%	17%	21%	–	11%
Santos	3%	4%	5%	4%	10%	6%	2%	6%	14%	26%	17%	–

Note que, no caso de segmentação bipolarizada ($K = 2$), a única análise que pode ser feita é medir o grau de separação entre o par de comunidades obtido – o que é feito por meio de métricas como modularidade [Zhang et al. 2008] ou mesmo métricas desenvolvidas com o propósito específico de medir polarização em redes sociais. No caso $K > 2$, podemos computar a proximidade entre cada um dos $\binom{K}{2}$ pares de comunidades e compará-las. A tabela 2 exhibe, para cada par de torcidas T_i e T_j , a proporção de arestas originadas na comunidade T_i que se destina a algum usuário da comunidade T_j . Para cada torcida de onde as arestas se originam (linhas da tabela), destacamos em negrito as porcentagens de arestas destinadas à torcidas conhecidas por compartilharem uma forte rivalidade regional.

Ao observar os números da tabela, concluímos que as torcidas tradicionalmente conhecidas como **rivais** entre si são **mais próximas** no grafo social em relação às outras torcidas, mesmo que a aresta tenha uma semântica predominantemente positiva, como o *retweet*. Note, por exemplo, que 40% dos *retweets* que saem da comunidade CRUZEIRO se direcionam à comunidade ATLÉTICO-MG. No caso da rivalidade entre as equipes do Rio Grande do Sul, as proporções são de 24% e 37% conectando torcedores de GRÊMIO e INTER, e vice-versa. Um padrão similar pode ser observado para as comunidades de torcidas das equipes cariocas (Flamengo, Vasco, Botafogo e Fluminense) e equipes paulistas (São Paulo, Palmeiras, Corinthians e Santos).

Chamamos esta observação experimental de que usuários com relações de antagonismo estão mais próximos entre si na estrutura da rede social de **Paradoxo dos Rivais**: mesmo sendo antagônicos, eles estão mais próximos entre si do que com relação a usuários cuja relação é de menor antagonismo, ou mesmo indiferença. Note que, no tradicional cenário bipolarizado na qual a literatura se concentra, tal inconsistência não é percebida; já que um usuário fará parte do seu grupo e conseqüentemente estará mais distante do grupo alternativo; e não há outras medições de distância com as quais este valor possa ser comparado. Entendemos, portanto, que as medições que mostramos na tabela 2 são relevantes por revelar inconsistências e premissas implícitas que são assumidas por modelos de redes sociais. Uma implicação direta dos resultados que encontramos aqui é que a ideia de que a proximidade em redes sociais implica em similaridade entre usuários, largamente assumida em tarefas de recomendação [Backstrom and Leskovec 2011], nem sempre se sustenta.

Entendemos que o “paradoxo dos rivais” decorre de, pelo menos, quatro fatores:

1. **Endosso irônico.** É comum um usuário propagar a mensagem de outro, do qual

discorda ou tem uma posição diferente, fora do contexto em que a mensagem foi direcionada. Por exemplo, um político que tuite que "escândalos de corrupção são absurdos" poderá ter essa mensagem retuitada, meses mais tarde, por partidários da oposição, caso um escândalo contra ele mesmo seja revelado.

2. **Endosso falso.** Outra prática comum entre usuários do Twitter que possuem ideias opostas é criar endossos falsos, no formato "*RT @usuario mensagem falsa*", atribuindo a *@usuario* algo que ele nunca disse. Endossos falsos (*fake retweets*) já foram investigados no contexto de spam no Twitter [Mowbray 2010], em que spammers tentam se valer da reputação de celebridades. O objetivo, no contexto de discussões polarizadas, é outro – criar humor ou criticar alguém.
3. **Endosso pra exibir a opinião contrária.** Muitas vezes, um usuário propaga uma mensagem com a qual ele discorda com o objetivo de exibir para seus seguidores o conteúdo. Em geral, isso faz com que os partidários do usuário que discordou da mensagem retuitada se virem contra o usuário endossado, enviando a ele mensagens que expressem essa discordância.
4. **Endosso a neutros.** Pares de comunidades caracterizadas por usuários que possuem uma relação antagônica costumam ter usuários neutros que se comportam como pontes entre as duas comunidades. Por exemplo, no futebol, órgãos de imprensa locais podem ser endossados por torcedores das equipes locais, aumentando a proximidade desses grupos em relação a grupos de outras torcidas.

Acreditamos que este resultado tem implicações diretas na noção amplamente aceita de que, em redes sociais, a proximidade de dois nós na rede é um sinal de similaridade e compartilhamento de interesses entre eles [Lerman et al. 2012, Koren et al. 2007] e também de que membros de dois grupos/comunidades mais próximas também devem ser mais similares [Saha and Getoor 2008].

4. Entendendo a multipolarização via interações de endosso e comunicação

Nesta seção, apresentamos nossa proposta em como analisar as relações entre usuários em cenários multipolarizados de modo a restaurar a coerência das medições de proximidade entre os grupos. Partimos da observação de que o paradoxo dos rivais ocorre porque a noção de **importância** que uma determinada comunidade dá a outra se mistura com a **polaridade** dessa interação (positiva ou negativa) em um contexto multipolarizado; essa "importância" é materializada nos 4 fatores descritos na seção anterior. Para fatorar esses dois componentes (importância e polaridade), lançamos mão de dois tipos de interação que têm propósito distinto e que indicam relações diferentes entre o originador e o alvo da interação: **endosso** e **comunicação**. As principais mídias sociais online, como *Facebook* e *Twitter*, permitem aos seus usuários estabelecerem ambos os tipos:

1. **Endosso:** o endosso é uma interação em que um usuário implicitamente concorda com outro usuário a respeito de um conteúdo. No Twitter, *retweets* são endossos em que um usuário propaga para seus seguidores o conteúdo de uma mensagem; no Facebook, o apoio a um conteúdo é feito a partir do botão "Curtir". Nem sempre o *retweet* ou o *curtir* indica endosso, como explicado na seção anterior; porém, o endosso é o objetivo original e natural dessas interações.
2. **Comunicação:** Comunicação usuário-usuário é implementada na maior parte das mídias sociais como uma forma de troca de mensagens entre os usuários, seja de

forma particular ou privada. No Twitter, esta comunicação se dá por meio de mensagens diretas ou *menções*, um mecanismo por meio do qual um usuário se refere a outro usuário com a intenção de conversar com ele ou citá-lo em uma argumentação. Isso é feito escrevendo o nome do usuário (@usuario) em qualquer parte de uma mensagem.

Qualitativamente, o tipo de informação que as interações de endosso e comunicação provêm são diferentes, como mostrado na tabela 3. O endosso tende a indicar apoio, mas um usuário não endossar o outro pode significar antagonismo ou indiferença. A falta de interações de comunicação, por outro lado, indica que dois usuários são indiferentes um em relação ao outro, mas não indica a polaridade da relação – de apoio ou de oposição.

	se há a interação	se não há a interação
endosso	indica apoio	antagonismo ou indiferença?
comunicação	apoio ou antagonismo?	indica indiferença

Tabela 3. Endosso e Comunicação são interações em mídias sociais que provêm diferentes informações sobre a relação entre dois usuários e o tipo de opiniões que esperamos que eles troquem.

Nossa proposta é combinar ambas as interações de forma a determinar relações de apoio, antagonismo e indiferença entre usuários, a partir da informação diferente que cada uma provê. Para cada uma das $\binom{K}{2}$ pares comunidades de uma rede multipolarizada, computamos 3 medidas: $P_{RT}(i, j)$ é a porcentagem de nós da comunidade j que foram endossados por algum membro da comunidade j ; $P_{COMM}(i, j)$ é a porcentagem de nós da comunidade j que recebem alguma mensagem (comunicação) de algum membro da comunidade i ; e, finalmente, $R = \frac{P_{COMM}}{P_{RT}}$ é a razão entre essas duas medidas.

A tabela 4 exhibe os valores que obtivemos para cada uma dessas medidas computadas para cada par de comunidades. A tabela está ordenada em uma sequência decrescente da razão comunicação/endosso R . Note que os valores mais altos de R tendem a capturar as principais rivalidades entre as torcidas do futebol brasileiro, como Palmeiras x Corinthians, Cruzeiro x Atlético-MG e Vasco x Flamengo, entre outras. A título de ilustração, mostramos na segunda parte da tabela alguns pares de comunidades cujo valor de R é baixo: é o caso dos pares de comunidades Atlético-MG x Palmeiras, Botafogo x Cruzeiro e Internacional x Santos. Essas torcidas não são conhecidas por serem rivais, mas por estabelecerem uma relação de neutralidade ou indiferença.

A partir da tabela, concluímos que valores de R grande estão ligados a antagonismo: quando, proporcionalmente, um lado comunica mais com o outro lado de que o endossa, este é um sinal de que os lados tem posições divergentes, e as comunicações indicam troca de críticas e discussões. Note que, no caso dos pares de comunidades antagonônicas, as interações de endosso também ocorrem (o que resultou no paradoxo dos rivais, discutido na seção anterior), mas são menos frequentes do que as interações de comunicação.

A análise conjunta de endossos e comunicações, portanto, restaura as relações de apoio, antagonismo e indiferença esperados para as torcidas analisadas. A quantidade de arestas que conecta um grupo ao outro, exibidos na tabela 2 da seção anterior, cap-

Tabela 4. Proporção de nós alvo de endosso, comunicação e a razão entre essas duas medidas. Os pares de comunidades em que há um conhecido antagonismo (devido à rivalidade regional entre os clubes) apresentam razões comunicação/endosso mais altas.

Grupo 1	Grupo 2	P_{RT}	P_{COMM}	razão comunicação/endosso
Palmeiras	Corinthians	0,046	0,125	2,717
Atlético-MG	Cruzeiro	0,082	0,196	2,390
Santos	Corinthians	0,029	0,063	2,172
São Paulo	Corinthians	0,080	0,169	2,113
Vasco da Gama	Flamengo	0,058	0,122	2,103
Internacional	Grêmio	0,058	0,101	1,741
Fluminense	Flamengo	0,038	0,063	1,658
Vasco	Corinthians	0,021	0,033	1,571
Vasco	Fluminense	0,119	0,172	1,445
Atlético-MG	Palmeiras	0,026	0,019	0,731
Botafogo	Cruzeiro	0,041	0,029	0,707
Internacional	Santos	0,024	0,016	0,667

turam a **importância** que uma comunidade dá a outra; a razão comunicação/endosso R captura o grau de **antagonismo** entre elas, a partir da premissa de que comunicações frequentes, sem endosso, indicam oposição de ideias. Um grupo que estabeleça poucas interações com o outro é indiferente a ele; e antagonico a um grupo em que a proporção de comunicação supere significativamente a proporção de endossos.

Entendemos que os resultados aqui obtidos e que o melhor entendimento de relações de de apoio, antagonismo e indiferença estabelecidos em redes sociais são relevantes para a construção de modelos mais efetivos em, pelo menos, duas tarefas-chave em mineração de dados na Web:

1. **Análise de Sentimento.** Diversos algoritmos de análise de sentimento empregam informação da estrutura da rede social subjacente [Calais et al. 2011, Tan et al. 2011b]. O melhor entendimento das relações sociais pode prover a estes modelos informações valiosas para a interpretação das opiniões, principalmente quando as opiniões e relações entre os diversos grupos varia ao longo do tempo;
2. **Recomendação.** O melhor entendimento das relações entre usuários pode ser incorporado a algoritmos de recomendação em rede sociais utilizados para recomendar novas amizades, seguidores e mesmo o conteúdo gerado por eles [Aranda et al. 2003]. Em particular, modelos de recomendação baseados apenas em proximidade tendem a recomendar nós populares, mas que são antagonicos com o alvo da recomendação. Em particular, um conteúdo popular e que não é consumido por um usuário provavelmente não o é por antagonismo, não por indiferença ou ignorância (pois o conteúdo é popular) [Herlocker et al. 2004]. Por isso, acreditamos que a modelagem das relações de similaridade, antagonismo e indiferença podem beneficiar algoritmos de recomendação.

5. Conclusões e Próximos Passos

Neste trabalho, exploramos rede sociais que emergem em contextos em que múltiplos grupos de usuários interagem em relações de apoio, antagonismo e indiferença. Nessas redes, os quais chamamos de multipolarizadas, inconsistências e sutilezas que não são

percebidas no caso clássico de bipolarização são tornadas mais claras. Em particular, demonstramos que que usuários que se encontram mais próximos na rede social nem sempre são mais similares – anomalia que chamamos de “Paradoxo dos Rivais”. Em seguida, demonstramos que a análise conjunta das interações de endosso (*retweet*) e comunicação (menção) no grafo social do Twitter ajuda a entender as relações de apoio, antagonismo e indiferença entre comunidades de torcidas de futebol.

Note que modelos aprendidos no caso multipolarizado vão apoiar o desenvolvimento de algoritmos que também são aplicáveis no caso bipolarizado; queremos explorar redes multipolarizadas não apenas pela relevância dos domínios multipolarizados, mas também porque os resultados e modelos podem ser aplicados em redes sociais bipolarizadas e mesmo redes sociais não-polarizadas.

Os próximos passos do trabalho incluem, além do estudo de novas redes e domínios, o levantamento de novas características que indiquem a presença de antagonismo, como, por exemplo, o tempo até que um *tweet* seja retuitado: tweets irônicos, em geral, são retuitados dias ou meses depois de originalmente postados, em um outro contexto. Um outro caminho que iremos investigar é tentar diferenciar indiferença de antagonismo a partir da investigação da proximidade entre um par de usuários: se eles são próximos um do outro no grafo e são populares, a ausência de uma conexão entre eles é um sinal de antagonismo, no sentido de que eles conscientemente evitam um ao outro. Um de nossos objetivos também é modelar a dinâmica temporal entre as relações entre os grupos de uma rede multipolarizada. Acreditamos que os eventos multipolarizados que ocorrerão no Brasil em 2014, como as Eleições Presidenciais e a Copa do Mundo de futebol, irão proporcionar dados e oportunidades interessantes de análise.

6. Agradecimentos

O presente trabalho foi realizado com o apoio do CNPq e do Google, por meio do *Google Brazil Focused Research Grants Program*.

Referências

- Abramowitz, A. and Saunders, K. (2005). Why can't we all just get along? the reality of a polarized america. *The Forum: A Journal of Applied Research in Contemporary Politics*, 2.
- Adamic, L. A. and Glance, N. (2005). The political blogosphere and the 2004 u.s. election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, LinkKDD '05, pages 36–43, New York, NY, USA. ACM.
- Aranda, J., Givoni, I., Handcock, J., and Tarlow, D. (2003). An online social network-based recommendation system. *Science*, pages 4–5.
- Backstrom, L. and Leskovec, J. (2011). Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 635–644, New York, NY, USA. ACM.
- Cai, D., Shao, Z., He, X., Yan, X., and Han, J. (2005). Community mining from multi-relational networks. In *Proceedings of the 9th European conference on Principles and*

- Practice of Knowledge Discovery in Databases*, PKDD'05, pages 445–452, Berlin, Heidelberg. Springer-Verlag.
- Calais, P. H., Veloso, A., Meira, Jr, W., and Almeida, V. (2011). From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In *Proc. of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, San Diego, CA.
- Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6):066111+.
- Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., and Menczer, F. (2011). Political polarization on twitter. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- Davis, D., Lichtenwalter, R., and Chawla, N. V. (2011). Multi-relational link prediction in heterogeneous information networks. In *Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '11, pages 281–288, Washington, DC, USA. IEEE Computer Society.
- Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 211–220, New York, NY, USA. ACM.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53.
- Hunter, J. (1992). *Culture Wars: The Struggle to Define America*. BasicBooks.
- Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50(6):1141–1151.
- Koren, Y., North, S. C., and Volinsky, C. (2007). Measuring and extracting proximity graphs in networks. *ACM Trans. Knowl. Discov. Data*, 1(3).
- Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., Luca, E. W. D., and Albayrak, S. (2010). Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM*, pages 559–. SIAM.
- Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *WWW '10: Proceedings of the 19th international conference on World wide web*, pages 591–600, New York, NY, USA. ACM.
- Lerman, K., Intagorn, S., Kang, J.-H., and Ghosh, R. (2012). Using proximity to predict activity in social networks. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 555–556, New York, NY, USA. ACM.
- Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010a). Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 641–650, New York, NY, USA. ACM.
- Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010b). Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1361–1370, New York, NY, USA. ACM.
- Lin, F. and Cohen, W. W. (2010). Semi-supervised classification of network data using very few labels. In *ASONAM*, pages 192–199.

- Livne, A., Simmons, M. P., Adar, E., and Adamic, L. A. (2011). The party is over here: Structure and content in the 2010 election. In *Proceedings of the 5th Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*.
- McCright, A. M. and Dunlap, R. E. (2011). The politicization of climate change and polarization in the american public's views of global warming, 2001-2010. *The Sociological Quarterly*.
- Mouw, T. and Sobel, M. (2001). Culture wars and opinion polarization: The case of abortion. *American Journal of Sociology*, 106(4):913–943.
- Mowbray, M. (2010). The twittering machine. In Filipe, J. and Cordeiro, J., editors, *WEBIST (2)*, pages 299–304. INSTICC Press.
- Paul DiMaggio, John Evans, B. B. (1996). Have American's Social Attitudes Become More Polarized? *American Journal of Sociology*, 102(3):690–755.
- Runchang, X. (1998). Multi-polarization in world politics and strategic competitions among big powers after the cold war. 4:20–24.
- Saha, B. and Getoor, L. (2008). Group proximity measure for recommending groups in online social networks. In *2nd ACM SIGKDD Workshop on Social Network Mining and Analysis (SNA-KDD)*.
- Sato, E. (2000). A agenda internacional depois da Guerra Fria: novos temas e novas percepções. *Revista Brasileira de Política Internacional*, 43:138 – 169.
- Sunstein, C. R. (2002). The Law of Group Polarization. *Journal of Political Philosophy*, 10(2):175–195.
- Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., and Li, P. (2011a). User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11*, pages 1397–1405, New York, NY, USA. ACM.
- Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., and Li, P. (2011b). User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11*, pages 1397–1405, New York, NY, USA. ACM.
- Tang, J., Lou, T., and Kleinberg, J. (2012). Inferring social ties across heterogenous networks. In *Proceedings of the fifth ACM international conference on Web search and data mining, WSDM '12*, pages 743–752, New York, NY, USA. ACM.
- Tang, L., Wang, X., and Liu, H. (2009). Uncovering groups via heterogeneous interaction analysis. In *ICDM*, Miami, FL, USA.
- Tang, W., Zhuang, H., and Tang, J. (2011). Learning to infer social ties in large networks. In *Proceedings of the 2011 European conference on Machine learning and knowledge discovery in databases - Volume Part III, ECML PKDD'11*, pages 381–397, Berlin, Heidelberg. Springer-Verlag.
- Zhang, Y., Friend, A. J., Traud, A. L., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2008). Community structure in congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications*, 387(7):1705–1712.