The Role of Social Capital in Information Diffusion over Twitter: a Study Case over Brazilian posts

Hercules Sandim[1,2], Danilo Azevedo[1], Ana Paula Couto da Silva[1], and Mirella M. Moro[1]

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Brasil
{herculessandim,danilo-va}@ufmg.br, {ana.coutosilva,mirella}@dcc.ufmg.br
[2] Universidade Federal de Mato Grosso do Sul, Campo Grande, Brasil

Abstract. Social Capital is the resulting advantage of the individual's localization in a social structure. It can be measured by traditional complex networks metrics, or specific ones, such as information capital, brokerage and bridging. Our goal is to verify which users have high information capital, bridging and brokerage for providing and spreading information. To do so, we first categorize Twitter users into seven types: typical users, primary media, secondary media, independent experts, fan accounts, fake accounts and potential bots. Then, we analyze their profiles on trending topics. Our results show potential bots and fan accounts as the main information spreaders in Brazil, a very concerning result given the upcoming presidential election in October 2018.

Keywords: Social Networks · Social Capital · Information Diffusion.

1 Introduction

Online Social Networks (OSNs), such as Facebook, Twitter, LinkedIn and Instagram, have achieved unprecedent growth in recent years. Current statistical data show Facebook has over 2.2 billion monthly active users, while Instagram, Twitter and LinkedIn have over 813 million, 330 million and 260 million monthly active users respectively [38]. Such huge volume of users and relationships is a motivation for several researches in the areas of Complex Networks, Big Social Data, and Urban Computing [16,22,23,36].

Among many usages, OSNs are powerful media for information diffusion [1,18,30,27]. Information diffusion occurs when there is a flow of information from one individual to another. The information may be retained by an individual or spread out on the network [39]. Individuals who are in a privileged position in the OSN have more social capital for information diffusion.

Social capital is a comprehensive concept, without one single definition or metric to capture all its facets [31]. In Social Network Analysis (SNA), it is the resulting advantage of the individual's right localization in a social structure [9]. For instance, the individual who holds information has the power to change what happens in one environment and to understand its surroundings [42].

In this scenario, new players arise and become relevant: the "human sensors", or citizens who share information about their environment via OSNs, supplementing, complementing or even replacing information as measured by physical sensors [13]. Human involvement is particularly useful in detecting multiple processes in complex personal, social and urban spaces where traditional embedded sensor networks have limitations [37]. The society engagement in the OSNs, whether individually or in small groups, can be facilitated through its social capital [34]. Overall, humans are relevant data sources, acquiring and spreading information on their own [37].

Such human sensing for information diffusion can be useful for emergency events management, crime detection, urban administration, intelligent transportation, smart traffic control, public healthcare, political engagement, among others [26,37]. In addition, the information can change the view of the recipient, motivating him/her to join the social network of the information provider [37]. Also, trust among communicating individuals strongly affects the reach and impact of information [19], and trust is an important social capital aspect [3,34].

However, OSNs have become significant spreaders of false facts, urban legends, fake news, or, more generally, *misinformation*. Misinformation drives the emergence of a post-truth society, where the debate becomes damaged by the repetition of discussions refuting the primary media or independent experts (official sources of truthful information) [27].

Even so, people still rely on the news published in social media. In 2017, according to Reuters [32], despite the predominance of TV stations in the media environment, social media played an essential role in the consumption of news, being the primary sources of news within the Brazilian urban context. For instance, in 2016, the impeachment of President Dilma Rousseff drew attention for its repercussion in Brazilian social media. However, only 30% of people believe that the social media is free from undue political influence.

Overall, the accurate information diffusion contributes to building smart urban spaces [29], promoting improvement in the quality of life of their citizens. Consequently, to ensure reliability to the citizen, the primary media and independent experts must take the lead in OSNs. Henceforth, we address the social capital forms as related to individuals' abilities to acquire and spread information. We model a network of Twitter users relating them through retweets in messages that contribute to a subject to become Brazilian Twitter Trend. We analyze and compare social capital metrics to verify the importance of the primary media and independent experts compared to typical users, fan accounts, and potential bots, in the Brazilian information diffusion context. Although considering only messages written in Brazilian Portuguese, our methodology is broad enough to be applied to any language without loss of generality.

2 Related Work

This section is divided in two parts: general work on social capital and online social networks (Section 2.1), and specific metrics for social capital (Section 2.2).

2.1 Social Capital and Online Social Networks

In SNA, social capital is the resulting advantage of the individual's right localization in a social structure [9]. There are many definitions of social capital. For instance, Bourdieu [6] emphasizes social capital as the collective resources used by individual members to obtain services and benefits either in the absence or conjunction with their economic capital. For Coleman [11], social capital is a neutral resource that facilitates any action, where the individuals are responsible for achieving their objectives. In turn, Putnam [34] defines social capital as characteristics of the social organization, such as social networks, social norms, and social trust, which facilitate coordination and cooperation for mutual benefit and civil engagement. Moreover, Bertolini and Bravo [3] define social capital as the resources individuals have access in a given network.

Authors also have different ways of thinking about how to classify social capital types. Putnam [33] defines two forms of social capital: bonding and bridging. In [21], Jackson introduces seven types: information, brokerage, bridging, coordination, favor, reputation, and community capital. Moreover, Burt [9] defines two activities related to social capital: brokerage (an individual self placing in a privileged position in the network), and closure (the coordination of a closed group of individuals in the network).

Regarding social capital over OSNs, there are metrics to visibility, reputation, popularity, and authority. Bertolini and Bravo [3] describe that reputation and authority comprise the sum of knowledge and information disseminated within a given group. Measuring the relationships' strength is also relevant (identifying weak ties), as well identifying hubs and influencers. Then, there are traditional metrics such as: ego-network measure, the structural hole measure [10], homophilia measure [5], and the standard centrality measures [14]. Also, Kang and Shen [25] define quantitative metrics for social capital, not validated in OSNs, and Michalak et al. [31] present measures of social capital based on techniques of cooperative game theory.

Social capital is also used for explaining information diffusion processes in social networks. Authors in [2,17] show that the presence of weak ties and hubs accelerates the information diffusion on social networks. In turn, Kleinberg [14] models the networks' cascading behavior like an influence process among individuals. Differently, many authors deal with the essential nodes detection (influencers) in OSNs as being the individuals who maximize spreading [20,24,35].

The aforementioned works do not rely on social capital for understanding the process of information diffusion over OSNs. Here, we assume that information capital, bridging and brokerage are skills from individuals that are well placed on the network structure and take advantage of their position to acquire and control the information flow. Furthermore, we apply the hub and authorities concepts proposed by Kleinberg in [28] as tool for identifying the use of social capital on information diffusion process.

2.2 Metrics for Social Capital

The previous section presented general work on social capital and networks. We now discuss existing metrics for social capital. We begin by providing notation that helps to define the metrics. Let G denote a directed graph, where $V(G)$ is its set of vertices (or nodes), and $E(G)$ its set of edges. Also, $n = |V(G)|$ and $m = |E(G)|$. G is represented by its adjacency matrix $A \in [0,1]^{n\times n}$, where $A_{u,v}$=1 indicates the existence of an edge between u and v, and $A_{u,v}$=0 otherwise. Given $N_o(u)$ as the set of u's outgoing neighbors and $N_i(u)$ the set of u's incoming neighbors, then $|N_o(u)|$ is u's *out-degree* and $|N_i(u)|$ its *in-degree*. Finally, $N(u)$ is the set of u's neighbors, where $N(u) = N_o(u) \cup N_i(u)$ and $|N(u)|$ is u's *degree*.

A *path* in G between two nodes u and v is a succession of distinct nodes $u \Rightarrow u^0, u^1, u^2, ..., u^p \Rightarrow v$ such that $A_{k,k+1} = 1$ $(\forall k \mid 0 \leq k < p)$. A *geodesic* (shortest *path*) between nodes u and v is a *path* such that no other *path* between them involves a smaller number of edges. Let g_{uv} the number of *geodesics* connecting u to v, and $g_{uv}(i)$ the number of *geodesics* that node i is on. Lastly, let d the graph diameter, that is, the largest geodesic distance between any pair of nodes.

Information Capital. Decay Centrality (DC) has been applied to measure individual's information capital [21]. In summary, DC measures the number of individuals reached by a specific individual, regardless the path length to arrive to them. DC counts *paths* of different lengths, i.e., how many people one can reach at different distances. The decay of information with distance is captured via a parameter p, with $0 < p < 1$. Furthermore, DC favors individuals that reach the largest number of neighbors in up to T hops, known as information's endurance. The DC metric of a node i, is given by Equation 1, where $N^l(i)$ is the set of individuals at maximum distance l from i in G.

$$DC(i) = \sum_{l=1}^{T} p^l.|N^l(i)|. \tag{1}$$

From Equation 1, a node has high information capital if its broadcast information reaches a large number of nodes. Otherwise, a node has low information capital. Although it has a simple calculation, DC does not consider all the possible paths that information might take [21]. The Eigenvector Centrality (EC) [4], or simply EigenCentrality, is an alternative to solve such a limitation [21]. For EC metric, the node importance depends on its neighborhood importance. Moreover, EC is a node influence measure in the network and given by Equation 2, where λ is the largest eigenvalue of R, and R is an eigenvector of A (the graph's adjacency matrix).

$$EC(i) = \frac{1}{\lambda} \sum_{t \in N(i)} EC(t). \tag{2}$$

Bridging Capital. Individuals who are bridges in the network topology are special nodes that can control the information flow [21], and/or accelerate its diffusion [17]. Here, we follow Granovetter's theory [17] that defines a bridge as a weak tie in the network. For labeling a node as a bridge, we apply the Neighborhood overlap metric (NO) as defined in Equation 3. Then, following Brandao and Moro [7], we classify a tie (edge, link) between nodes u and v as weak if: $0 \leq \mathrm{NO}(u, v) \leq 0.2$.

$$NO(u, v) = \frac{|N_o(u) \cap N_o(v)|}{|N_o(u) \cup N_o(v) - \{u, v\}|} \tag{3}$$

Brokerage Capital. From [10,21], nodes with high Betweenees Centrality (BC) play the role of brokers in the network. A normalized BC metric (for directed graphs) is given by [15] as Equation 4.

$$BC(i) = \frac{\sum_{j,k \in E(G)} \frac{g_{jk}(i)}{g_{jk}}}{(n-1).(n-2)}. \tag{4}$$

Hubs and Authorities. In order to rank the most important nodes based on their outgoing and incoming links, we apply the HITS algorithm proposed in [28]. As we discuss in Section 2.1, this approach enables profiling nodes with special roles on the information diffusion process. In summary, HITS algorithm computes two types of ranking: (*i*) the authority ranking estimates the node importance based on the incoming links; and (*ii*) the hub ranking estimates the node importance based on the outgoing links. We refer the reader to [28] for more details on such an algorithm.

2.3 Main Contributions

Our contributions over the related work are:

- We apply network topological metrics for analyzing information capital (eigenvector centrality), bridging (neighborhood overlap), and brokerage (betweenness centrality);
- We apply decay centrality for measuring information capital;
- We apply the HITS algorithm for profiling nodes in Twitter; and
- We provide a comparative analysis of the top 10 Twitter users, regarding network topological metrics, decay centrality, and the HITS algorithm.

3 Methodology

Our work evaluates social capital over a social network as extracted from posts in an online microblogging platform. To do so, we first build a dataset through collecting real posts in Twitter over a week. In order to be able to easily qualify the posts and our analyses results, our collecting process is limited to tweets in Brazilian Portuguese, our native language. Nonetheless, our methodology is broad enough to be applied to any language without loss of generality. Then, in order to better qualify our analysis, we categorize the users in two different forms. Next, we implement the previously discussed metrics and apply them over a network modeling. Finally, we are able to evaluate such metrics and the potential relevance of users according to their behavior. The next sections discuss respectively: the dataset building process, our categorization for user types, metrics implementations and parameters setting, and the network modeling considered.

3.1 Dataset

Our work evaluates social capital over a social network as extracted from tweets. Tweet is a Twitter[3] post (limited to 280 characters, in Brazil), and retweet (RT) is a re-posting of a tweet, which helps quickly sharing a given tweet. Moreover, Twitter Trends (TT) are topics that have become immediately popular (as opposed to topics have been popular for a while or on a daily basis)[4].

We initially collect Brazil's top 50 TTs. For each TT, we collect 100 most popular tweets. For each tweet, we collect 100 most recent RTs. The number of collected posts is 100 due to search limitations in the Twitter API[5]. We collect continuously for seven days (2018/04/20 to 2018/04/26), to acquire a more significant amount of data.

After cleaning and data processing, the dataset[6] contains 165,936 and 371,612 distinct users and messages, respectively. In this fase, 1,648 distinct subjects appeared in TTs. In the top 10 TTs collected, there are subjects related to TV, entertainment, musicians, sports, and a single related to commemorative dates ("Tiradentes"). There is no political or economical subject in the overall top 10 (surprinsingly given the current Brazilian crisis), as presented in Table 1.

Now, considering only political/economic TT, the top 10 topics regard the current Brazilian political crises and the upcoming presidential election in October 2018, as shown in the bottom half of Table 1. In summary, the TT cover: *Lula Livre* and *Prisão de Lula* are popular clamor over the arrest of former Brazilian president *Luís Inácio Lula da Silva*; *Ciro Nogueira*, *Palloci*, *Rocha Loures*, *Mantega*, and *Temer* are Brazilian politicians who are currently targets for corruption investigations; *Odebrecht* is a large Brazilian construction company, also investigation target; *Marcos Valério* is a publicist involved (and delator) in corruption schemes; *PSDB* is a political party; and *STF* and *STJ* are the federal supreme and superior courts in Brazil.

3.2 Twitter User Types

We categorize Twitter users in two ways. The first type is called *posting categories* and regards user behavior over posting tweets and retweets. There are

[3] https://www.twitter.com
[4] https://help.twitter.com/en/using-twitter
[5] https://developer.twitter.com/en/docs
[6] Dataset available at http://homepages.dcc.ufmg.br/~mirella/projs/apoena/datasets.html

Table 1: Number of occurrences of top TTs over dataset.

Rank	TT	Topic	Occurrences
1º	#MaratonaRedeBBB	TV, Entertainment	26
2º	Júlio César	Football	24
	Biel	Musician	24
	#PowerCoupleBrasil	TV, Entertainment	24
5º	Tiradentes	Commemorative date	20
6º	#FinalBBB18	TV, Entertainment	19
	Diego Souza	Football	19
8º	Ferrugem	Musician	17
9º	Dia de Grêmio	Football	16
10º	#SuperligaNoSporTV	Sports	15
81º	Lula Livre	Political/Economic	7
	Ciro Nogueira	Political/Economic	7
	Palocci	Political/Economic	7
138º	Lula e Mantega	Politica/Economic	5
	Rocha Loures	Political/Economic	5
	Marcos Valério	Political/Economic	5
	Temer e PSDB	Political/Economic	5
	Odebrecht	Political/Economic	5
	Prisão de Lula	Political/Economic	5
180º	STJ e STF	Political/Economic	4

users who tweet more than retweet (*providers*), users who retweet more than tweet (*spreaders*), and users who balance both actions (*neutral*). Therefore, we define *p_ratio* as being the ratio between the tweets (*out-degree*) of a user, and the total number of tweets and retweets (*degree*), as shown in Equation 5. Thus, users can be spreaders ($0 \leq p\_ratio \leq 0.25$), neutral ($0.25 < p\_ratio < 0.75$), or providers ($0.75 \leq p\_ratio \leq 1$). In our specific dataset (Section 3.1), most users are spreaders (164,884 users), followed by providers (1,043 users), and then neutral (9 users).

$$p\_ratio = \frac{out\text{-}degree}{out\text{-}degree + in\text{-}degree}. \tag{5}$$

The second way, called *social categories*, distributes users into seven types regarding the social characteristics observed in the user's timeline on Twitter:

i *Potential bots* for users who have *Botometer Score*[7] bigger then 2.5;
ii *Independent experts* for users who are experts in a given theme, which usually include journalists acting independently;
iii *Fake accounts* for users who assume a false identity, passing through another personality;

[7] *Botometer* checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. *Botometer Score* ranges from 0 to 5. Higher scores are more bot-like (http://botometer.iuni.iu.edu) [12].

iv *Fan accounts* for users who identify themselves as celebrity fan accounts, or use the social network just to demonstrate their fanaticism (about a football team or a political party for example);
v *Primary media* for users who represent the major channels of communication in the Brazilian context of news publishing;
vi *Secondary media* for users who act as lower relevance media, comparing to primary media; and
vii *Typical users* for other users who do not fall into the previous types.

3.3 Implementations and Parameters Setting

Our experiments use implementations[8] and parameters settings as follows.

- **DC**: we implement DC in Python[9], and we run it setting p=0.5 (moderated decay) and T=d=8 (maximum endurance). The DC calculated values are normalized ($0 \leq DC(i) \leq 1$);
- **EC**: we use Python NetworkX[10] library, which is an iterative algorithm with two parameters: (i) max_iter = 100, which defines the maximum number of iterations in power method; and (ii) ε =1e-06, which is the error tolerance used to check convergence in power method iteration. Calculated values are also normalized ($0 \leq EC(i) \leq 1$);
- **BC**: we use Brandes Algorithm[8] with normalization ($0 \leq BC(i) \leq 1$);
- **NO**: we implement NO in Python; and
- **HITS**: we use Python NetworkX library, an iterative algorithm with three parameters: (i) max_iter = 100, which defines the maximum number of iterations in power method; (ii) ε =1e-06, which is the error tolerance used to check convergence in power method iteration; and (iii) $normalized = True$, which normalizes results by the sum of all of the values.

3.4 Network Modeling

We model the online social network as a directed graph G, where nodes u and v ($u, v \in V(G)$) represent Twitter users and an edge $(v \rightarrow u) \in E(G)$ means that user u retweeted a message from user v. Such a modeling provides social capital for users who publish (*out-degree*) or retweet (*in-degree*) messages. The graph G built from our dataset has $|V(G)|$=165,936 nodes and $|E(G)|$ =280,898 edges.

Fig. 1 shows a toy example of our network modeling. It illustrates *providers* (red color nodes), *spreaders* (blue color nodes), and *neutral* (yellow color nodes). Another example is given by Fig. 2 with a small piece of our dataset that shows a potential bot (@Felipe__100) acting on the network.

[8] Code available at `http://homepages.dcc.ufmg.br/~mirella/projs/apoena/datasets.html`
[9] `https://python.org`
[10] `https://networkx.github.io/documentation`

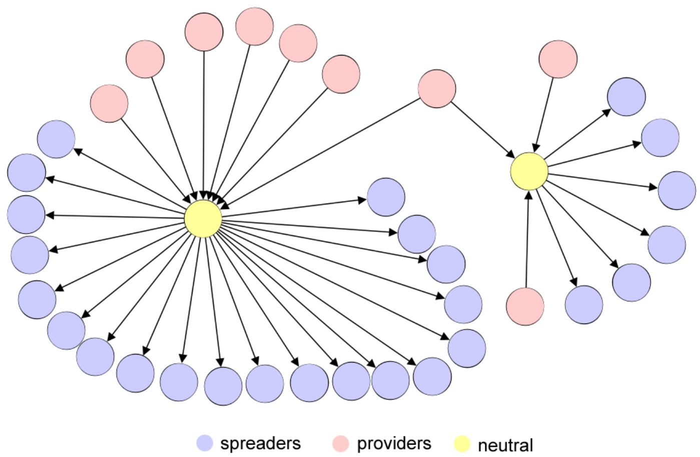

Fig. 1: Toy example with the three types of nodes regarding posting behavior.

Fig. 2: Example of a potential bot (@Felipe__100) retweeting a message.

4 Data Analysis

Here, we analyze and compare metrics for social capital facets regarding the information diffusion process to explore the importance of primary media and independent experts compared to typical users, fan accounts and potential bots in the Brazilian context of information diffusion. First, we analyze our types of users according to the information capital measures (Section 4.1), an important contribution of our work in the context of misinformation in urban centers. Then, we expand such evaluation by considering the other forms of calculating social capital measure regarding: bridging (Section 4.2), brokerage (Section 4.3), and hubs and authority index (Section 4.4).

4.1 Information Capital Measures

We start our analyses with Decay Centrality (DC), which favors users that reach the largest number of neighbors in up to T hops. Thus, the *providers* that transmit information for other providers have higher DC. Table 2 presents the results, in which sports and entertainment news providers stand out (seven out of 10). Therefore, users have high EC if they are related to other users who are also well

Table 2: Top 10 users for Decay Centrality

Rank	User	Posting Category	Social Category	DC
1º	@Flamengo	provider	primary media (sports news)	1
2º	@FoxSportsBrasil	provider	primary media (sports news)	0.951
3º	@bbb	provider	primary media (entertainment news)	0.862
4º	@Esp_Interativo	provider	primary media (sports news)	0.672
5º	@VascodaGama	provider	primary media (sports news)	0.610
6º	@globoesportecom	provider	primary media (sports news)	0.563
7º	@HugoGloss	provider	independent expert (entertainment news)	0.350
8º	@MomentsBrasil	provider	independent expert (general news)	0.335
9º	@g1	provider	primary media (general news)	0.323
10º	@RedeGlobo	provider	primary media (telecommunication)	0.305

Table 3: Top 10 users for EigenCentrality

Rank	User	Posting Category	Social Category	EC
1º	@ClaraGuimarae	spreader	potential bot	1
2º	@hey_dann	spreader	fan account	0.723
3º	@pastorasandram5	spreader	typical user	0.699
4º	@JonahWhite30	spreader	typical user	0.675
5º	@_DiasFabio	spreader	typical user	0.673
6º	@AntroReality	spreader	fan account	0.666
7º	@PedroAlvesFer12	spreader	typical user	0.665
8º	@jeanbuenodumke	spreader	potential bot	0.652
9º	@arte_prima	spreader	typical user	0.647
10º	@limalblue_ofc	spreader	typical user	0.645

connected to the network. In our context, users who retweet the highly retweeted information are privileged. Hence, typical users, fan accounts and potential bots stand out, as the results in Table 3. Unfortunately, these are the user types that contribute to turning a subject into a trend, depreciating the quality of the main information served on Twitter, in the midst of the Brazilian upcoming elections.

As seen earlier, DC and EC are centrality metrics that reveal information capital of individuals in the network. However, both metrics present different results. DC reveals the main providers because they are nodes that start long sequences of broadcasts and relays. Meanwhile, EC reveals the main *spreaders* because they retweet many messages of important nodes (*providers*). Therefore, DC and EC do not measure the same event, and there is no correlation between them ($\rho = -0,095$, $p\text{-}value = 0$)[11].

4.2 Bridging Capital Measure

Our modeling process induces the formation of disconnected components because we create a component for each message, where this message points to its RTs. However, users tend to connect as a new TT arises. Then, bridges link weakly connected components, allowing broadcast through such weak tie. Overall, in our dataset, there are seven weakly connected components and 165,901 strongly connected components.

Fig 3 shows a subgraph that contains the top weak ties (red color edges) and their neighborhood. Each weak tie is an edge ($v \rightarrow u$), where v is a *provider* (represented as red nodes), and u is a *spreader* (represented as blue nodes). Nodes

[11] ρ is Spearman Rank Correlation Coefficient [41]

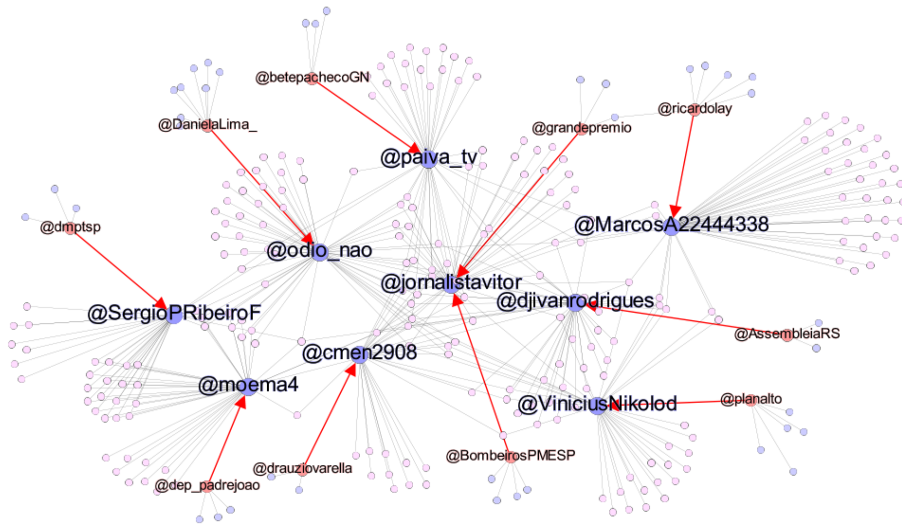

Fig. 3: Example of weak ties (bridges) in our dataset emphasized by red edges.

u and v are bridges because they allow the connection between u and v's neighborhoods. Since bridging capital refers to the ties' strength, it does not make sense to correlate it with the aforementioned node centrality metrics. Specifically in our dataset, bridging capital highlights five political party fan accounts as *spreaders* [12], two independent experts as *spreaders*[13], three primary media as *providers*[14], and three independent experts as *providers*[15]. Moreover, the potential bot "@paiva_tv" stands out acting as secondary media in the network.

Bridging and brokerage are similar in the sense that both focus on the singular position of an individual in the network. Nevertheless, they are still different and measured through distinct forms.

4.3 Brokerage Capital Measure

Nodes with high BC are important brokers in communication and information diffusion [21,40]. Table 4 shows that BC presents the primary media and independent experts as main brokers in the Brazilian information diffusion context, where six out of 10 users belong to the "Grupo Globo"[16], and all top 10 are *providers*. Thus, the information tends to circulate in the network passing through the *providers*, mainly belonging to a single group. We also analyze the

[12] @cmen2908, @ViniciusNikolod, @SergioPRibeiroF, @odio_nao, @moema4
[13] @djivanrodrigues, @jornalistavitor
[14] @drauziovarella, @planalto, @AssembleiaRS
[15] @betepachecoGN, @ricardolay, @DanielaLima_
[16] Grupo Globo, the largest media broadcaster in Brazil, `http://grupoglobo.globo.com/ingles`

Table 4: Top 10 users for Betweenness Centrality (BC)

Rank	User	Posting Category	Social Category	BC
1º	@RedeGlobo	providers	primary media (telecommunication)	1
2º	@g1	providers	primary media (general news)	0.619
3º	@Esp_Interativo	providers	primary media (sports news)	0.464
4º	@SBTonline	providers	primary media (telecommunication)	0.450
5º	@HugoGloss	providers	independent expert (entertainment news)	0.408
6º	@gshow	providers	primary media (entertainment news)	0.323
7º	@SporTV	providers	primary media (sports news)	0.267
8º	@GloboNews	providers	primary media (general news)	0.225
9º	@MaisVoce_Globo	providers	primary media (TV program)	0.126
	@gustavovillani	providers	independent expert (sports news)	0.126

correlation between BC and DC (not shown due to space constraints). There is a weak positive correlation between them ($\rho = 0,322$, *p-value* $= 0$), as both measures identify information providers over network.

4.4 Hubs and Authority for Social Capital

The HITS algorithm [28] calculates Hub Index (HI) and Authority Index (AI), where HI measures social capital of users who produce much information (like providers) and AI measures social capital of users who mostly retweet, that is, users who spend a lot of time browsing the network (like potential bots). Table 5 presents the results in two parts. First, Table 5a shows the hubs as the main providers, where the sports information providers stand out (nine out of 10). Meanwhile, Table 5b shows the potential bots and football fan accounts stand out (seven out of 10) as authorities. Interestingly, all football fan accounts are fans of the "Clube de Regatas Flamengo (CRF)" (the largest football fan club in Brazil). The potential bots suspended[17] by Twitter were also CRF fans. Probably, the suspended accounts were linked to *tweetdecking*[18] practice to inflate popularity (social capital facet).

Furthermore, there is a weak positive correlation between AI and EC ($\rho = 0,261$, *p-value* $= 0$) because both metrics use eigenvector concept and capture similar events (not equivalent). As seen in Section 4.1, providers who broadcast information for another providers have high DC values (like a hub). Hence, there is a very strong positive correlation between HI and DC ($\rho = 0,995$, *p-value* $= 0$), and there is a weak positive correlation between HI and BC ($\rho = 0,324$, *p-value* $= 0$) – not illustrated due to space constraints.

5 Conclusion

Categorizing users in OSNs is an important task for understanding how information spreads over society. Humans as data sources can streamline the sharing of relevant events, whether for emergency events management, crime detection, urban administration, intelligent transportation, smart traffic control,

[17] `https://help.twitter.com/pt/managing-your-account/suspended-twitter-accounts`

[18] `http://www.newsweek.com/tweetdecking-why-twitter-suspended-multiple-accounts-840031`

Table 5: Top 10 users for Hubs and Authority

(a) Hub Index (HI)

Rank	User	Posting Category	Social Category	HI
1º	@Flamengo	provider	primary media (sports news)	1
2º	@FoxSportsBrasil	provider	primary media (sports news)	0.837
3º	@Esp_Interativo	provider	primary media (sports news)	0.498
4º	@globoesportecom	provider	primary media (sports news)	0.426
5º	@VascodaGama	provider	primary media (sports news)	0.243
6º	@bbb	provider	primary media (entertainment news)	0.185
7º	@venecasagrande	provider	independent expert (sports news)	0.176
8º	@SporTV	provider	primary media (sports news)	0.160
9º	@maurocezar	provider	independent expert (sports news)	0.108
10º	@lucaspedrosaEI	provider	independent expert (sports news)	0.101

(b) Authority Index (AI)

Rank	User	Posting Category	Social Category	AI
1º	@MarcosA22444338	spreader	(football) fan account	1
2º	@_DiasFabio	spreader	typical user	0.995
3º	@jhonatalima355	spreader	(football) fan account	0.904
4º	@PHRN1895	spreader	potential bot (suspended)	0.889
5º	@_arthurpassos	spreader	potential bot	0.880
6º	@PedroAlvesFer12	spreader	typical user	0.879
7º	@Rgo17_	spreader	potential bot (suspended)	0.870
8º	@FlavioM32255797	spreader	(football) fan account	0.859
9º	@RlPenha	spreader	typical user	0.851
10º	@AllanRN1981	spreader	(football) fan account	0.842

public healthcare, or even political engagement. It is a matter of major concern when bots or malicious users assume such activities.

Here, we analyzed the Twitter Brazilian users behavior in publishing and sharing Twitter trend topics to understand how important information flows over the network. We do so by addressing the social capital forms as related to individuals' abilities to acquire and spread information. We analyzed and compared social capital metrics to verify the importance of the primary media and independent experts compared to typical users, fan accounts, and potential bots, in the Brazilian information diffusion context.

In general, potential bots and fan accounts are users who spread information through retweets, and they are the main authorities in the social network. Potential bots have automated behavior. Then, they can be programmed for malicious purpose. Furthermore, fan accounts exacerbate a fanaticism sentiment. For instance, the retweets may inflate the "hate speech" or spread fake news. This is very concerning given the upcoming Brazilian presidential election in October 2018. However, despite the *Media Groups* monopoly, we also found the primary media and independent experts as the main information providers, which may represent there is still hope in controlling misinformation over the network.

Acknowledgements. Work funded by CAPES, CNPq and FAPEMIG, Brazil.

References

1. Ahsan, M., Kumari, M., Singh, T., Pal, T.L.: Sentiment based information diffusion in online social networks. IJKDB **8**(1), 60–74 (2018)
2. Barabási, A.; Frangos, J.: Linked: the new science of networks science of networks. Perseus Books Group (2014)
3. Bertolini, S., Bravo, G.: Social capital, a multidimensional concept. In: Euresco Conference (2001)
4. Bonacich, P.: Power and centrality: A family of measures. American journal of sociology **92**(5), 1170–1182 (1987)
5. Borgatti, S., Jones, C., Everett, M.: Network measures of social capital. Connections **21**(1), 27–36 (1998)
6. Bourdieu, P.: The forms of capital. In: Richardson, J. (ed.) Handbook of Theory and Research for the Sociology of Education, pp. 241–258. Greenwood, New York (1986)
7. Brandão, M.A., Moro, M.M.: Analyzing the strength of co-authorship ties with neighborhood overlap. In: International Conference on Database and Expert Systems Applications. pp. 527–542. Springer (2015)
8. Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology **25**(2), 163–177 (2001)
9. Burt, R.S.: Brokerage and closure: An introduction to social capital. Oxford University Press (2005)
10. Burt, R.S.: Structural holes: The social structure of competition. Harvard University Press (2009)
11. Coleman, J.S.: Social capital in the creation of human capital. American Journal of Sociology **94**, S95–S120 (1988)
12. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: Botornot: A system to evaluate social bots. In: Int'l Conf on World Wide Web, Companion Volume. pp. 273–274. Montreal, Canada (2016)
13. Doran, D., Severin, K., Gokhale, S., Dagnino, A.: Social media enabled human sensing for smart cities. AI Communications **29**(1), 57–75 (2016)
14. Easley, D., Kleinberg, J.: Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press (2010)
15. Freeman, L.C.: A Set of Measures of Centrality Based on Betweenness. Sociometry **40**(1), 35–41 (1977)
16. Gao, S., Janowicz, K., Couclelis, H.: Extracting urban functional regions from points of interest and human activities on location-based social networks. Trans. GIS **21**(3), 446–467 (2017)
17. Granovetter, M.S.: The Strength of Weak Ties. vol. 78, p. 1360–1380. JSTOR (1973)
18. Hu, Y., Song, R.J., Chen, M.: Modeling for information diffusion in online social networks via hydrodynamics. IEEE Access **5**, 128–135 (2017)
19. Hui, C., Goldberg, M., Magdon-Ismail, M., Wallace, W.A.: Simulating the diffusion of information: An agent-based modeling approach. IJATS **2**(3), 31–46 (2010)
20. Iannelli, F., Mariani, M.S., Sokolov, I.M.: Network centrality based on reaction-diffusion dynamics reveals influential spreaders. CoRR **abs/1803.01212** (2018)
21. Jackson, M.O.: A typology of social capital and associated network measures. CoRR **abs/1711.09504** (2018)
22. Jerônimo, C.L.M., Campelo, C.E.C., de Souza Baptista, C.: Using open data to analyze urban mobility from social networks. JIDM **8**(1), 83 (2017)

23. Kadar, C., Brüngger, R.R., Pletikosa, I.: Measuring ambient population from location-based social networks to describe urban crime. In: Social Informatics. pp. 521–535 (2017)
24. Kang, C., Kraus, S., Molinaro, C., Spezzano, F., Subrahmanian, V.: Diffusion centrality: A paradigm to maximize spread in social networks. Artificial Intelligence **239**, 70 – 96 (2016)
25. Kang, S., Shen, L.: A quantitative measure for meal-mate social capital networks. In: Int'l Conf. on Intelligent Environments. pp. 124–131 (2016)
26. Kim, J., Bae, J., Hastak, M.: Emergency information diffusion on online social media during storm cindy in U.S. Int J. Information Management **40**, 153–165 (2018)
27. Kim, J., Tabibian, B., Oh, A., Schölkopf, B., Gomez-Rodriguez, M.: Leveraging the crowd to detect and reduce the spread of fake news and misinformation. CoRR **abs/1711.09918** (2017)
28. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5), 604–632 (Sep 1999)
29. Kukka, H., Kostakos, V., Ojala, T., Ylipulli, J., Suopajärvi, T., Jurmu, M., Hosio, S.: This is not classified: everyday information seeking and encountering in smart urban spaces. Personal and Ubiquitous Computing **17**(1), 15–27 (2013)
30. Li, M., Wang, X., Gao, K., Zhang, S.: A survey on information diffusion in online social networks: Models and methods. Information **8**(4), 118 (2017)
31. Michalak, T.P., Rahwan, T., Moretti, S., Narayanam, R., Skibski, O., Szczepanski, P.L., Wooldridge, M.: A new approach to measure social capital using game-theoretic techniques. SIGecom Exchanges **14**(1), 95–100 (2015)
32. Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D.A., Nielsen, R.K.: Reuters institute digital news report 2017. Reuters (2017)
33. Putnam, R.D.: Tuning in, tuning out: The strange disappearance of social capital in america. PS: Political science & politics **28**(4), 664–683 (1995)
34. Putnam, R.D.: Bowling alone: the collapse and revival of american community. In: ACM Conference on Computer Supported Cooperative Work. p. 357 (2000)
35. Saito, K., Kimura, M., Ohara, K., Motoda, H.: Super mediator - a new centrality measure of node importance for information diffusion over social network. Inf. Sci. **329**(C), 985–1000 (2016)
36. Smarzaro, R., de Lima, T.F.M., Jr., C.A.D.: Could data from location-based social networks be used to support urban planning? In: Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017. pp. 1463–1468 (2017)
37. Srivastava, M., Abdelzaher, T., Szymanski, B.: Human-centric sensing. Phil. Trans. R. Soc. A **370**(1958), 176–197 (2012)
38. Statista: Most famous social network sites worldwide as of april 2018, ranked by number of active users, `https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users`, accessed: 2018-04-24
39. Stimmel, C.L.: Building smart cities: analytics, ICT, and design thinking. CRC Press (2015)
40. Tang, L., Liu, H.: Community detection and mining in social media. Synthesis lectures on data mining and knowledge discovery. Morgan and Claypool (2010)
41. Wayne, W.D., et al.: Applied nonparametric statistics. Boston, MA: PWS-Kent (1990)
42. Zheng, Y., Capra, L., Wolfson, O., Yang, H.: Urban computing: Concepts, methodologies, and applications. ACM Trans. Intell. Syst. Technol. **5**(3), 38:1–38:55 (2014)