Article

American Behavioral Scientist
2014, Vol. 58(10) 1260–1277

DOI: 10.1177/0002764214527088
abs.sagepub.com

The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter

Elizabeth Dubois[1] and Devin Gaffney[2]

Abstract

This study compares six metrics commonly used to identify influential players in two of Canada's largest political Twitter communities based on the users, and ranking order of users, identified by each metric. All tweets containing the hashtag #CPC, representing the Conservative Party of Canada (government), and #NDP, representing the New Democratic Party of Canada (official opposition), were collected over a 2-week period in March 2013 and a follower network graph was created. Social network analysis and content analysis were employed to identify influentials. Kendall's τ was the primary quantitative measure for comparison. Categorization of Twitter profiles of users found within the top 20 most influential lists, according to each metric of influence, made up the qualitative portion of analysis. The authors find that measures of centrality—indegree and eigenvector centrality—identify the traditional political elite (media outlets, journalists, politicians) as influential, whereas measures considering the quality of messages and interactions provide a different group of influencers, including political commentators and bloggers. Finally, the authors investigate the possibility of using the local clustering coefficient of nodes to identify those who are both aware of the traditional elite and embedded in tightly knit communities, similar to the "opinion leader," described in the Two-Step Flow Hypothesis.

Keywords

Twitter, online communities, influence, social media, social network analysis

[1]University of Oxford, Oxford, UK
[2]Little Bird, Portland, OR, USA

Corresponding Author:
Elizabeth Dubois, University of Oxford, Oxford, UK.
Email: elizabeth.dubois@oii.ox.ac.uk

Introduction

The debate concerning the potential for the Internet to empower individuals is well documented (Dahlgren, 2005; Hindman, 2009), with many scholars noting the importance of examining changes in political systems given the increasing pervasiveness of digital technologies (Agre, 2002). Digital technologies make it possible for new players, for example, average citizens, to become involved in political decisions (Dubois & Dutton, 2012). Understanding how these players interact is crucial to understanding how our political system works.

The ability to influence (convincing an individual to change his or her opinion, attitude, and/or behavior) is a powerful skill. Two theories emerge explaining an individual's political influence: The "opinion leader" uses social support and social pressure to influence his or her personal network (Katz & Lazarsfeld, 1955), and the "influential" uses his or her visible position in a large network to spread messages widely (Rogers, 2010). The former speaks to individuals occupying an important role in political discussion communities, which are thought to be integral to democratic integrity (Dillard, Segrin, & Harden, 1989). The latter is tied to work on diffusion of innovations and information cascades (Bakshy, Hofman, Watts, & Mason, 2011; Lerman & Ghosh, 2010; Rogers, 2010).

Our study relies on both interpretations in order to answer the question, Which political players are the most influential within the two largest Canadian political communities on Twitter (specifically within the #CPC and #NDP hashtag communities)?

Our literature review traces the theoretical history of opinion leadership and influentials before identifying the most common ways in which these individuals are identified in Twitter research. Measures of network centrality, interaction, knowledge, and local embeddedness in the network are compared. The top 20 influential users are then examined.

Canadians are among the highest Internet users in the world (Duong, 2012), making this study an instructive case for understanding political communities in a digital environment. Twitter was selected because it is a popular site among Canadians, provides a clear set of boundaries for data collection, and has been studied in some depth.

The Canadian political Twittersphere is large enough to provide sufficient data but small enough to conduct meaningful qualitative analysis, which serves as a baseline for comparison. A number of scholars have begun to investigate the Canadian political Twittersphere and are providing useful basic descriptions (Gruzd, 2012; Small, 2011).

In this study, the use of two distinct hashtags, #CPC (Conservative Party of Canada) and #NDP (New Democratic Party of Canada), provides opportunity to compare different communities within the Canadian Twittersphere and increases reliability (Yin, 2008).

Our findings suggest that, when considering network-wide placement, those who traditionally hold power—media and politicians—maintain power. When indicators of expertise and interaction are considered, bloggers and political commentators gain prominence. Guided by the Two-Step Flow Hypothesis, we propose and test a way to identify those who are locally influential. These results are specific to this case. However, the broader notion that various measures of influence in fact identify very different kinds of influencers is beneficial for researchers defining influence and selecting measures for future studies. Our discussion of local measures of influence, as they relate to the notion of opinion leaders, provides direction for future studies of political discussion communities.

Defining Influence

Katz and Lazarsfeld's *opinion leader*, as described in their Two-Step Flow Hypothesis, is able to influence his or her close personal ties by exerting social pressure and social support. They are important political players because they transmit messages to a wider public who do not choose to access messages directly from the political elite. They are knowledgeable and trusted on a specific topic (Katz & Lazarsfeld, 1955). Four core facets of influence are suggested: *having a following, seen as an expert, knowledgeable/have expertise*, and in a position within their local community to exert social pressure and social support/*social embeddedness*.

In response to theories of direct mass media effects, Katz and Lazarsfeld argued that mass media influence a small segment of the population, opinion leaders, who then influence the wider public. The hypothesis has been tested in multiple settings (Katz, 1957). Scholars have modified the hypothesis to include multiple steps (Weimann, 1982), combine it with other theories of media effects like agenda setting (Brosius & Weimann, 1996), and apply it in a digital media environment (Norris & Curtice, 2008).

Studies of political discussion build off this work. Although identifying influencers is less important in political discussion work than in work on the Two-Step Flow, the importance of strong ties and quality of discussion remains crucial (Huckfeldt, Sprague, Kuklinski, Wyer, & Feldman, 1995). A second direction focuses on how messages flow through social networks. Early research traced the effects of a message from its sender (originally mass media) through a chosen network (e.g., a Twitter hashtag community or geographically bounded Twitter communities). This analysis is useful to market researchers interested in who is best placed in a network to reach a wide following (Bakshy et al., 2011; Watts & Dodds, 2007).

These broad theoretical backdrops for identifying influentials remain important. Political discussion happens in a hybrid media environment in which various players all have access to a variety of tools (Chadwick, 2011). Politicians, journalists, activists, bloggers, and average users may use multiple strategies, such as broadcasting, asking and responding to questions, and posting links, to interact. The context in which one is influential is uncertain in today's hybrid media environment.

This means that describing someone as an "influential" or "opinion leader" can be problematic because it is difficult to identify traceable practices, specific tools or strategies, or even structures of social connections that are necessarily unique to influencers. To address this problem, researchers use a range of different metrics to identify influentials. Whereas most advocate for a certain combination of metrics (González-Bailón, Borge-Holthoefer, & Moreno, 2013; Lee, Kwak, Park, & Moon, 2010), there are multiple ways to operationalize influence based on the inclusion or exclusion of various facets of influence.

Measuring Influence

The most labor-intensive, yet most common, method within studies of the Two-Step Flow (Katz, 1957) is to ask people who they are influenced by and if they believe that they themselves are influential. However, the presence of social media and online social networking sites has meant that much trace data exist from which networks of influence can be constructed (Welser, Smith, Fisher, & Gleave, 2008). Although digital techniques and approaches are becoming increasingly popular, they do not provide a perfect substitute for interview- or survey-based methods of identifying opinion leaders.

Recently, studies using access to large-scale social data have assumed a more easily quantifiable definition of influence. They tend to measure influence by the number of followers and/or how far a message travels (Rattanaritnont, Toyoda, & Kitsuregawa, 2012; Wu, Hofman, Mason, & Watts, 2011).

These studies use social network analysis to compute metrics. The assumption is when a given member of a network (called a node) is placed in that network in such a way that they could be heard by many others also in that network, that node is likely to be influential (e.g., Bakshy et al., 2011; Subbian & Melville, 2011). The facet of influence that these studies rely on in order to provide an operational definition is *having a following*.

Another approach is to consider interaction in the network. In the context of Twitter research, this approach often involves either counting the times a user is mentioned (Cha & Gummadi, 2010) or applying metrics to a re-tweet or mention network (Sousa, 2010). The main facet of influence that these studies are concerned with is *being seen as an expert*.

Content analysis provides an alternate option for identifying influentials. Some studies use complex methods, such as ranking quality of language or tracking URLs over time, for assigning levels of influence to individuals (Bakshy et al., 2011). A simpler base form is counting keywords in tweets. Using keywords is more contextually bound to any particular study and also more rare in work in this area. However, this kind of content analysis is a potentially useful measure for identifying opinion leaders since it is possible to assess a user's knowledge/expertise on a subject based on the content of the user's tweets. This approach theoretically addresses a similar facet of influence to the interaction measures noted above. Yet, instead of considering how a user is treated by the audience, it considers the quality of messages a user sends (i.e., knowledge and expertise).

Although some studies have used the measures mentioned above in order to identify influentials within smaller networks (e.g., Gruzd, Wellman, & Takhteyev, 2011), identification of influentials based on social embeddedness in their local community, is largely absent.

The local clustering coefficient is one social network analysis measure that could provide insight into the role of social embeddedness. The measure scores nodes in terms of the degree of completeness of the graph among their immediate neighbors (Watts & Strogatz, 1998). Despite being used in only a handful of studies (e.g., Sousa,

Table 1. Network Summaries.

Metric	#CPC	#NDP
Users (nodes)	3,860	3,536
Friendships (edges)	163,506	144,658
Statuses (tweets)	730,562	653,989

Note. CPC = Conservative Party of Canada; NDP = New Democratic Party of Canada.

2010), this information could be particularly useful to those interested in identifying opinion leaders and political discussants who are locally influential.

We have a range of metrics with which to define and operationalize influence but little sense of how they empirically play out or how they compare to one another. In addition to asking who is most influential, we ask, Do different operational definitions of influence actually measure the same general trends when considering a Twitter network constructed in this way, and if not, how do they differ and why?

Accessing Political Discussion Communities

Given the importance of context when identifying influence, we set a specific topic, Canadian politics, and collected tweets from a specific time period, March 12 to 26, 2013. We chose two distinct online communities as denoted by the use of the #CPC and #NDP hashtags.

A Twitter Streaming API[1] connection was established, selecting tweets that matched each hashtag to create a base index of users of interest. A network graph was generated where connections among users who followed one another were used as the edges between nodes (users in the dataset). Up to 200 of the most recent tweets were collected for every user. Table 1 provides a summary of the number of nodes, edges, and tweets in the two datasets, as well as aggregate network statistics.

A given Twitter hashtag is sometimes used by more than one community for different reasons (Conover, Gonçalves, Ratkiewicz, Flammini, & Menczer, 2011). The portion of total tweets using either hashtag outside of the context of Canadian politics is estimated to be less than 10% in each network,[2] and no top influencers across any metric included users who used the hashtag outside of the Canadian political context. Further, when one follows a given hashtag community, he or she is subject to all messages with that tag. For these reasons, we left all tweets/users in the networks.

Next, we applied six different metrics, summarized in Table 2, to all users in each community in order to assign influence scores. Using a rank correlation coefficient to compare metrics, we examined the degree to which different metrics agree, and in turn, what different facets of influence may exist.

Of the metrics employed, the majority are standard network analysis metrics. *Interaction* counts the times a user is mentioned by other users during the 2-week sampling period. *Knowledge* was developed for this study to provide further insight about influence based on the content produced by the user in question.

Table 2. Metrics Summary.

Metric	Description
Indegree	A simple importance rank expressed by the number of nodes with a directed edge pointing toward the given node (i.e., followers within sampled network).
Eigenvector centrality	A measure that quantifies importance of a node. A score is higher when a node's connections are in turn highly connected.
Clustering coefficient	A metric conferring the degree to which a given node is embedded within a tightly bound set of other nodes.
Knowledge	The number of tweets that a user posts containing context-specific terms divided by the number of tweets in the sample terms derived from a random sample of tweets collected during the sampling period from both networks.
Interaction	The total number of times that all other users mentioned the given user within the dataset.

The knowledge metric ranks users based on language used. A coding schedule was created using a random subsample of tweets in order to identify keywords deemed to indicate knowledge of Canadian politics. The assumption is that those using specific terms are likely to have higher levels of expertise and be portrayed as more knowledgeable. Operationally, this number is the number of times a user posted a tweet containing one or more keywords divided by the number of tweets the user posted within the dataset.

Comparing Rankings

Quantitatively, we were most interested in whether or not different metrics ranked users in similar ways. All metrics were run against the two separate datasets. A non-parametric ranking statistic, Kendall's τ, was used to draw out pairwise comparisons of all metrics and analyze the relative degrees to which metrics agree (when a high τ is found), disagree (when a low τ is found), and/or diverge across the entire set (when a τ approaching zero is found). Results are reported in Table 3.

The ranking statistics bore similar trends in both the #CPC and #NDP networks: Eigenvector centrality and indegree ranked highly together, interaction count conferred minor agreement with other rankings, and knowledge and clustering coefficient scores seemed independent from other metrics. This does not reveal the distinctions or whether they are substantive but instead provides a way to compare how each metric ranks users at an aggregate level.

Metrics that ranked highly together suggest that they are either measuring the same facet of influence or measuring different facets that tell a similar story. Consider the pair with the highest Kendall's τ: eigenvector centrality and indegree. These both indicate how central a node is within a network; the facet of influence addressed is *having a following*. The knowledge metric ranking highly with these measures would suggest

Table 3. Kendall's τ Ranks.

		Kendall's τ Full Network		Kendall's τ Network Excluding Top Elbow by Eigenvector Centrality	
First Metric	Paired Metric	#CPC	#NDP	#CPC	#NDP
Indegree	eigenvector centrality	0.856	0.7968	0.8251	0.8518
Indegree	interaction count	0.4933	0.476	0.4178	0.5182
Eigenvector centrality	interaction count	0.4345	0.4246	0.3466	0.4599
Eigenvector centrality	knowledge	0.2431	0.1502	0.1959	0.2355
Indegree	knowledge	0.2310	0.1446	0.1848	0.2189
Knowledge	interaction count	0.1365	0.0678	0.0800	0.1219
Clustering coefficient	knowledge	0.0544	0.0373	0.0835	0.0786
Eigenvector centrality	clustering coefficient	0.0148	0.1482	0.1065	0.0625
Indegree	clustering coefficient	–0.0156	0.1219	0.0720	0.0191
Clustering coefficient	interaction count	–0.0659	0.0738	–0.0045	–0.0618

Note. CPC = Conservative Party of Canada; NDP = New Democratic Party of Canada.

that *expertise/influence* overlaps with *having a following*. This is not the case. It appears that those who have a large following do not necessarily use terms we would expect experts to use. Qualitative analysis is required to understand the differences between the measures.

In short, we interpret these findings to suggest that indegree and eigenvector centrality rank according to a similar facet of influence, one that is quite different from the facets that our other metrics measure. At this point, we can say that different metrics, sensitive to different facets of influence, do indeed identify users differently.

Identifying Influential Political Players

We have discussed quantitatively the relationship between these various measures, noting that they do not identify influentials in the same order but that some tend to agree more than others. Are these differences substantive?

We conducted a content analysis of the profiles of all accounts found among the top 20 influencers per metric. This sample is large enough to provide variety but small enough to conduct meaningful qualitative analysis and is in line with past studies (e.g., Cha & Gummadi, 2010; Wu et al., 2011).

Table 4. Top 20 Accounts by Metric.

Facet	Centrality			
Metric	Indegree		Eigenvector	
Community	#NDP	#CPC	#NDP	#CPC
1	ThomasMulcair	ElizabethMay	kady	kady
2	ElizabethMay	kady	ThomasMulcair	SusanDelacourt
3	kady	acoyne	nathancullen	ElizabethMay
4	oliviachow	bobraeMP	SusanDelacourt	acoyne
5	nathancullen	SusanDelacourt	oliviachow	davidakin
6	bobraeMP	davidakin	davidakin	bobraeMP
7	SusanDelacourt	oliviachow	punditsguide	punditsguide
8	MeganLeslieMP	TheHillTimes	ElizabethMay	natnewswatch
9	davidakin	punditsguide	MeganLeslieMP	TheHillTimes
10	PeggyNashNDP	MeganLeslieMP	iancapstick	althiaraj
11	PaulDewar	rabbleca	PaulDewar	dgardner
12	punditsguide	dgardner	PeggyNashNDP	oliviachow
13	iPoliticsca	natnewswatch	MPJulian	Carolyn_Bennett
14	LibbyDavies	Carolyn_Bennett	LibbyDavies	MeganLeslieMP
15	nikiashton	PaulDewar	laura_payton	wicary
16	NDP_HQ	PeggyNashNDP	althiaraj	rabbleca
17	iancapstick	althiaraj	nikiashton	smithjoanna
18	laura_payton	stephenlautens	iPoliticsca	Paul Dewar
19	TorontoStar	stephen_taylor	natnewswatch	stephenlautens
20	MPJulian	leadnowca	smithjoanna	AntoniaZ

Legend	
	Politician/party
	Journalist/media outlet
	Commentator/blogger
	Other notable (includes political staff)
	Likely opinion leader
	Unlikely opinion leader
	Spam, nonpolitical organization, bot, etc.

Note. CPC = Conservative Party of Canada; NDP = New Democratic Party of Canada.

A coder and one of the authors, familiar with Canadian politics, classified accounts based on whether they were media, partisan, activist, commentator/blogger, other notable, opinion leader, or average.[3]

Our qualitative analysis suggests that, yes, different metrics identify different kinds of political players. Tables 4 and 5 provide the username of each of the top 20 ranked accounts given each metric.

Table 5. Top 20 Accounts by Metric.

Facet	Interaction	
Metric	Mentions	
Community	#NDP	#CPC
1	kady	kady
2	ThomasMulcair	acoyne
3	ElizabethMay	bobraeMP
4	joycemurray	ElizabethMay
5	TorontoStar	joycemurray
6	NSNDP	davidakin
7	davidakin	**HenrikSandbergJ**
8	bobraeMP	Beari8it
9	SheilaGunnReid	dgardner
10	RealMattHopkins	natnewswatch
11	MeganLeslieMP	Min_Reyes
12	nspector4	Bergg69
13	leadnowca	RealMattHopkins
14	natnewswatch	stephenlautens
15	PeggyNashNDP	nspector4
16	nathancullen	SheilaGunnReid
17	DBECanada	rabbleca
18	Bergg69	trapdinawrpool
19	Beari8it	MHallFindlay
20	MHallFindlay	SusanDelacourt
Facet	Knowledge	
Metric	Key Words	
Community	#NDP	#CPC
1	**Tenaciousceeee**	**Tenaciousceeee**
2	NDPHoC_NPDCdC	DickieAnginson
3	truthmashup	NDPHoC_NPDCdC
4	alexboulerice	ValckeNDP
5	CharleyCanucky	SLangeneggerCBC
6	samdinicol	JeffreyGriese
7	ctvqp	InfoAlerteBot
8	kismith	futurecpleaders
9	journo_dale	truthmashup
10	G_Soule	alexboulerice
11	ValckeNDP	CharleyCanucky
12	colewhogan	journo_dale
13	jessebrady	G_Soule
14	sedouglas	BrandanRowe
15	jaystor	samdinicol
16	*	ParmGill
17	Buswell_FedNDP	DaveJHL
18	SavannahNDP	NDPInduNPD
19	NDPInduNPD	jaystor
20	jasbirsandhu	markabel5

(continued)

Table 5. (continued)

Facet	Locally Embedded	
Metric	Clustering Coefficient (adjusted)	
Community	#NDP	#CPC
1	Darkcreations	miss_nik
2	stevem20	*
3	Quebecois100	CarolynMLB
4	schultzphoto	smokeycrow
5	**RoyalShipyard**	Rambojks
6	philhampton	CanadasPredsFan
7	TRStephen	mikofox
8	orangeclockwork	lynneakin
9	JimFare	Starfishlb
10	NickLane	*
11	greentak	Omarheaps
12	**ESG_Solutions**	Smoker420Kush
13	YTPaleo	LW709
14	*	ArezooC
15	smilesomiles	DLeowinata
16	**MyTorontoHome**	aliceedunn
17	Remzer	chrisdarimont
18	**rojas_lily**	OrlandoMisfit
19	jay9mac	**ottawa_rt**
20	ialuddington	savagestate

Legend	
	Politician/party
	Journalist/media outlet
	Commentator/blogger
	Other notable (includes political staff)
	Likely opinion leader
	Unlikely opinion leader
	Spam, nonpolitical organization, bot, etc.

Note. Deleted accounts are marked with an asterisk (*) and refer to those that no longer existed at the time of writing. This could be due to removal by Twitter for spamming or by an independent decision made by the account owner. CPC = Conservative Party of Canada; NDP = New Democratic Party of Canada.

Metrics concerned with the following of a Twitter user tended to identify traditionally important and highly visible political players such as media outlets, journalists, and politicians. Metrics with embedded assumptions about the importance of being seen an expert identified political commentators and bloggers as particularly influential as did metrics, which embedded assumptions about the importance of acting like an expert. Finally, our metric, which prioritizes local social embeddedness, identified primarily average users.

Cha and Gummadi (2010) noted that indegree is the basic measure of influence used across disciplines, although it may better measure popularity than actual influence. The assumption made by those using indegree to measure influence is that the most important facet of the influence is a large following. Eigenvector centrality further suggests that it is a matter of having a large following of those who also have a large following (Shamma, Kennedy, & Churchill, 2009).

It is not surprising that the 20 most highly ranked users in the indegree and eigenvector centrality lists (across both cases) were largely media outlets, journalists, and politicians. For example, live-blogger and CBC journalist @kady was found near the top of both #CPC and #NDP lists, as was Green Party leader @elizabethmay. The only exceptions were @iancapstick, a former communications director for the NDP appearing in both #NDP indegree and eigenvector centrality lists, @stephen_taylor, a well-known CPC blogger appearing in the #CPC indegree list, and @leadnowca, an activist group appearing in the indegree #CPC list.

When we think about the #CPC and #NDP Twitter communities as opportunities for getting news (Kwak, Lee, Park, & Moon, 2010), it is not surprising that central accounts are those of politicians and journalists involved in politics. They are most likely to send out firsthand and/or reliable information and have a professional reputation. The general public are interested in who they are and what they say, as are their counterparts. However, are these journalists and other political elite necessarily best placed to lead political discussions or to convince others to change opinions?

Content-based rankings, like our interaction metric, offer an alternate route to network structure metrics. In both cases, three quarters of the 20 most mentioned Twitter users in the #CPC and #NDP lists were media and politicians. The media and politicians who made it on each list were not the exact same for each hashtag, although some overlap did exist. For example, @NSNDP, a Twitter feed run by the Nova Scotia New Democrats, was sixth most mentioned in the #NDP network but was not found to be among the 20 most influential by any other metric. Similarly, Liberal Leadership candidates @JoyceMurray and @MhallFindlay were among the most mentioned in each network, yet neither appeared influential by other measures.

Next, we consider keyword ranking, what we call knowledge. Slightly less than half of the most highly ranked accounts were deemed average users or opinion leaders with a mix of politicians, parties, journalists, and bloggers filling in the list. This list reveals the prominence of political staffers. Staffers in the CPC and NDP lists are all affiliated with the NDP. This is likely a result of the NDP's communication strategy, where those within the party were prompted to provide the public with consistent language that our coding scheme identified.

Testing our coding schedule repeatedly and using a random sample of actual data to develop a hierarchical coding schedule increased validity and reliability (Richards & Richards, 1995; Rourke & Anderson, 2004). However, it remains possible that results were skewed in favor of certain groups or individuals using particular language. For example, if the NDP were pushing Senate reform while the CPC were attempting to divert conversation, we might see skewed results. This skew is not due to level of knowledge but to specific words becoming politically positive or negative depending on partisan leaning. Consulting news coverage over our sampling period showed that there were no major indicators of such situations, but an exhaustive analysis of non-Twitter-based content is beyond the scope of this article. Such analysis is not required for our basic comparison of measures, particularly since qualitative analysis is able to contextualize findings.

Nevertheless, the knowledge approach offers an opportunity to engage with the community in a deeper way, considering not who has the largest audience but what they are saying and how it may be received. These are theoretically important facets of the influence process.

Looking to the clustering coefficient ranking, almost every highly ranked user was considered an average account. Yet, average users appeared only once or twice among the top 20 most highly ranked for most other metrics. Since more than 20 accounts had the highest possible clustering coefficient (1), we randomly sampled the top users. No user ranked within the top 20 on any metric has a clustering coefficient of 1. The highest score found among the top ranked user given that knowledge was at approximately 0.6 in both the #CPC and #NDP networks. Since a clustering coefficient of 1 means that every follower that a user is connected to also follows every other of those followers, this is not surprising. Unless you follow only very few other users, it is hard to achieve a perfectly connected local community/neighborhood.

What this meant within the #CPC and #NDP networks is that users with very small, local communities who did not follow politicians, journalists, or other visible influentials made up the vast majority of users with a clustering coefficient of 1. When a user follows a very visible account, for example, a journalist, his or her clustering coefficient diminishes unless that journalist happens to be connected to everyone else in the user's direct network. This is unlikely because political elites, like journalists, tend to not follow many non-elite users. The alternate route to a fully connected neighborhood, assuming an undirected graph, would be for all of the users' connections to follow that same journalist. This, however, presents a theoretical challenge because it is assumed that influentials will have some form of access to information with which to develop opinion and then influence others who do not access that information (Katz & Lazarsfeld, 1955). Put simply, local influencers (like opinion leaders in Two-Step Flow work) are expected to follow elites. As such, the clustering coefficient of the broad network is not optimal for identifying influencers, whether they are the very visible political elite or the local influencers who are embedded in a community.

In our comparison of measures, we have noted that traditional measures of centrality tend to agree on how to rank influencers. These network-wide measures have identified political elites like politicians, media outlets, and journalists. Arguably, journalist @kady and politician @elizabethmay, found near the top of these lists in both #CPC and #NDP communities, are highly influential in this sense. Measures of interaction and other content-based metrics help identify political commentators and bloggers outside the traditional elite, for example, @ValckeNDP, who is an NDP staffer and appeared in the top ranking list based on knowledge in both communities. Although the clustering coefficient as applied to the full network did not identify individuals who appear likely to be influential, this does not mean that position in one's local network or the clustering coefficient are inconsequential. Given the theoretical importance of one's close personal ties and placement within a community, the following section investigates the potential use of the clustering coefficient.

The Local Context

Given the algorithm behind the clustering coefficient, the general characteristics of users following very visible accounts, and the fact that following visible accounts is crucial for potential local influencers, we reason the following: Those most likely to be local influencers will have lower clustering coefficients within the wider network because they follow very visible influentials. Simply, their clustering coefficients will be artificially low. Should those very visible influentials be removed, the clustering coefficients of likely local influencers will increase. Users whose clustering coefficient increases the most are most likely to be locally influential because they have access to information and are well positioned to disseminate that information to their local network.

We tested this hypothesis by removing very visible users and creating a derivative network. We chose to remove nodes based on whether their eigenvector centrality score was at or above the elbow of the distribution of all eigenvector centrality scores within the network.[4] Whereas indegree may remove most popular users, eigenvector centrality will remove popular users within the network who are in turn followed by popular users. In effect, this removes users who are followed by likely candidates for opinion leadership. The elbow of the distribution of eigenvector centrality values provides a readily interpretable cutoff point and is based on the dataset itself rather than an arbitrary figure.

A qualitative analysis of the 20 users whose clustering coefficient increased the most supports our reasoning. No users were political elite. We then separated average users who are likely to be influential from those less likely. Posting political content, mentioning the political elite, and having political conversations were all considered positive indicators of influence and were consistently found among this top 20 list, whereas it was not for the network-wide clustering coefficient list.

Although the results of our clustering coefficient analysis are promising, there remain methodological concerns. The clustering coefficient applied to the derivative network still favors users with small neighborhoods. They may be well positioned to influence locally but that locale may be quite small. It may be advisable to set a minimum indegree level as other studies of influence on Twitter have done (Cha & Gummadi, 2010). Although the clustering coefficient increases, this does not mean that the new clustering coefficient is high. In our case, only three users of our top 20 in the #CPC, and none in the #NDP, derivative networks had clustering coefficients below 1, whereas the others scored relatively high, above 0.83. In another network, this could be different. For example, if a network is very sparse, even the largest increase in the clustering coefficient could result in a value that is still low and indicates very little local social embeddedness. Finally, those with already high clustering coefficients are less likely to be found in the list of users whose score increased the most, even if they have a high or perfect score in the derivative network. This is justified because those with the highest original scores tend to be those who are not connected to elite players and do not access political information. Depending on the specific network, it is possible that users who occupy a middle ground could be ranked lower than is ideal.

The utility of the clustering coefficient for identifying local influentials is context dependent. As is the case with all the measures of influence we have used in this study, specific assumptions about which facets of the influence process are most important are embedded in operationalization. These assumptions have been justified based on theory and tested by considering users who are qualitatively influential.

Discussion

Typically, the most important facets of influence are assumed to be based on who follows a given user and how often they talk about that user or if the user is treated as an expert. Other facets of influence are routinely ignored. The role of expertise and knowledge and the importance of interpersonal interaction and personal connection are factors deemed theoretically relevant to the process of influencing someone (Katz & Lazarsfeld, 1955). It is not that having a following and being trusted, knowledgeable, and socially connected are in opposition, rather, it is a matter of placing more theoretical and operational importance on some facets over others. Although decisions about how to categorize users and which metrics are most appropriate must be made, using measures of influence out of context could lead to confusing or inaccurate results.

For example, we might assume that journalists and politicians topping most measures of influence, particularly the standard measures of centrality, are trusted experts. Indeed, @kady and @elizabethmay, two consistently highly rated accounts, are known for providing authoritative information, insight, and commentary on political issues. It is unlikely, however, that they fulfill the need for close personal interaction to help interpret information and actively convince someone to change his or her attitude, opinion, or behavior.

Another interesting question is raised when we consider the example of Liberal Leadership candidates within the #CPC and #NDP networks. Two candidates ranked as highly influential according to the number of times they were mentioned in both the #CPC and #NDP networks without appearing in any of the other lists of most influential.

Is it interaction with certain individuals or is it a user's position within a wider community that is more telling of a user's capacity to influence? Put differently, when we talk about a given "political community," are we discussing those who are most active only, or also those who may passively exist within its bounds?

From the perspective of the Two-Step Flow, if we think only of those who actively engage, we are already limiting ourselves to likely opinion leaders and public figures, both of which have been described as influentials (Katz & Lazarsfeld, 1955). We eliminate those most likely to be primarily followers when examining those who actively engage. Without including followers, the theoretical context shifts. Since we rank people in terms of influence, the top bracket retain their title as influential but with the bottom bracket eliminated, the middle group become the "followers" and their type of influence is lost. *Influential* becomes a simpler concept that can be useful, but it also means that we then lack clarity concerning the complexity of the social process that is influence.

With that said, it is not theoretically sufficient to take the list of users ranked by interaction and assign the top portion the title of public influential, bottom portion opinion leader, and any user not on the list follower. The notion of opinion leadership was seminal in the field of media studies and political communication because it connected theories of community and group dynamics to theories of mass media and political messaging. Social support and social pressure, applied by the opinion leader on his or her "everyday associates" (Katz, 1957), were the mechanisms through which influence happened. Opinions changed when someone in a group paid attention to a mass message and then used their position within that group to personally influence the other members (Katz & Lazarsfeld, 1955).

Thus, interaction within a network is indeed important, and having some following is necessary, but structural position within one's local neighborhood is also important. Whereas the majority of influence metrics overlook this factor, the use of clustering coefficient may be valuable.

We are not advocating for the clustering coefficient as a stand-alone measure of influence. No single measure we have assessed is sufficient for identifying the different kinds of influentials found within a political discussion network on Twitter. Influence is a contextualized phenomenon. Measuring communicator influence presupposes an ability to isolate the components of influence and weigh them accurately. The reason that some measures vary so greatly is that the components of influence are very different. Clear understandings of what these measures qualitatively represent can be used to help guide theory development and influential identification.

In summation, our study has used multiple measures of influence to identify the most influential members of the #CPC and #NDP Twitter communities. We have found that in terms of network placement, political elites such as media outlets, journalists, and politicians are most influential in each network. When interaction and content are considered both at the network level and globally, the political elite remain prominent but political commentators and bloggers are integrated into the lists of most influential. Finally, considering how socially embedded a user is within his or her local neighborhood, we are able to identify likely opinion leaders. Unlike our key journalists and politicians, who have network-wide patterns of influence, these opinion leaders influence those in their personal network. There are many opinion leaders, likely far more than we examined in our top 20 qualitative analysis. This presents an interesting avenue for future research.

Although specific to our case, these results are instructive for future studies of influence on Twitter and potentially other sites. As the ease with which we can trace interactions among users increases, we need to remain aware of how operational definitions can affect the theoretical context of our research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Elizabeth Dubois's DPhil (PhD) research was funded through a Clarendon Fund Scholarship (http://www.clarendon.ox.ac.uk/). This article was researched and written during the funding period.

Notes

1. For extended discussion of researcher data access on Twitter, see Gaffney and Puschmann (2014).
2. This is based on manual content analysis of a random sample of 10% of tweets in each network. Intercoder reliability was calculated in two ways following guidelines outlined in Lombard, Snyder-Duch, and Bracken (2002). Cohen's kappa was well above minimum acceptable levels at 0.8731 and 0.8942 for the #CPC and #NDP samples, respectively; agreement between coders was slightly less than 98% in both cases.
3. Cohen's kappa was 0.8737 and agreement between coders was slightly less than 90%, which is within an acceptable range for work of this nature (Lombard et al., 2002).
4. An elbow of a distribution is the point at which the value is furthest away from the expected linear decay. These scores, for #CPC and #NDP, respectively, were 0.1288 and 0.1311.

References

Agre, P. (2002). Real-time politics: The internet and the political process. *The Information Society*, *18*(5), 311-331.

Bakshy, E., Hofman, J. M., Watts, D. J., & Mason, W. A. (2011). Everyone's an influencer: Quantifying influence on Twitter categories and subject descriptors. In *WSDM '11 Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (pp. 65-74). *Hong Kong, China: WSDM*. doi:10.1145/1935826.1935845

Brosius, H.-B., & Weimann, G. (1996, October). Who sets the agenda: Agenda-setting as a two-step flow. *Communication Research*, *23*(5), 561-580.

Cha, M., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. *ICWSM*, *10*, 10-17.

Chadwick, A. (2011, January). The political information cycle in a hybrid news system: The British prime minister and the bullygate affair. *The International Journal of Press/Politics*, *16*(1), 3-29.

Conover, M., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the political alignment of Twitter users. In *Proceedings of the 3rd IEEE Conference on Social Computing*. Retrieved February 20, 2013, from IEEE Xplore.

Dahlgren, P. (2005). The internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, *22*(2), 147-162.

Dillard, J., Segrin, C., & Harden, J. (1989). Primary and secondary goals in the production of interpersonal influence messages. *Communication Monographs*, *56*, 19-38.

Dubois, E., & Dutton, W. H. (2012). The fifth estate in internet governance: Collective accountability of a Canadian policy initiative. *Revue Francaise d'Etudes Americaines*, *131*(1), 81-97.

Duong, K. (2012). *Canada digital future in focus 2012*. Available from http://www.comscore.com/2012CanadaDigitalFutureinFocus

Gaffney, D., & Puschmann, C. (2014). Data collection on Twitter. In A. Bruns, K. Weller, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 55-68). New York, NY: Peter Lang.

González-Bailón, S., Borge-Holthoefer, J., & Moreno, Y. (2013, March 8). Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist*. Advance online publication. doi:10.1177/0002764213479371

Gruzd, A. (2012, September 20-21). *Investigating political polarization on Twitter: A Canadian perspective*. Paper presented at Internet, Politics, Policy 2012: Big Data, Big Challenges?, University of Oxford, Oxford, UK.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an imagined community. *American Behavioral Scientist*, *55*(10), 1294-1318.

Hindman, M. S. (2009). *The myth of digital democracy*. Princeton, NJ: Princeton University Press.

Huckfeldt, R. R., Sprague, J., Kuklinski, J. H., Wyer, R. S., & Feldman, S. (1995). *Citizens, politics and social communication: Information and influence in an election campaign*. Cambridge, UK: Cambridge University Press.

Katz, E. (1957, March). The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, *21*(1), 61-78.

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Piscataway, NJ: Transaction Publishers.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600). *New York, NY: ACM*.

Lee, C., Kwak, H., Park, H., & Moon, S. (2010). Finding influentials based on the temporal order of information adoption in Twitter. In *Proceedings of the 19th International Conference on World Wide Web* (p. 1137). *New York, NY: ACM*.

Lerman, K., & Ghosh, R. (2010). Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM)*. *Washington, DC: ICWSM*.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, *28*(4), 587-604.

Norris, P., & Curtice, J. (2008, May). Getting the message out: A two-step model of the role of the internet in campaign communication flows during the 2005 British general election. *Journal of Information Technology & Politics*, *4*(4), 3-13.

Rattanaritnont, G., Toyoda, M., & Kitsuregawa, M. (2012). Characterizing topic-specific hashtag cascade in Twitter based on distributions. *Web Technologies and Applications*, *7235*, 735-742.

Richards, T., & Richards, L. (1995). Using hierarchical categories in qualitative data analysis. In U. Kelle (Ed.), *Computer-aided qualitative data analysis* (pp. 88-95). Thousand Oaks, CA: SAGE.

Rogers, E. M. (2010). *Diffusion of innovations* (4th ed.). New York, NY: Free Press.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development*, *52*(1), 5-18.

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009). Tweet the debates: Understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM Workshop on Social Media* (pp. 3-10). *New York, NY: ACM*.

Small, T. A. (2011). What the hashtag? A content analysis of Canadian politics on Twitter. *Information, Communication & Society*, *14*(6), 872-895.

Sousa, D. (2010). Characterization of the Twitter @ replies network: Are user ties social or topical? In *SMUC '10 Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents* (pp. 63-70). *Toronto: CIKM*.

Subbian, K., & Melville, P. (2011). Supervised rank aggregation for predicting influencers in Twitter. In *Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk, and Trust (PASSAT)* (pp. 661-665). Retrieved March 3, 2013, from IEEE Xplore.

Watts, D., & Dodds, P. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, *34*(4), 441-458.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 409-410.

Weimann, G. (1982). On the importance of marginality: One more step into the two-step flow of communication. *American Sociological Review*, *47*(6), 764-773.

Welser, H., Smith, M., Fisher, D., & Gleave, E. (2008). Distilling digital traces: Computational social science approaches to studying the internet. In L. R. Blank Grant & N. Fielding (Eds.), *SAGE handbook of online research methods* (pp. 116-140). Thousand Oaks, CA: SAGE.

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 705-714). *New York, NY: ACM*.

Yin, R. (2008). *Case study research: Design and methods (applied social research methods)* (4th ed.). Thousand Oaks, CA: SAGE.

Author Biographies

Elizabeth Dubois is a doctoral candidate and Clarendon Scholar at the Oxford Internet Institute, University of Oxford. Her research focuses on the ways in which citizens use technology to exert political influence.

Devin Gaffney is a senior developer at the online influencer discovery platform Little Bird in Portland, Oregon. A graduate of the Oxford Internet Institute, he co-founded 140kit, an open source Twitter research platform, and is an incoming PhD candidate at Northeastern University concentrating in Network Science. His principal research interests focus on geodata, data access, privacy, and how to make a great frittata.