

Portland State University

PDXScholar

Engineering and Technology Management
Faculty Publications and Presentations

Engineering and Technology Management

2019

Do Directionality and Network Size Affect Network Structure in Online Social Networks?

Nitin V. Mayande

Portland State University

Charles Weber

Portland State University, webercm@pdx.edu

Follow this and additional works at: https://pdxscholar.library.pdx.edu/etm_fac



Part of the [Communication Technology and New Media Commons](#), and the [Social Media Commons](#)

Let us know how access to this document benefits you.

Citation Details

N. Mayande and C. Weber, "Do Directionality and Network Size Affect Network Structure in Online Social Networks?," 2019 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, 2019, pp. 1-18.

This Article is brought to you for free and open access. It has been accepted for inclusion in Engineering and Technology Management Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Do Directionality and Network Size Affect Network Structure in Online Social Networks?

Nitin Mayande¹, Charles Weber²

¹ Tellagence Corporation, Portland, Oregon, USA

² Dept. of Engineering and Technology Management, Portland State University, Portland, Oregon, USA

Abstract—A study of the online social networks of six Twitter conversations about six YouTube product categories reveals that directionality and network size affect the structure of online social networks. Our results indicate that large networks tend to be non-random, regardless of whether they are directional or not, suggesting that structural attributes of the online networks under study are a true reflection of network's features. Smaller non-directional networks also tend to be non-random, whereas smaller directional networks tend to be random in nature. However, very small networks tend to be random in nature, whether they are directional or not.

Our results suggest that larger online networks undergo different generation mechanisms than smaller real-world networks, especially if these networks are directional. Extant theory, which is almost exclusively derived from observation of real-world networks, may thus not adequately describe the behavior of online networks. We propose research to remedy this deficiency at the end of this paper.

I. INTRODUCTION

Traditional marketing models are swiftly being upended by the advent of online social networks (*Deighton [45]*). Yet, practicing firms that are engaging with online social networks neither have a reliable theory (*Mayande [76]*) nor enough practical experience (*Wiertz, et al. [131]*) to make sense of the phenomenon. Extant theory is based on observations of the real world and may thus not apply to online social networks (*Weber and Mayande [121]*). Practicing firms may consequently be misallocating a large amount of resources (*Edwards [52]*), simply because they do not know how the online social networks with which they interact are organized (*Aral, et al. [10]*, *Cha, et al. [40]* [41], *Li and Bernoff [71]*, *Mayande [76]*, *Weber and Mayande [121]*).

Practitioners in the online social media space are currently interacting with networks of unprecedented size. They may consequently encounter phenomena that have never been observed in the real world. Furthermore, propagating messages by 'word of mouth' online may utilize different mechanisms than doing so literally in the real world. Online conversations could be more reciprocal (bi-directional) than their counterparts in the real-world, or less so. In either case, directionality of information flow could affect network structure in online networks in ways that have hitherto not been observed. It has thus been argued that scale and directionality constitute two

attributes of social networks that have not been studied sufficiently in the online context (*Mayande [76]*).

The purpose of this paper is to investigate how online social networks that are in stark contrast to real-world social networks get organized. In particular, we (the authors of this paper) explore how scale and directionality affect the structure of online social networks. We test the following null hypothesis: Directionality and scale do not impact the structure of online social networks.

To test this null hypothesis, we have collected retrospective data from Twitter conversations about six YouTube product categories (Music, Entertainment, Comedy, Science, Howto and Sports) in continuous time for a period of three months. These Twitter conversations vary considerably in scale, as does the popularity of the product conversations that comprise the subjects of the conversations (*Mayande & Weber [79]*). We have calculated the clustering coefficients for the networks that pertain to each of these conversations and compared them to the clustering coefficients of random networks of equivalent size and density (*Erdős and Rényi [53]*). If the clustering coefficients of the network under observation are equal to naught or to the clustering coefficient of the random network, then the network under observation was generated at random. If the clustering coefficients of the networks under observation neither equal naught nor the clustering coefficients of their corresponding random networks, then networks under observation are not random and the clustering coefficients are a true reflection of network's features.

The results of our study indicate that large networks tend to be non-random, regardless of whether they are directional or not. Smaller non-directional networks also tend to be non-random, whereas smaller directional networks tend to be random in nature. However, very small networks (define very small) tend to be random in nature, whether they are directional or not. Thus, the null hypothesis had to be rejected.

Our results suggest that scale and directionality affect network structure. Larger networks undergo different generation mechanisms than smaller networks do, especially if these networks are directional. However, these generation mechanisms could not be characterized in this study. Given the importance of online social networks to modern commerce, we propose in depth, follow-on research of this subject matter at

the end of this paper. We also discuss how our findings potentially contribute to a structural theory of online social networks.

II. BACKGROUND

Extant theory of social networks was developed almost entirely from observing social networks in the real world (e.g., Bailey [13], Luhmann [74], Miller [81], Parson, [90]), where practitioners track the deliberate propagation of information from one user to another through word of mouth (*Granovetter* [60], Rogers [98], Tichy, et al. [105]). These *network flows* (Aral & Walker [11] [12], Burt et al. [36], Dellarocas et al. [46], Hodas and Lerman [63], Weber and Mayande [121]) take place between the source of information and the seeker of information, and they transpire exclusively within existing social relationships (*Bristor* [30], Burt [32], Burt and Doreaian [35], Duhan et al. [51], Money et al. [84]), which Weber and Mayande [121] have defined as *network structure*. Network structure determines the boundaries of communities ([13] [74] [81] [90] [121]), and it guides the network flows within communities ([13] [74] [90]). Network flows, in turn, enable network phenomena such as social capital (*Bourdieu* [27], Burt [33] [34], Coleman [43], Putnam [97]), social behavior (*Allen* [6], Burt [31], *Granovetter* [60]), economic benefit ([6] [27] [31] [33] [39] [43] [60]) and social influence, which is exerted when an actor adapts his/her behavior to the behavior of other actors in the community (*Cartwright* [39], *March* [75], *Simon* [101]).

A. Topological Organization of Social Networks

A variety of theories have attempted to explain the topological organization of social networks. These include the theory of social systems, autopoietic theory, living systems theory and social entropy theory. All these theories explain the organization of network structure (groups, societies, organizations, countries, etc...) through *network flows* (processes like communication, collaboration, reproduction, coordination, control, etc...) and through constraints that impact network structure and network flow (geographic distance/boundaries, land availability etc...).

The theory of social systems was initially proposed by Talcott Parson in 1951 [90]. The author advocated a functionalist approach and hypothesized that all social systems perform the following basic functions:

1. *Adaptation*: acquiring enough resources
2. *Goal Attainment*: setting and achieving goals
3. *Integration*: maintaining coordination amongst sub-systems.
4. *Latency*: creating, preserving and propagating systems distinct culture and values.

Parsons states that a social system comprises one of the three aspects of structuring a completely concrete system of social action [90]. The other two are the “personality system” of the individual actors and the “cultural system,” which is built into the individuals’ actions [90].

In the theory of social systems [90], individuals do not act as the fundamental units of society. Instead, society is based on the actions out of which personality systems and cultural systems are built. Therefore, the theory of social systems does not treat the personality systems and the cultural systems independently. Instead, it is concerned with how these components of the social system affect the overall structure of the social system and how it functions. The theory analyzes social processes in relation to the structure of social systems and their variability. It describes the mechanisms of socialization, patterns of orientation in social roles, tendencies toward deviant behavior and mechanisms of social control.

Niklas Luhman applied autopoietic theory to social systems and suggested that social systems use communication as their mechanism for autopoietic reproduction [74]. Communications are not living units; they are not conscious units; and they are not actions. A unit of communication consists of a synthesis of three components: information, utterance and understanding (including misunderstanding). Essentially, every actor within the social system must make three choices: 1) whether to accept or reject information; 2) understand (or perhaps misunderstand) the information; and 3) and propagate it to other actors. The synthesis that results in communication is produced by the network in which the communication takes place; it is not derived from an inherent power of consciousness or from the inherent quality of the information. In addition, the synthesis of information, utterance and understanding cannot be preprogrammed by language. It must be recreated from situation to situation by referring to previous communications and to the possibility of further communications. In every situation, communication is restricted by the actual event, requiring self-reference. Furthermore, information, utterance and understanding cannot reside independently in a system; they are inherently co-created.

Living systems theory (*Miller* [81]) is a general theory about how living systems work. It deals with the notion of emergence and interaction. A system is defined as a set of interacting units and the relationships among them. Miller’s model of living systems constitutes a hierarchy that consists of the following eight levels:

- *Cells*: the basic building block of life
- *Organs*: the principle components are cells, organized in simple, multi-cellular systems.
- *Organisms*: there are three kinds of organisms: fungi, plants and animals. Each has distinctive cells, tissues and body plans and carries out life processes differently.
- *Groups*: these contain two or more organisms and their relationships.
- *Organizations*: these involve one or more groups with their own control systems for doing work.
- *Communities*: these include individual persons and groups, as well as groups which are formed and are responsible for governing or providing services to them.
- *Societies*: these are loose associations of communities, with systematic relationships between and among them.

- *Supranational systems*: organizations of societies with a supra-ordinate system of influence and control.

The properties (behavior) of a system emerge from the interaction between the components that comprise the system. Regardless of their complexity, they each depend upon the same essential twenty subsystems that perform specific processes, in order to survive and to continue the propagation of their species or types beyond a single generation. The twenty subsystems and the processes of all living systems are arranged by input-throughput-output processes. Some of these processes deal with material and energy for the metabolic processes of the system. Other subsystems process information for the coordination, guidance and control of the system. Some subsystems and their processes are concerned with both. [81]

Social Entropy Theory (SET) (Bailey [13]) uses the system's internal entropy level as an indicator of system state, where entropy is a measure of system disorder. Entropy can show up in the system as various indicators of system disorder, such as faulty communication, errors, inadequate supply levels, lack of energy, resources, or even clutter. If entropy gets too high, the functionality of the system is impaired or even threatened. From the standpoint of SET, entropy can best be properly managed by a self-steering process, where the chief goal of self-steering is to keep system entropy levels from getting too high.

B. Network Flows in Social Phenomena

Theories of social phenomena show that social phenomena within a social network are caused by network flows. These theories, in contrast to theories of social organization, do not address the structure of a social network. Instead, they try to explain a specific social phenomenon within the broader context of social networks that occur in the real world.

In his book *Diffusion of Innovations* [98], Everett Rogers describes the process of adoption of new innovations. He emphasizes the role of interpersonal communication in the adoption of innovations. According to Rogers, diffusion is "the process in which an innovation is communicated through certain channels over time among the members of a social system" ([98], p. 5), the key components in this definition being innovation, communication channel, time and social system.

For Rogers, "diffusion is a very social process that involves interpersonal communication relationships" ([98], p.19). He defines communication as "a process in which participants create and share information with one another, in order to reach a mutual understanding" ([98], p. 5). This communication occurs through channels between sources. Rogers defined a source as "an individual or an institution that originates the message and an interpersonal channel consists of two-way communication between two or more individuals through which the message gets to the receiver" ([98], p. 204). These interpersonal channels are powerful enough to create or change strong attitudes held by an individual.

Rogers defined the social system as "a set of interrelated units engaged in joint problem solving to accomplish a common

goal" ([98], p. 23). Since diffusion of innovations takes place in the social system, it is influenced by the social structure of the social system. For Rogers, structure is "the patterned arrangements of the units in a system" ([98], p. 24). He further claimed that the nature of the social system affects individuals' innovativeness, which is the main criterion for categorizing adopters into innovators, early adopters, early majority, late majority and laggards ([98], p. 22).

Granovetter [60] asserts that acquaintances (weak ties) are less likely to be socially involved with one another than close friends (strong ties). Thus, the set of people made up of any individual and his or her acquaintances comprises a low-density network (one in which many of the possible relational lines are absent), whereas the set consisting of the same individual and his or her close friends will be densely knit (many of the possible lines are present).

Burt [31]-[34] proposes the construct of structural holes, which suggests that social capital is created by a network in which people can broker connections between disconnected network segments. He views society as a network in which people or groups of people can exchange all types of goods and ideas in order to achieve their goals. Some of these people or groups of people achieve better returns in lieu of their efforts than others do. For example, some people earn a better remuneration, some become more important and some lead more important projects. The human capital explanation of this inequity is that people who do better are more able people, more intelligent, more articulate, more attractive or more skilled. Social capital is a contextual complement of human capital, suggesting that people who are better connected should be more successful. Thus, holding a specific position in the network structure is associated with a certain level of social capital.

Burt [32] defines structural holes as weaker connections between two groups in a social structure. These holes in the structure create competitive advantage for the people who have relationship that span these holes. This does not mean that the people in each group are unaware of the existence of the other group. Instead, the people in each group are more focused on their own activities and do not participate in the activities of the other group. Thus, structural holes are an opportunity to broker and control the flow of information across groups [76].

Coleman's network closure argument [43] suggests that networks in which everybody is connected to everybody and no one can escape notice of the other (in other words dense networks) are the source of social capital. He defines social capital as a resource for action within a social structure [43]. Network closure does two things for people in a network. First, it affects access to information. Second, network closure facilitates collective sanctions, and fear of sanctions for behavior that is out of the norm fosters conformity. It also reinforces trust between those who already conform.

The theories of social phenomena described above identify a social phenomenon within a network and explain the phenomenon within the broader context of a social network that exists in the real world. These theories do not attempt to explain

the organization of the social network. In all instances, the social phenomena under observation within a network structure are caused by network flow. For example, Rogers [98] talks about the importance of interpersonal communication within a social system for diffusion of innovation. Granovetter [60] suggests that weak ties are the sources of new information that flows into the network from the outside. In his structural holes theory [31]-[34], Burt talks about competitive advantage being derived by creating network flows between two different cliques. This suggests competitive advantage is obtained from being on the fringe of a network (Burt [31]-[34]). In contradiction to Burt, Coleman [43] talks about the advantage of being in the middle of network flows within a clique and the risks of being on the fringe of a network. In summary, theories of social phenomena are different elaborations of the impact of network flow on social phenomena.

C. Real World versus Online

A key factor that differentiates online social networks from their real-world counterparts is scale (Weber and Mayande [121]). Social networks in the real world tend to consist of fewer than 1000 people (e.g., Burt [32], Granovetter [60], Rogers [98], Tichy et al. [105]), whereas the scale of online networks is, in principle unlimited; they may contain hundreds of thousands or millions of people (Dodds et al. [47] [48], Mislove et al. [82]). Very large online networks may thus be driven by processes that are not observed in real-world networks. Conversely, real-world networks may be driven by social processes that cannot be found in online networks. Scale could consequently be a factor that influences network structure [121].

In most analyses of real-world social networks, relationships have been treated as reciprocal (e.g., Burt, [31], Burt and Doreian [35], Granovetter, [60]), i.e. we assume that a person that propagates information to another also receives information from that person. However, directionality has been a factor in some studies (e.g., Allen [6], Rogers [98]) because social relationships are not necessarily reciprocal. This is especially true online, where a few highly influential people act as opinion leaders for large communities (e.g., Weber, Hasenauer and Mayande, [119] [120]), and the structure of an online social network in which directionality is considered may vary significantly from one where it is not.

III. RESEARCH METHODS

A. Research Design

The methods deployed in this study approximately follow those of prior investigations of dynamic networks (e.g., Mayande and Weber [79]). We have chosen Twitter conversations about six YouTube product categories as the unit of analysis of our study for a variety of reasons. First, the selected YouTube product categories allow us to advantage of natural discrepancies in scale. In addition, Twitter is the only social media platform that can capture changes in the context and content of online conversations at the rate at which they

occur. Furthermore, all data on Twitter are available in the public domain. Finally, Twitter is popular enough for it to cover enough conversations to enable a comprehensive analysis of the product categories under study (e.g., Huberman et al. [65]). As early as 2014, Twitter received almost 190 million unique visits every month (Alexa [7]), which makes it the eighth most popular website in the world, and over 1 billion tweets were generated on Twitter every 5 days (Statisticbrain [102]). All product categories under study are sufficiently mature to avoid any bias associated with startup effects [79]. They are also not in rapid decline; thus, any bias pertaining to rapid decay of the social network under observation can be avoided [79].

B. Data Collection and Filtration

The study described in this paper utilizes the dataset from [79], a retrospective population study for which data were collected in continuous time. Under these circumstances, the number and sequence of events, as well as the duration between them, can be calculated. The main advantage of this approach lies in the greater detail and precision of information (Blossfeld and Rohwer [18]). It also reduces time required to collect data, and it enhances the chances of recognizing the overall patterns (Leonard-Barton [70]). In addition, modern data extraction capabilities on the Internet enable the study of whole populations. “This approach not only eliminates sample selection bias; it also ensures that the results observed are valid and generalizable to the entire population under study. This is especially important in studies that involve networks, as selecting only a sample instead of the population can break a network into multiple small networks (Goggins and Petakovic [58]), which can lead to faulty results. Furthermore, the data collection method deployed in this study allows us to extract large amount of data from which statistically significant conclusions can be drawn” (Mayande and Weber [79], p. 3).

Twitter data are readily available through application programming interfaces (API’s) from which the networks forming within a context can easily be deduced. For the sake of simplicity, we use keyword searches as a means of finding contextual networks (Jansen, et al. [66]). Both the Twitter platform as a data source and keyword search as data filter have been used in previous studies [66] [79] [132].

As mentioned in Mayande and Weber [79], p. 3, “data were gathered for a period of three months (a total of 91 days), from December 31st, 2013 to March 31st, 2014, in order to control for any monthly periodicity in the data (Gonçalves and Ramasco [59], Meiss, et al. [80]). The time period for data collection was chosen at random. The data have been analyzed in daily intervals, in order to capture tweet volatility patterns caused by daily routine (Dodds et al., [48]). (For example, Twitter users in Tokyo tweet a lot less during working hours (Gigaom [57]).) The 24 hours started in accordance with Greenwich Meridian Time (GMT).”

We collected the cumulative sum of the daily volume of tweets that were associated with a topic of interest (for example, YouTube + Music). However, only the tweets with an @mention or a re-tweet were included in the sample, because

these tweets could be associated with the engaged activity pertaining to a topic. Of this set of ‘engaged tweets’ we only kept the set of ‘community tweets’ in the sample. These tweets were associated with the largest network engaged in the collective conversations within a topic. Finally, we have filtered out the redundancies generated by multiple tweets per day. Participants are only counted once as a ‘daily unique’, even if they tweet more than once on a specific day. Thus, the largest community of a network consists of the cumulative sum of daily unique people associated with the largest network engaged in the collective conversation within a topic. [79]

C. Variables and Measures (Wassermann and Faust [109])

1) Networks, Vertices and Edges

In this study, anyone that has been active in the largest community during the 91-day period of study constitutes a vertex. Any @mention or retweet that occurred between any two vertices V_i and V_j in the largest community would denote an edge between vertex V_i and V_j . We can then let the graph G be a network such that

$$G = (V, E), \quad (1)$$

where

V is a finite and non-empty set of vertices;

$V = \{1, 2, 3, \dots, N\}$;

V determines the size of the network; and

E is a finite and non-empty set of edges.

Therefore,

the edge $(i, j) \in E$ is incident with vertices V_i and V_j ;

$(i, j) \in E$ is a link, if $i \neq j$; (2)

$(i, j) \in E$ is a loop, if $i = j$. (3)

2) Adjacency Matrixes

If two vertices are incident with the same edge, then they are adjacent or neighbors. We can then define the $N \times N$ adjacency matrix $A = (a_{ij})$ by setting a_{ij} equal to 1, if $(i, j) \in E$, and to 0, if it is not. Therefore, the adjacency matrix is a matrix representation of a graph that displays the connectivity of the graph. The rows and columns of the graph are labeled by the vertices. If there is an edge between two vertices, then the edge is indicated in the matrix as 1; otherwise the link takes the value of 0. This is also the first order adjacency matrix, i.e., it defines vertices that are connected directly. The first order adjacency matrix does not define relations that are not direct. For that, a higher order of adjacency matrices is required, which can be achieved by the multiplying the first order adjacency matrix with itself. For example, to identify vertices that have just one vertex between them, a first order adjacency is multiplied with itself. The resultant matrix is called the second order adjacency matrix.

Similarly, let “A” be an $N \times N$ adjacency matrix; then a degree matrix “D” is a second order adjacency matrix in which

all the elements except the diagonal elements are non-zero. Then the second adjacency matrix is given by

$$A^2 (\alpha_{i,j}) = A^1 * A^1 \quad (4)$$

Hence, an $N \times N$ degree matrix (D) is given by

$$D = A^2, \quad (5)$$

where

$$\alpha_{i,j} = \alpha_{i,j}, \quad (6)$$

iff $i=j$ and $(i,j) \in A^2$,

and

$$\alpha_{i,j} = 0, \quad (7)$$

iff $i \neq j$ and $(i,j) \in A^2$.

3) Clustering Coefficient

Clustering is a typical property of acquaintance networks, where two individuals with a common friend are likely to know each other [109]. Watts and Strogatz [111] describe the clustering coefficient in the context of social networking as the degree to which the vertices in the graph cluster together. Newman et al. [87] describe the clustering coefficient to be same as the transitivity of a graph and defined it as follows:

$$C(G) = [3 * \Delta(G)] / \tau(G), \quad (8)$$

where $C(G)$ denotes the clustering coefficient of the graph, $\Delta(G)$ is total number of triangles in the graph, and $\tau(G)$ represents the total number of connected triples in the graph.

The total number of triangles is calculated by letting A^3 denote third order adjacency matrix of a graph. The diagonal elements of A^3 contain elements that start from vertex V_i and, after passing through 2 other vertices, end at the same vertex V_i . This can happen only if it is triangle. The diagonal element counts each triangle 3 times. Example: triangle ijk is counted i to j to k and i to k to j . Thus, every triangle is counted 6 times, and

$$\Delta(G) = \frac{1}{6} \sum_{(i,j)=1}^n A^3_{i,j} \dots \dots \dots (9)$$

where $i = j$.

To calculate the total number of connected triples, let A^2 represent the second order adjacency matrix of a graph. The elements of A^2 contain elements that start from vertex V_i and after passing through 2 other vertices ends at the same vertex V_i or any other vertex in the network V_j . These are called connected triples. Thus, every connected triple is counted 4 times. Therefore,

$$\tau(G) = \frac{1}{4} \sum_{(i,j)=1}^n A^2_{i,j} \dots \dots \dots (10)$$

where $i \neq j$.

4) Density

The graph density (D) measures the fullness of a graph. It looks at all the existing edges in the graph (E) and compares

them all to the all the possible edges in a graph (E_T) (Wasserman and Faust [109]). Therefore,

$$D = E / E_T. \quad (11)$$

5) Reciprocity

Reciprocity is an important characteristic of directed networks, which helps quantify tendency of vertex pairs to form mutual connections with each other (Newman et al. [87]). The reciprocity R is a ratio of the quantity of bi-directional edges (E_{i,j_2dir}) in the network to the quantity of non-bi-directional edges (E_{i,j_n2dir}) in the network. Therefore,

$$R = E_{i,j_2dir} / E_{i,j_n2dir}. \quad (12)$$

6) Control Variable: Network Size

TABLE 1: SIZES OF TWITTER CONVERSATIONS OF THE LARGEST COMMUNITIES AND THEIR ASSOCIATED YOUTUBE PRODUCT CATEGORIES. [79]

<u>Product Category</u>	<u>Community Tweets</u>	<u>Largest Community</u>
Music	1,586,149	1,456,770
Sports	32,778	29,998
Comedy	25,624	24,555
Science	22,598	21,277
Entertainment	16,365	15,822
Howto	4,299	4,203

Table 1 illustrates why the number of tweets in the conversations of the largest community of the six selected YouTube product categories constitutes a control variable (Table 1 covers the complete 91-day period of observation). The largest community in Music, the biggest conversation in the sample, is more than an order of magnitude larger than largest community of Sports, the next biggest conversation in the sample, and more than two orders of magnitude larger than the largest community in Howto, the smallest conversation in the sample. If the clustering coefficients of these largest communities vary as a function of scale, then this variation should become clearly discernable.

It should be noted that online social networks can be highly dynamic. The degree of activity can vary by orders of magnitude from day to day; network structure and the information flow within the network can undergo dramatic daily changes as well. For example, in the ‘‘Music’’ category, the maximum of the total number of daily tweets and the maximum of the total number of daily unique people observed on a single day (the daily uniques) are 62,380 and 59,666, respectively. Similarly, the minimum of the total number of daily tweets and the minimum of the number of daily uniques are 19,700 and 18,333, respectively. The size of the largest community on a particular day and the largest number of community tweets on that day also seem to follow the trend of total people and total tweets. The largest number of daily community tweets and the largest number of daily unique people are 48,720 and 47,630, respectively. Similarly, the smallest number of daily community tweets and the smallest number of daily unique people are 10,830 and 10,324,

respectively. Thus, we cannot simply take the mean and standard deviation of the community tweets across the 91-day period of observation and assume that reflects network behavior. Instead, we chose to analyze the networks under study on a daily basis.

D. Hypothesis Testing in Networks

In hypothesis testing, it is important to minimize the odds of an α -error (Type-1 error) or a β -error (Type-2 error), which Montgomery [85] defined as follows:

1. α -error: The study results lead to the rejection of null hypothesis even though it’s true.
2. β -error: The study results lead to the acceptance of null hypothesis even though it’s false.

In this study, an α -error would consist of concluding that scale or directionality affect network structure even though this is not the case. Concluding that scale or directionality do not affect network structure even though it does would constitute a β -error.

In principle, one can compute any network measure for any network that is built on empirical data. Many conclusions can be drawn based on these network measures. Unfortunately, one cannot be confident that the network measure that has been computed is a true reflection of the network’s structural features or a random variation. To overcome this predicament, Erdős and Rényi [53] proposed comparing the network and the network measures of the network in question to the network and the network measures of a randomly generated network with same number of vertices and edges such that every edge is chosen with equal probability. (Kejzar et al. [67] used such networks as the basis for modeling the dynamics of acquaintanceships. Donninger [49] deployed this approach to derive the distribution of degree centralization. Anderson et al. [9] used it to simulate the distribution of degree centralization and betweenness centralization.)

The deployed method is very similar to testing a mean with a z-test. In a z-test, a sample of data is taken where the value of each data point is considered a random value. One single mean is calculated for the sample and every random value is compared with this mean. The mean in question is the expected value. If the random value is different from the expected value, then one rejects the hypothesis that the true mean is equal to expected mean.

In case of a network, the random values are the edges. A single network is observed, and the network measures are calculated for this network. These network measures are compared with the expected network measures of a network in which every edge has an equal probability. If the expected and the random network measures are different, then one can conclude that random network has different characteristics than the expected network. If the network measure is non-existent for the network in question, then the network is considered random.

This method also prevents one from drawing the wrong conclusions because of a lack of reference point. For example, let's consider a network A with a clustering coefficient of 0.25 and a network B with a clustering coefficient of 0.5. It is easy to conclude that network B is highly clustered when compared to network A, because 0.5 is greater than 0.25. But it may be that network A has a higher clustering coefficient than one would expect in a random network, whereas network B has a clustering coefficient that is same as that of a random network. As a result, one would consider the clustering coefficient of network A as a true network feature, whereas the clustering coefficient of network B could be considered equivalent to what one might observe in a random network.

To perform meaningful comparisons of a network measure and to eliminate α - error and β - error, we generate for every network under study (Music, Sports, Comedy, Entertainment, Science and Howto) an Erdős-Rényi (E-R) random network with the same number of vertices, the same number of edges and the same density. We subsequently compare the clustering coefficients of the undirected and directed networks under study to their corresponding random E-R networks. If the clustering coefficients of the undirected and directed (CC_ud, CC_d) networks are equal to those of the corresponding E-R random network (CCudran, CCdran), then the directed and undirected networks are considered random. If they are not equal, then they are not random, and the computed variables are a true reflection of network's features. If CC_ud or CC_d equal naught, then the networks with which these cluster coefficients are associated are also considered random.

To further elucidate, consider the "Howto" community that formed on January 6th, 2014. The directed network, which is displayed in fig. 1a, has 15 vertices and 15 directed edges (self-edges, i.e. edges where both ends connect to one vertex, are ignored). In the undirected version of the network, which is displayed in fig. 1b, all edges are considered symmetric. Therefore, there are 30 edges in the network in fig. 1b, two for each connection that is displayed.

In the undirected network, in which all relationships are bi-directional, the information flows beyond the connected vertex. There are reciprocal relationships and instances where two different vertices connected to a vertex and exchange information with each other. The clustering coefficient of the undirected network in fig. 1b (CC_ud) is 0.07894. The undirected network still needs to be compared to an equivalent E-R random undirected network (CCudran) to ascertain if it's random or not.

In the directed version of the network, there is no transitivity, i.e. the information only moves from a vertex to the connected vertex in a single direction. The information does not go beyond the connected vertex. There are no reciprocal relationships or instances where two different vertices connected to a vertex exchanging information with each other. The clustering coefficient (CC_d) of the network equals 0. Therefore, the directed network in fig. 1a is a random network.

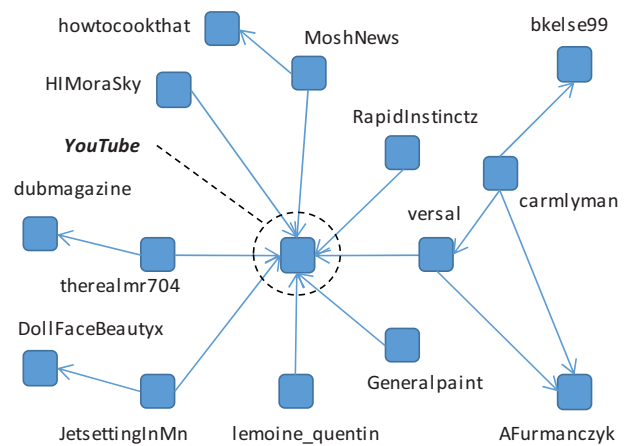


Fig. 1a. The directed "Howto" network on Jan. 6th, 2014.

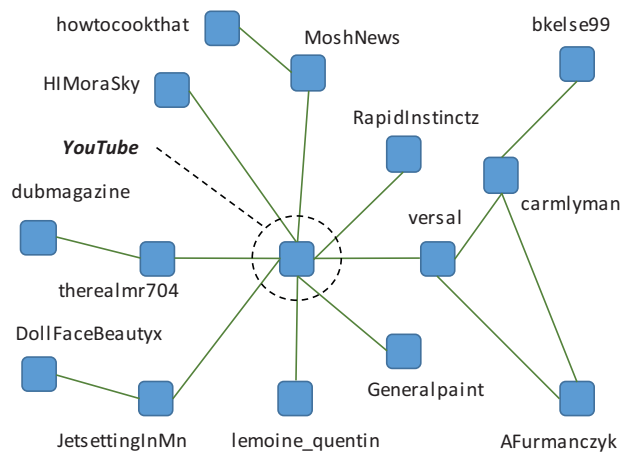


Fig. 1b. The undirected "Howto" network on Jan. 6th, 2014.

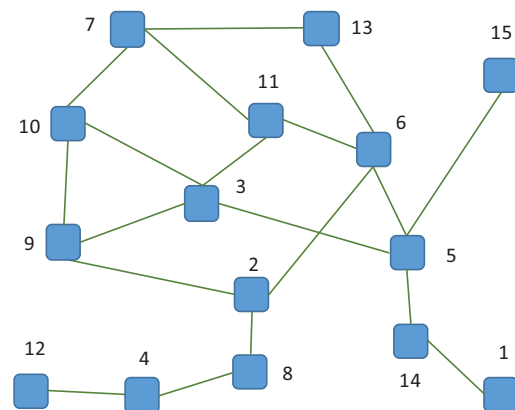


Fig. 2. Equivalent E-R network of the undirected network in fig. 1b.

Fig. 2 displays the equivalent E-R random of the undirected network fig. 1b. Fig. 2 illustrates that there are instances where two different vertices connected to one vertex and exchange information with each other. The clustering coefficient (CCudran) of the random undirected network in fig. 2 is 0.133333, which is higher than CC_ud. Therefore,

comparing the clustering coefficients of the undirected network in fig. 1b to that of its equivalent random undirected network clearly illustrates that the undirected network is not a random network.

TABLE 2. NUMBER OF DAYS (OUT OF 91) FOR WHICH THE CLUSTERING COEFFICIENTS OF THE LARGEST COMMUNITY EQUALS OR DOES NOT EQUAL 0.

Product Category	Un-directed		Directed	
	CC_ud≠0 (days)	CC_ud=0 (days)	CC_d≠0 (days)	CC_d=0 (days)
Music	91	0	91	0
Sports	79	12	2	89
Comedy	76	15	2	89
Science	67	26	0	91
Entertainment	45	46	2	89
Howto	7	84	3	88

IV. RESULTS

Table 2 and fig. 3 display the results of our study. Table 2 denotes the number of days (out of 91) for which the clustering coefficients of the largest community that discusses each product category equals or does not equal naught, and it does so for the undirected and directed networks. Fig. 3 graphs the percentage of days (out of 91) that the clustering coefficient does not equal naught as a function of the size of the largest community (over 91 days) that discusses each product category under study, and it does so for the undirected and directed networks. The sizes of these largest communities are given in table 1.

Table 2 and fig. 3 clearly illustrate that the directed networks and the undirected networks behave differently, and the regression results in table 3 show that the difference is significant. The clustering coefficient for the very large Music conversation does not equal naught on any of the 91 days under study, whether the network is directed or not. However, for the conversations about the five out of the six product categories whose largest community is smaller than 30,000 over 91 days (Sports, Comedy, Science, Entertainment and Howto), CC_ud behaves very differently from CC_d. Fig. 3 shows that CC_d deviates from naught on relatively rare occasions, whereas CC_ud increases linearly as a function of the largest community and begins to saturate at about 25,000.

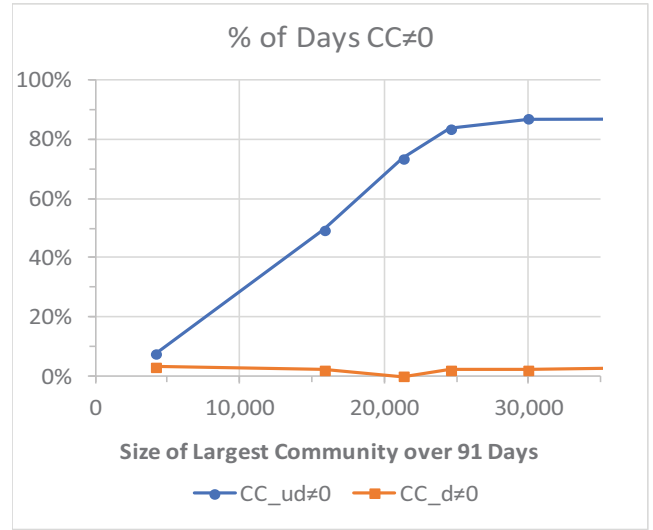


Fig. 3. Percentage of days (out of 91) for which the clustering coefficient does not equal naught versus size of the largest community (over 91 days) that discusses each product category.

TABLE 3: RESULTS OF REGRESSION OF DATA IN FIG. 3

	Undirected	Directed
Best-fit values		
Slope	0.002980 ± 0.0003439	-4.970e-005 ± 5.764e-005
Y-intercept	-2.337 ± 7.253	2.753 ± 1.216
X-intercept	784.1	55387
1/Slope	335.5	-20120
95% Confidence Intervals		
Slope	0.001886 to 0.004075	-0.0002331 to 0.0001337
Y-intercept	-25.42 to 20.74	-1.115 to 6.621
X-intercept	-10622 to 6459	21675 to +infinity
R square	0.9616	0.1986
Sy.x	6.757	1.132
Is slope significantly non-zero?		
F	75.09	0.7434
DFn,DFd	1,3	1,3
P Value	0.0032	0.452
Deviation from horizontal?	Significant	Not Significant
Data		
Number of XY pairs	5	5
Equation	Y = 0.002980*X - 2.337	Y = -4.970e-005*X + 2.753

Figs. 4a through 4f in appendix A show that CC_ud does not equal CCudran on all days for which CC_ud does not equal naught. Fig. 5a in appendix B shows that CC_d does not equal CCdran on all days for which CC_d does not equal naught. This suggests the following:

- The directed and the undirected networks for Twitter conversations about the product category Music are not random networks. Their clustering coefficients are a true reflection of networks' features.

- The undirected networks for Twitter conversations about the Sports, Comedy, Science and Entertainment product categories tend to be non-random on a significant amount of the days that they were observed. On these days, their clustering coefficients are a true reflection of networks' features. By contrast, the undirected networks for Twitter conversations about Howto tend to be random on most days.
- The directed networks for Twitter conversations about the other product category Music are random networks on almost all days of the 91 days under study. Their clustering coefficients are not a true reflection of network's features.
- Finally, our day-to-day observation of the social networks under study imply evolutionary path dependence as described by Utterback and Suárez [107]. No network in our study changes structure like any other. And, very importantly, online social networks can change structure from minute to minute, perhaps at a pace that exceed our ability to analyze them [121].

V. DISCUSSION

We believe the research described in this paper is unique in a variety of ways. First and foremost, our study is the first that measures the impact of scale and directionality on online network phenomena. It is also the first study that attempts to measure the degree of randomness in online social networks by comparing the reciprocal and directional networks in our sample to randomly generated Erdős-Rényi [53] networks. This allowed us to discover that the degree of randomness within a network is a function of directionality and scale.

We can infer from the results of our study that larger social networks undergo different generation mechanisms than smaller networks do, especially if these networks are directed. Our paper consequently makes a contribution to theory by demonstrating that extant theoretical frameworks, which were developed from observing relatively smaller social networks in the real world (e.g., [6] [13] [27] [31] [33] [34] [39] [43] [60] [74] [81] [90] [98] [105]), cannot explain network structure or phenomena associated with online social networks in their entirety, especially if directionality is not taken into account in the analysis of said networks. We also consider these discoveries a contribution to management practice. Henceforth, all managers and analysts of online social networks need to be aware of the need to analyze online communities both as reciprocal and directed networks. Managers and analysts that fail to do so, are likely to misallocate a large amount of resources (Edwards [52]), simply because they do not understand how the online social networks with which they interact are organized (Aral, et al. [10], Cha, et al. [40] [41], Li and Bernoff [71], Mayande [76], Weber and Mayande [121]).

We have observed in this study that, due to the ability to conduct searches in online social networks (Adamic and Adar [2], Watts et al. [110]), the structure of online social networks is emergent (in the sense of Drazin and Sandelands [50] [99]).

As a consequence, online social networks exhibit a highly dynamic network structure (Dodds et al. [47], Wiertz et al. [131]), in which real-world constraints such as connectedness and distance may not affect the behavior of online networks in a significant way (Borgatti et al. [22]-[26]). Instead, factors such as "scalefreeness" (Barabási, et al. [4] [5] [14] [15]), "assortativity" (Newman et al. [87]) and "smallworldness" (Watts and Strogatz [111]), which can alter the topological organization of network structure, could be paramount. Thus, we cannot attribute all network flows in online social networks to social relationships (Greenemeier [61], Pei et al. [91]). For example, we know from the study of hashtag communities (Weng et al. [130]) that a two people can discuss a topic of mutual interest online without having a permanent connection. By contrast, the structure of real-world social networks, upon which extant theory is based, tends to be static (Burt [32], Moffitt [83]). Furthermore, existing measures of the network phenomenon of influence, such as centrality metrics derived from graph theory (e.g., Freeman [54] [55]), may be inadequate; they may not really measure influence (Mayande [76]). By contrast, eigenvector centrality (Bonacich [20] [21]) measures influence as a function of the spread of information but not as a function of how fast the information spreads through the network [76]. All this reinforces the notion that extant theory does not characterize the behavior of online social networks to a sufficient degree of satisfaction.

While the study described in this paper has pointed out the deficiencies of extant theory, it has provided little insight into the generation mechanisms that are specific to online social networks. This constitutes a limitation the study, and further research of this topic is consequently warranted. Specifically, the relationships between network flows, network structure and network phenomena like influence need to be explored and characterized. The authors of this paper and their colleagues are particularly interested in the following research questions [76]-[79]:

1. How do network flows impact network structure?
2. How does network structure impact network flows?
3. How do network flows and network structure impact network phenomena such as influence?

To answer these questions, we propose building on research that has been described in this paper and in prior studies on the behavior of online social networks (Mayande, et al. [76]-[79]). The proposed research will primarily consist of exploratory population studies of online communities. We intend to include all the basic network metrics that were utilized for the study in this paper: the number of nodes and ties; reciprocity; network density; and the clustering coefficient. We will add structural metrics such as scalefreeness, smallworldness and assortativity, as well as variables provided by graph theory (Borgatti [22]) and information theory (e.g., [1] [16] [44] [62] [69] [100] [123] [124]), which act as proxies for measuring the degree of or impedance to network flow [121]. Information theory also provides a set of metrics for influence, which measure how much information can

concentrate at a specific node (*Mayande and Weber [77], Nikolaev, et al. [88], Tutzauer [106]*). Eigenvector centrality (*Bonacich [20] [21]*) complements entropy centrality as a measure of influence. It is a function of the extent that information has spread, but it does not measure the speed at which information spreads [76].

The results of this study also suggest that scale and directionality should be added as control variables. Including network size as a control variable will facilitate the observation of effects that result from scale. The impact of directionality can be assessed by analyzing directed and undirected networks as has been done in the study in this paper. However, it has been suggested that a person within a social network who influences how information is consumed may not necessarily influence how it is propagated, and conversely [121]. Identifying key influencers and deducing the behavioral characteristics of individuals within a network may consequently mandate an analysis of four networks: the undirected network, the directed network, the network that is focused on consumption and another that is focused on propagation.

When addressing the abovementioned research questions, it is of paramount importance to recognize that network flows are knowledge flows rather than information flows [121]. Thus, research that attempts to answer the questions from above is highly likely to be context-specific. Fortunately, research pertaining to knowledge flows in the real world is plentiful and has been conducted in a variety of contexts ([3] [8] [17] [19] [29] [37] [38] [42] [56] [86] [89] [92]-[96] [103]-[105] [112]-[120] [122]-[129] [133]-[136]). The question is, do its conclusions carry over into the online domain?

To address issues pertaining to knowledge we propose engaging in what Weber and Mayande [121] call *virtual field work*, an approach that has been utilized successfully by a variety of researchers (*e.g., Jansen, et al. [66], Williams, et al. [132]*). We collect data about a social network from the Internet, which we analyze using statistical methods. We repeat this analysis for the various contexts in which the network operates. These contexts will be explored in keyword searches, in which each set of keywords acts as a proxy for different kinds of knowledge.

Once again, most of these limitations can be addressed by further study. For example, issues pertaining to limited number of Twitter conversations could be addressed by repeating this study in a setting in which many more Twitter conversations are available. Broader generalizability could be achieved, if similar studies would be conducted across a variety of social media platforms. We consequently view our paper as an impetus for further research into online phenomena, which—despite a recent surge in attention—still constitute a topic that remains underexplored. Given the increasing importance of social media in fields such as advertising and marketing (*e.g., Bressler and Grantham [28], Honeycutt and Herring [64], Huberman et al. [65], Jansen et al. [66], Khammash and Griffith [68], Li and Bernoff [71], Lindsay et al [72] Longart [73]*), the proposed research and novel theory resulting

therefrom should make a significant contribution to management practice.

Finally, it should be noted that our study is subject to a limitation that could affect the conclusions from section V. Our study is based exclusively on conversations from one social media platform—Twitter—about six product categories from one entertainment platform—YouTube [79]. The conclusion about directed networks and scale could thus constitute an oversimplification. The propensity for directed networks to be non-random as a function of the size of the largest community could instead take the shape of an S-curve in the domain between 30,000 and 1.5 million—we have no way of knowing from this study. One could also conclude for analogous reasons that one result of our study—the one which states that scale increases linearly as a function of the largest community for undirected networks and subsequently saturates—is not necessarily statistically significant. Moreover, as Goggins and Petakovic [58] have shown, the behavior of online social networks can vary greatly from platform to platform, suggesting that the results of our Twitter-based study may not be valid for discussions that are carried out on, say, Facebook.

VI. CONCLUSION

The results of our study lead to the conclusion that the null hypothesis that was proposed at the beginning of this paper must be rejected—clearly, directionality and scale do impact network structure. The two networks that pertained to Music, whose largest conversations exceed a total of 1 million over 91 days, were not random on any of these days, regardless of whether they were directed or undirected. For the undirected networks of the Twitter conversations about the remaining product categories, non-randomness was a function of scale. Undirected networks with bigger largest communities tended to be random on fewer days, than undirected networks with smaller largest communities. By contrast, the directed networks associated with the Twitter conversations that involved smaller largest communities were random on almost all days of the 91-day period under observation. This suggests that directed networks tend to be random unless they are very large, whereas the propensity for undirected networks to be non-random tends to increase with network size and saturate.

REFERENCES

- [1] N. Abramson, *Information Theory and Coding*, New York, NY, USA: McGraw-Hill, 1963.
- [2] L. Adamic and E. Adar, "How to search a social network," *Social Networks*, Vol. 27, 2005, pp. 187-203.
- [3] F. M. Albar and A. J. Jetter, "Fast and frugal heuristics for new product screening - is managerial judgment 'good enough?'," *International Journal of Management and Decision Making*, Vol. 12, No. 2, 2013, pp. 165-189.
- [4] R. Albert and A. L. Barabási, "Topology of evolving networks: Local events and universality," *Physical Review Letters*, Vol. 85, 2000, pp. 5234-5237.
- [5] R. Albert, H. Jeong and A. L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, Vol. 406, 2000, pp. 378-382.

- [6] T. J. Allen, *Managing the Flow of Technology*, Boston, MA, USA: MIT Press, 1977.
- [7] <http://preview.alexacom/siteinfo/youtube.com> (accessed 04/09/2014).
- [8] D. G. Ancona and C. L. Chong, "Entrainment: Pace, cycle, and rhythm in organizational behavior", *Research in Organizational Behavior*, Vol. 18, 1996, pp. 251-284.
- [9] B. S. Anderson, C. Butts and K. Carley, "The interaction of size and density with graph-level indices," *Social Networks*, Vol. 21, 1999, pp. 239-267.
- [10] S. Aral, C. Dellarocas and D. Godes, "Introduction to the special issue-social media and business transformation: A framework for research," *Information Systems Research*, Vol. 24, 2013, pp. 3-13.
- [11] S. Aral and D. Walker, "Creating social contagion through viral product design: A randomized trial of peer influence in networks," *Management Science*, Vol. 57, 2011, pp. 1623-1639.
- [12] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," *Science*, Vol. 337, 2012, pp. 337-341.
- [13] K. D. Bailey, "Social entropy theory: An overview," *Systemic Practice and Action Research*, Vol. 3, 1990, pp. 365-382.
- [14] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, Vol. 286, 1999, pp. 509-512.
- [15] A. L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific American*, Vol. 288, 2003, pp. 60-69.
- [16] P. Beckmann, *Probability in Communication Engineering*, New York, NY, USA: Harcourt, Brace & World, Inc., 1967.
- [17] C. N. Berglund, C. M. Weber and P. Gabella, "Benchmarking the productivity of photomask manufacturers," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 22, No. 4, November 2009, pp. 499-506.
- [18] H.-P. Blossfeld and G. Rohwer, *Techniques of event history modeling: New approaches to causal analysis (2nd ed.)*, Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers, 1995.
- [19] R. E. Bohn, "Measuring and Managing Technological Knowledge," *Sloan Management Review*, Vol. 36, No. 1, Fall 1994, pp. 61-73.
- [20] P. Bonacich, "Factoring and weighing approaches to status scores and clique identification," *Journal of Mathematical Sociology*, Vol. 2, 1972, pp. 113-120.
- [21] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Networks*, Vol. 29, 2007, pp. 555-564.
- [22] S. P. Borgatti, "Centrality and network flow," *Social Networks*, Vol. 27, 2005, pp. 55-71.
- [23] S. P. Borgatti and R. Cross, "A relational view of information seeking and learning in social networks," *Management Science*, Vol. 49, 2003, pp. 432-445.
- [24] S. P. Borgatti and P. C. Foster, "The network paradigm in organizational research: A review and typology," *Journal of Management*, Vol. 29, 2003, pp. 991-1013.
- [25] S. P. Borgatti and D. S. Halgin, "On network theory," *Organization Science*, Vol. 22, 2011, pp. 1168-1181.
- [26] S. P. Borgatti and V. Kidwell, *Network Theorizing*, Thousand Oaks, CA, USA: Sage Publications, 2011.
- [27] P. Bourdieu, *The Forms of Capital*, New York, NY, USA: Greenwood, 1986.
- [28] S. E. Bressler and C. Grantham, *Communities of Commerce: Building Internet Business Communities to Accelerate Growth, Minimize Risk, and Increase Customer Loyalty*, New York, NY, USA, McGraw-Hill, Inc., 2000.
- [29] H. Bresman, "External learning activities and team performance: A multi-method field study," *Organization Science*, Vol. 21, No. 1, 2010, pp. 81-96.
- [30] J. M. Bristor, *Enhanced Explanations of Word of Mouth Communications: the Power of Relationships*, Research and Publications, School of Business Administration, University of Western Ontario, 1989.
- [31] R. S. Burt, "Positions in networks," *Social Forces*, Vol. 55, 1976, pp. 93-122.
- [32] R. S. Burt, "Social contagion and innovation - Cohesion versus structural equivalence," *American Journal of Sociology*, Vol. 92, 1987, pp. 1287-1335.
- [33] R. S. Burt, *Structural Holes: The Social Structure of Competition*, Cambridge, Mass., USA: Harvard University Press, 1992.
- [34] R. S. Burt, *Brokerage and Closure: An Introduction to Social Capital*, Oxford -New York: Oxford University Press, 2005.
- [35] R. S. Burt and P. Doreian, "Testing a structural model of perception - Conformity and deviance with respect to journal norms in elite sociological methodology," *Quality and Quantity*, Vol. 16, 1982, pp. 109-150.
- [36] R. S. Burt, M. Kilduff and S. Tasselli, "Social network analysis: Foundations and frontiers on advantage," *Annual Review of Psychology*, Vol. 64, 2013, pp. 527-547.
- [37] P. R. Carlile, "A pragmatic view of knowledge and boundaries: Boundary objects in new product development," *Organization Science*, Vol. 13, No. 4, July/August 2002, pp. 442-455.
- [38] P. R. Carlile, "Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries," *Organization Science*, Vol. 15, No. 5, September-October 2004, pp. 555-568.
- [39] D. Cartwright, *Influence, Leadership, Control*, Chicago, IL, USA: Rand McNally, 1965.
- [40] M. Cha, H. Haddadi, F. Benevenuto and P. K. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Vol. 10, 2010, pp. 10-17.
- [41] M. Cha, A. Mislove and K. P. Gummadi, "A measurement-driven analysis of information propagation in the flickr social network," *Proceedings of the 18th international conference on World Wide Web, 2009. ACM*, pp. 721-730.
- [42] W. M. Cohen and D. A. Levinthal, "Absorptive capacity: A new perspective on learning and innovation," *Administrative Science Quarterly*, Vol. 35, No. 1, March 1990, pp. 128-152.
- [43] J. Coleman, "Social capital in the creation of human capital," *American Journal of Sociology*, Vol. 94, 1988, S95-S120.
- [44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, NY, USA: Wiley, 1991.
- [45] J. Deighton, "What are marketers' top concerns?" (MSI Research Priorities 2012-14), Jun 25, 2012 ed.: Marketing Science Institute.
- [46] C. Dellarocas, Z. Katona and W. Rand, "Media, aggregators and the link economy: Strategic hyperlink formation in content networks," *Management Science*, vol. 59, 2013, pp. 2360-2379.
- [47] P. S. Dodds, R. Muhamad and D. J. Watts, "An experimental study of search in global social Networks," *Science*, Vol. 301, 2003, pp. 827-829.
- [48] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss and C. M. Danforth, "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter," *PloS one*, Vol. 6, Dec. 7, 2011, e26752.
- [49] C. Donnering, "The distribution of centrality in social networks," *Social Networks*, Vol. 8, 1986, pp. 191-203.
- [50] R. Drazin and L. Sandelands, "Autogenesis - A perspective on the process of organizing," *Organization Science*, Vol. 3, 1992, pp. 230-249.

- [51] D. F. Duhan, S. D. Johnson, J. B. Wilcox and G. D. Harrell, "Influences on consumer use of word-of-mouth recommendation sources," *Journal of the Academy of Marketing Science*, Vol. 25, 1997, pp. 283-295.
- [52] J. Edwards, "Priceline CEO, who has a \$1.8 billion online ad budget, says Facebook and Twitter are useless," *Business Insider*, April
- [53] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Mathematicae (Debrecen)*, Vol. 6, 1959, pp. 290-297.
- [54] L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, Vol. 40, 1977, pp. 35-41.
- [55] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks*, Vol. 1, 1979, pp. 215-239.
- [56] P. Gabella and C. M. Weber, "Synchronizing innovation: The case of semiconductor lithography," *Proceedings of PICMET '10*, Phuket, Thailand, July 18-22, 2010, pp. 2646-2670.
- [57] <http://gigaom.com/2012/06/04/twitter-shows-when-we-tweet-and-explains-why-its-search-sucks/> (accessed on 4/27/2014).
- [58] S. P. Goggins and E. Petakovic, "Connecting theory to social technology platforms: A framework for measuring influence in context," *American Behavioral Scientist*, Vol. 58, No. 10, 2014, pp. 1376-1392.
- [59] B. Gonçalves and J. J. Ramasco, "Human dynamics revealed through Web analytics," *Physical Review E*, vol. 78, 2008, 026123.
- [60] M. S. Granovetter, "The strength of weak ties," *American Journal of Sociology*, Vol. 78, 1973, pp. 1360-1380.
- [61] L. Greenemeier, "Inflated expectations: Crowd-sourcing comes of age in the DARPA network challenge," *Scientific American*, Dec. 21, 2009.
- [62] R. V. L. Hartley, "Transmission of information," *Bell System Technology Journal*, Vol. 7, No. 3, July 1928, pp. 535-563.
- [63] N. O. Hodas and K. Lerman, "The simple rules of social contagion," *Science Reports*, Vol. 4, Article 4343, 2014.
- [64] C. Honeycutt and S. Herring, Beyond Microblogging: Conversation and Collaboration via Twitter. *Hawaii International Conference on System Sciences*, 0, 2009, pp. 1-10.
- [65] B. A. Huberman, D. M. Romero, and F. Wu, "Social Networks that Matter: Twitter Under the Microscope," *CoRR*, 2008, abs/0812.1045.
- [66] B. Jansen, M. Zhang, K. Sobel and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American Society for Information Science and Technology*, Vol. 60, 2009, pp. 2169-2188.
- [67] N. Kejzar, Z. Nikoloski and V. Batagelj, "Probabilistic inductive classes of graphs," *Journal of Mathematical Sociology*, Vol. 32, 2008, pp. 85-109.
- [68] M. Khammash and G. H. Griffiths, "'Arrivederci CIAO.com, Buongiorno Bing.com'—Electronic word-of-mouth (eWOM), antecedences and consequences," *International Journal of Information Management*, Vol. 31, 2011, pp. 82-87.
- [69] S. Kullback, *Information Theory and Statistics*, New York: Dover, 1968.
- [70] D. Leonard-Barton, "A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites," *Organization Science*, Vol. 1, 1990, pp. 248-266.
- [71] C. Li and J. Bernoff, *Groundswell: Winning in a World Transformed by Social Technologies*, Boston, MA, USA: Harvard Business Press, 2008.
- [72] A. Lindsay, J. Kaykas-Wolff and C. Mathwick, "Key influencers: Locating, measuring and creating programs to influence social media influencers," *Proceedings of the Academy of Marketing Science (AMS) Annual Conference*, 2010, p. 1.
- [73] P. Longart, What drives word-of-mouth in restaurants? *International Journal of Contemporary Hospitality Management*, Vol. 22, 2010, pp. 121-128.
- [74] N. Luhmann, "The autopoiesis of social systems," In: F. Geyer, a. J. V. d. Z. (ed.) *Sociocybernetic Paradoxes: Observation, Control and Evolution of Self-Steering System*. London, UK: Sage, 1986.
- [75] J. G. March, "An introduction to the theory and measurement of influence," *American Political Science Review*, Vol. 49, 1955, pp. 431-451.
- [76] N. Mayande, *Network Structure, Network Flows and the Phenomenon of Influence in Online Social Networks: An Exploratory Empirical Study of Twitter Conversations about YouTube Product Categories*, Dissertation in Engineering and Technology Management, Portland State University, July 2015. http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=3471&context=open_access_etds
- [77] N. Mayande and C. M. Weber, "Designing virtual social networks for for-profit open innovation," *Proceedings of PICMET '11*, Portland, Oregon, July 31-Aug. 4, 2011.
- [78] N. Mayande and C. M. Weber, "Directed interaction networks and their impact on social media," *Proceedings of PICMET '13*, San Jose, CA, USA, July 28 to August 1, 2013, pp. 1709-1722.
- [79] N. Mayande and C. M. Weber, "Product popularity versus size of conversation in social media: An analysis of Twitter conversations about YouTube product categories," *Proceedings of PICMET '18*, Honolulu, HI, USA, Aug. 18-22, 2018 (18R0132) 978-1-890843-37-3.
- [80] M. R. Meiss, F. Menczer, S. Fortunato, A. Flamini and A. Vespigiani, "Ranking web sites with real user traffic," *Proceedings of the ACM 2008 International Conference on Web Search and Data Mining*, Palo Alto, California, USA, Sept. 7-11, 2008, pp. 65-76
- [81] J. G. Miller, *Living Systems*, New York, NY, USA: McGraw-Hill, 1978.
- [82] A. Mislove, M. Marconi, K. P. Gummadi, P. Druschel, B. Bhattacharjee, "Measurement and analysis of online social networks," *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007, pp. 29-42.
- [83] R. A. Moffitt, "Policy interventions, low-level equilibria, and social interactions," *Social Dynamics*, Vol. 4, 2001. pp. 45-82.
- [84] R. B. Money, M. C. Gilly and J. L. Graham, "Explorations of national culture and word-of-mouth referral behavior in the purchase of industrial services in the United States and Japan," *Journal of Marketing*, Vol. 62, 1998, p. 76.
- [85] D. Montgomery, *Design and Analysis of Experiments*, Hoboken, NJ, USA: Wiley, 2008.
- [86] L. A. Nemanich, R. T. Keller, D. Vera and W. W. Chin, "Absorptive capacity in R&D project teams: A conceptualization and empirical test," *IEEE Transactions on Engineering Management*, Vol. 57, No. 4, 2010, pp. 674-688.
- [87] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, Vol. 89, 2002, pp. 208701.
- [88] A. G. Nikolaev, R. Razib and A. Kucheriya, "On efficient use of entropy centrality for social network analysis and community detection," *Social Networks*, Vol. 40, Jan. 2015, pp. 154-162.
- [89] W. J. Orlikowski, "Using technology and constituting structures: A practice lens for studying technology in organizations," *Organization Science*, Vol. 11, No. 4, 2000, pp. 404-428.
- [90] T. Parson, *The Social System*, New York, NY, USA: The Free Press, 1951.
- [91] S. Pei, L. Muchnik, J. J. S. Andrade, Z. Zheng and H. A. Makse, "Searching for superspreaders of information in real-world social media," *Scientific Reports*, Vol. 4, No. 5547, 2014.
- [92] P. Ploykitikoon, *The Impact of Knowledge Inflows on the Performance of National Laboratories in Technological Latecomer Countries*, Dissertation in Engineering and Technology Management, Portland State University, June 2013. http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=2070&context=open_access_etds
- [93] P. Ploykitikoon and C. M. Weber, "Selective absorption of knowledge and open innovation in the national laboratories," *11th Open and User Innovation Conference*, Brighton, UK, July 15, 2013.

- [94] P. Ploykitikoon and C. M. Weber, "Organizational knowledge and ambidexterity in the national laboratories," *Proceedings of PICMET '13* (13A0138), San Jose, CA, USA, July 28 to August 1, 2013, pp. 1370-1388.
- [95] P. Ploykitikoon and C. M. Weber, "Internal or external knowledge: Which is more important for the performance of national laboratories in technology latecomer countries?" *Proceedings of PICMET '15*, Portland, OR, USA, Aug. 2-5, 2015, pp. 1787 - 1797.
- [96] P. Ploykitikoon and C. M. Weber, "Knowledge pathways and performance: An empirical study of the national laboratories in a technology latecomer country," *International Journal of Innovation and Technology Management*, in press (online ready), 2019.
- [97] R. Putnam, "Bowling alone: America's declining social capital," *Journal of Democracy*, Vol. 6, No. 1, 1995, pp. 65-78.
- [98] E. M. Rogers, *Diffusion of Innovations*, New York, NY, USA: Free Press, 2003.
- [99] L. Sandelands and R. Drazin, "On the language of organization theory," *Organization Studies*, Vol. 10, 1989, pp. 457-478.
- [100] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL, USA: University of Illinois Press, 1949.
- [101] H. A. Simon, *Models of Man*, New York, NY, USA: Wiley, 1957.
- [102] <http://www.statisticbrain.com/twitter-statistics/> (accessed 04/09/2014)
- [103] G. Szulanski, "Exploring internal stickiness: Impediments to the transfer of best practice within the firm," *Strategic Management Journal*, Vol. 17, 1996, pp. 27-43.
- [104] G. Szulanski, D. Ringov and R. J. Jensen, "Overcoming stickiness: How the timing of knowledge transfer methods affects transfer difficulty," *Organization Science*, Vol. 27, No. 2, 2016, pp. 304-322.
- [105] N. M. Tichy, M. L. Tushman and C. Fombrun, "Social network analysis for organizations," *Academy of Management Review*, October 1979, Vol. 4, No. 4, pp. 507-519.
- [106] F. Tutzauer, "Entropy as a measure of centrality in networks characterized by path-transfer flow," *Social Networks*, Vol. 29, No. 2, May 2007, pp. 249-265.
- [107] J. M. Utterback and F. F. Suárez, "Innovation, competition, and industry structure," *Research Policy*, Vol. 22, No. 1, Feb. 1993, pp. 1-21.
- [105] E. von Hippel, "'Sticky information' and the locus of problem solving: Implications for innovation," *Management Science*, Vol. 40, No. 4, 1994, pp. 429-439.
- [109] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge, UK: Cambridge University Press, 1994.
- [110] D. J. Watts, P. S. Dodds and M. E. J. Newman, "Identity and search in social networks," *Science*, Vol. 296, 2002, pp. 1302-1305.
- [111] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, Vol. 393, 1998, pp. 440-442.
- [112] C. Weber, "Knowledge transfer and the limits to profitability: An empirical study of problem-solving practices in the semiconductor industry," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 15, No. 4, Nov. 2002, pp. 420-426.
- [113] C. Weber, "Yield learning and the sources of profitability in semiconductor manufacturing and process development," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 17, No. 4, November 2004, pp. 590-596.
- [114] C. M. Weber, "Do learning organizations have strokes of genius?" *Proceedings of PICMET '06*, Istanbul, Turkey, July 8-13, 2006, pp. 1220-1235.
- [115] C. M. Weber, "Characterizing the economic value of organizational learning in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 26, No. 1, Feb. 2013, pp. 42-52.
- [116] C. M. Weber, C. N. Berglund and P. Gabella, "Mask cost and profitability in photomask manufacturing: An empirical analysis," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 19, No. 4, November 2006, pp. 465-474.
- [117] C. M. Weber and A. Fayed, "Scale, scope and speed: Managing the challenges of semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 23, No. 1, February 2010, pp. 30-38.
- [118] C. M. Weber and A. Fayed, "Optimizing your position on the operating curve: How can a fab truly maximize its performance?" *IEEE Transactions on Semiconductor Manufacturing*, Vol. 23, No. 1, February 2010, pp. 21-29.
- [119] C. M. Weber, R. Hasenauer and N. V. Mayande, "Quantifying nescience: A decision-aid for practicing managers," *Proceedings of PICMET '17*, Portland, OR, USA, July 14-16, 2017.
- [120] C. M. Weber, R. Hasenauer and N. V. Mayande, "Toward a pragmatic theory for managing nescience," *International Journal of Innovation and Technology Management*, Vol. 15, No 5, 2018, p. 1850045.
- [121] C. M. Weber and N. V. Mayande, "Knowledge flows and influence in online social networks: Proposing a research agenda," *Proceedings of PICMET '17*, Portland, Oregon, July 9-13, 2017, pp. TBD.
- [122] C. Weber, B. Moslehi and M. Dutta, "An integrated framework for yield management and defect/fault reduction," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 8, no. 2, May 1995, pp. 110-120.
- [123] C. Weber, V. Sankaran, K. Tobin, and G. Scher, "A yield management strategy for semiconductor manufacturing based on information theory," *Proceedings of PICMET '99*, Portland, Oregon, USA, July 25-29, 1999, pp. 533-539.
- [124] C. Weber, V. Sankaran, K. Tobin and G. Scher, "Quantifying the value of ownership of yield analysis technologies," *IEEE Transactions of Semiconductor Manufacturing*, Vol. 15, No. 4, Nov. 2002, pp. 411-419.
- [125] C. M. Weber and J. Yang, "Managing Moore's Law: A survival guide for VLSI circuit manufacturers," *Proceedings of PICMET '12*, Vancouver, BC, Canada, Aug. 1, 2012, pp. 2715-2753.
- [126] C. M. Weber and J. Yang, "Organizational learning and capital productivity in semiconductor manufacturing," *IEEE/SEMI-ASMC*, Saratoga Springs, NY, USA, May 14-16, 2013, pp. 80-86.
- [127] C. M. Weber and J. Yang, "Managing knowledge impedance: A case from the office products industry," *Proceedings of PICMET '13* (13A0040), San Jose, CA, USA, July 28 to August 1, 2013, pp. 1309-1334.
- [128] C. M. Weber and J. Yang, "Organizational learning and capital productivity in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 27, No. 3, August 2014, pp. 316-326.
- [129] C. M. Weber and J. Yang, "Managing pattern-specific fixed costs in integrated device manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 24, No. 4, 2016, pp. 275-282.
- [130] L. Weng, A. Flammini, A. Vespignani and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Report*, Vol. 2, No. 335, 2012.
- [131] C. Wiertz, C. Mathwick, K. De Ruyter and B. Dellaert, "A balancing act: Exploring governance in a virtual P3 community," *Advances in Consumer Research*, Vol. 37, 2010, pp. 672 - 673.
- [132] S. A. Williams, M. M. Terras and C. Warwick, "What do people study when they study Twitter? Classifying Twitter-related academic papers," *Journal of Documentation*, Vol. 69, 2013, pp. 384-410.
- [133] J. Yang, C. M. Weber and P. Gabella, "Enabling collaborative solutions in the semiconductor manufacturing ecosystem," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 26, No. 4, Nov. 2013, pp. 465-475.
- [134] B. A. Zenobia and C. M. Weber, "Opening the black box of adoption: The Motive-Technology-Belief framework," *International Journal of Innovation and Technology Management*, Vol. 8, No. 4, Dec. 2011, pp. 535-555.

[135] B. A. Zenobia and C. M. Weber, "Bridging the gap between artificial market simulations and qualitative research in diffusion of innovation," *International Journal of Innovation and Technology Management*, Vol. 9 No. 4, Dec. 2012, 1250026, pp. 1-22.

[136] B. A. Zenobia and C. M. Weber, "Deciding to change: An event sequence analysis of technology adoption behavior in consumers," *International Journal of Innovation and Technology Management*, Vol. 10, No. 6, Dec. 2013, 1340029, pp. 1-21.

APPENDIX: VISUAL DISPLAYS OF COMPARISONS OF CLUSTERING COEFFICIENTS

For the six Twitter conversations about the product categories under study (Music, Sports, Comedy, Science, Entertainment and Howto), we compare the clustering coefficients of the largest communities to those of an equivalent E-R network. We do this for undirected and directed networks.

A. For Undirected Networks (CC_{ud} and CC_{udran})

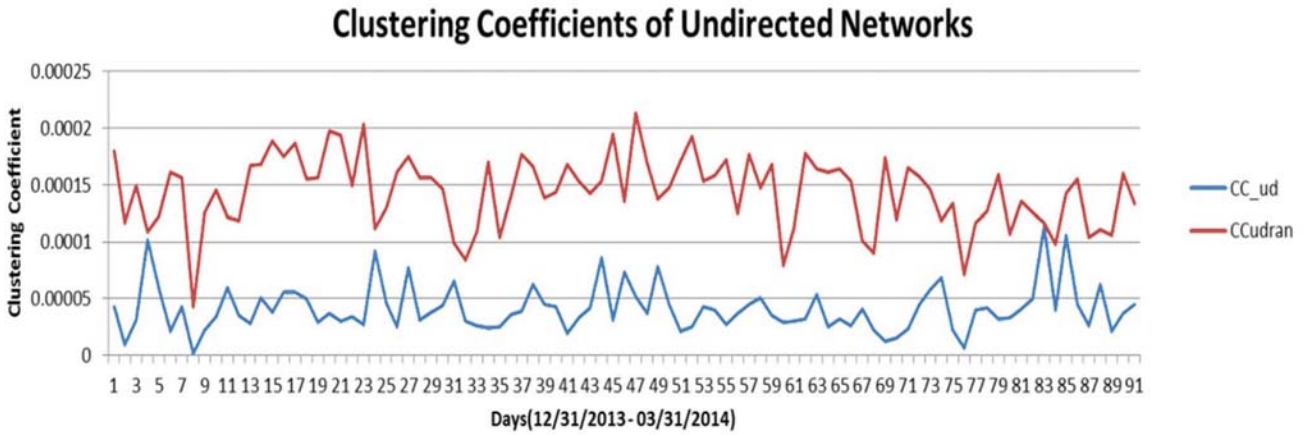


Fig. 4a. For the Twitter conversation about the Music product category.

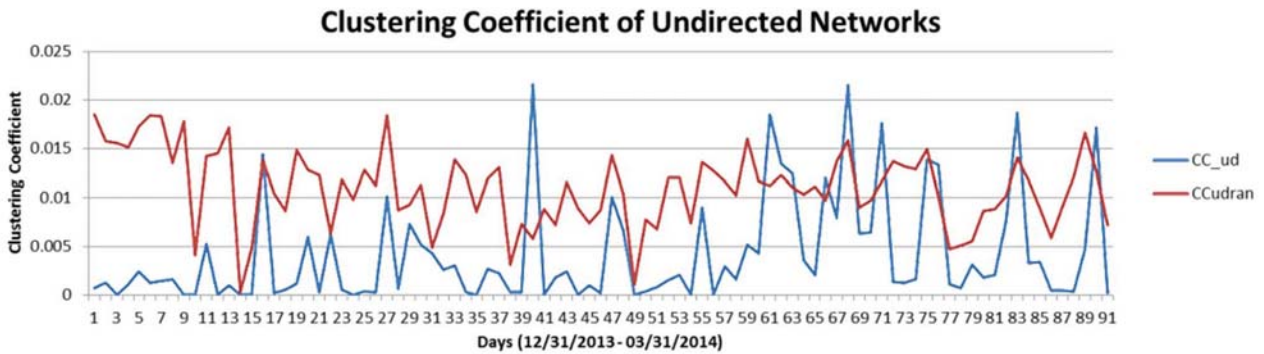


Fig. 4b. For the Twitter conversation about the Sports product category.

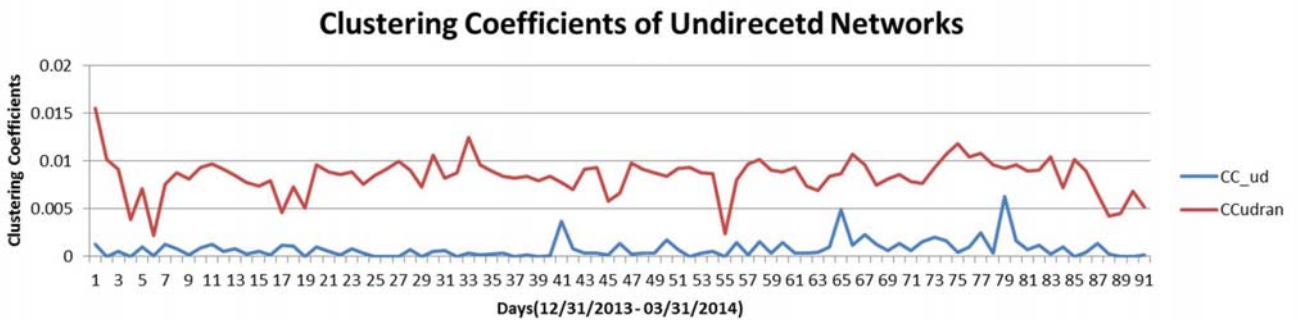


Fig. 4c. For the Twitter conversation about the Comedy product category.

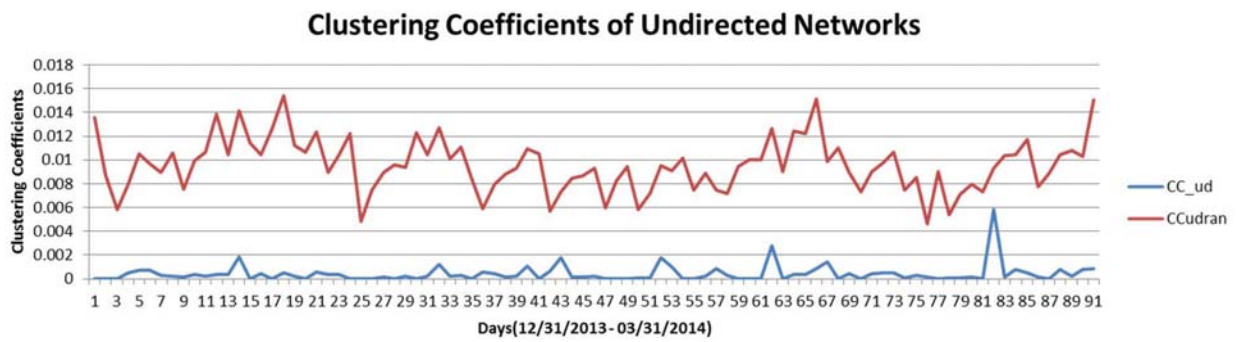


Fig. 4d. For the Twitter conversation about the Science product category.

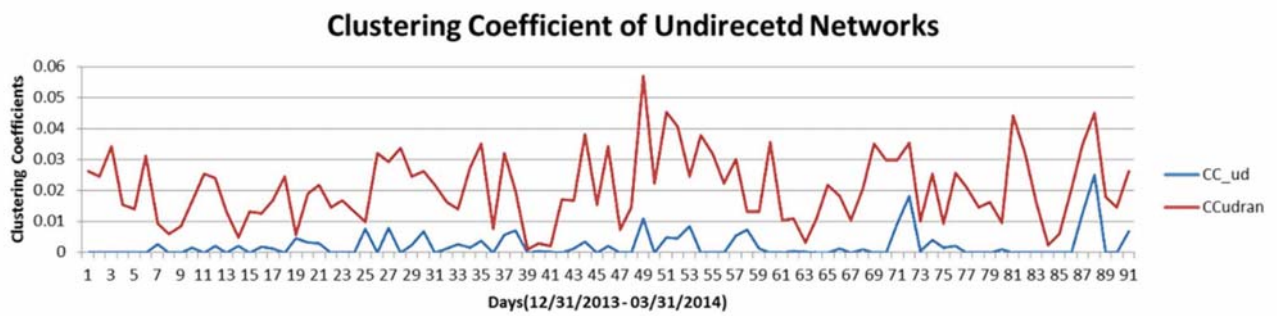


Fig. 4e. For the Twitter conversation about the Entertainment product category.

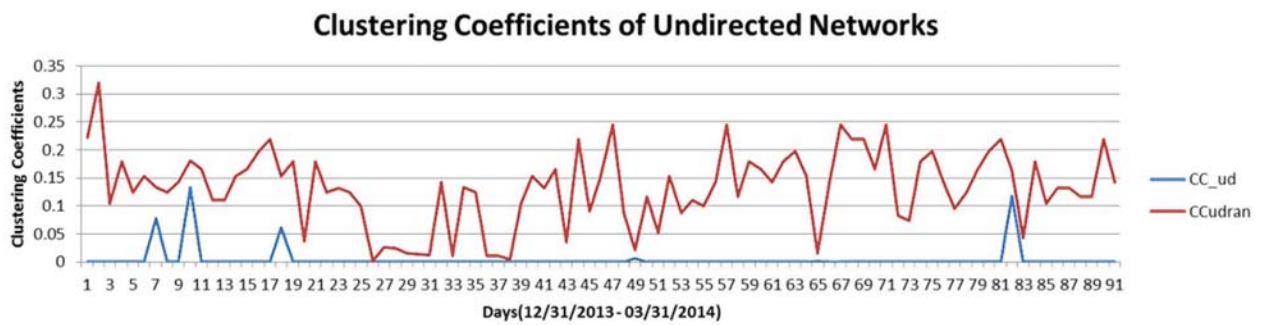


Fig. 4f. For the Twitter conversation about the Howto product category.

B. Directed Networks (CC_d and CC_{dran})

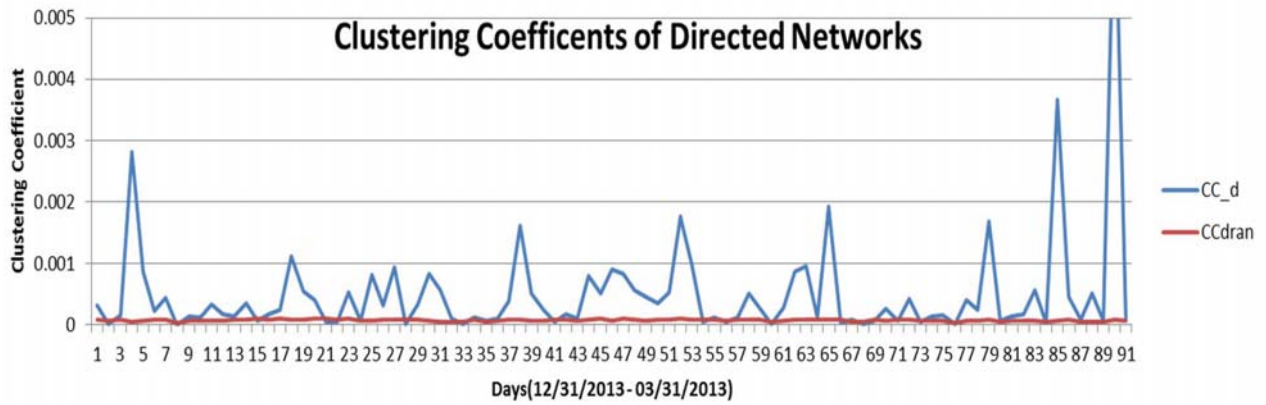


Fig. 5a. For the Twitter conversation about the Music product category.

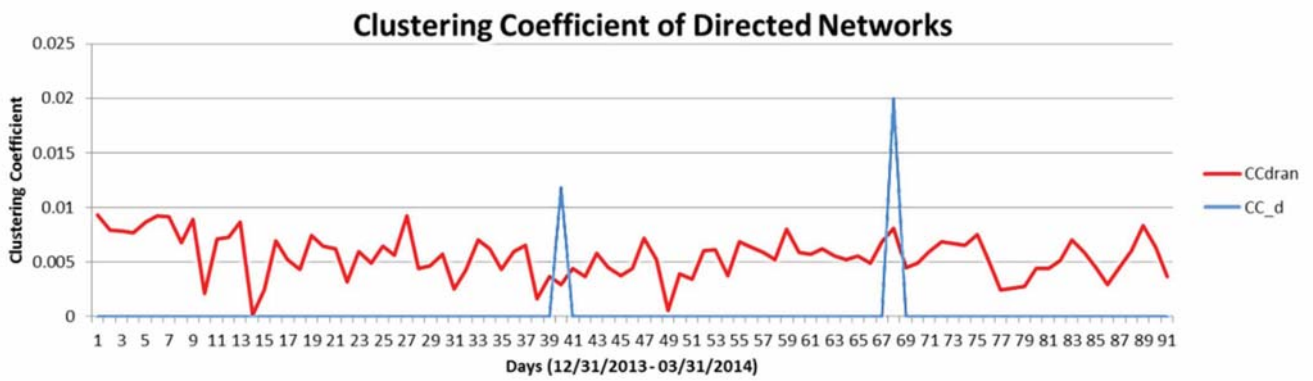


Fig. 5b. For the Twitter conversation about the Sports product category.

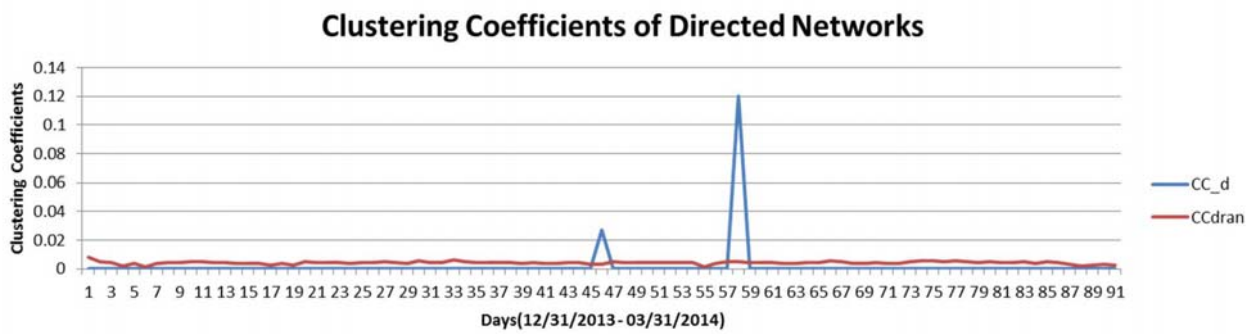


Fig. 5c. For the Twitter conversation about the Comedy product category.

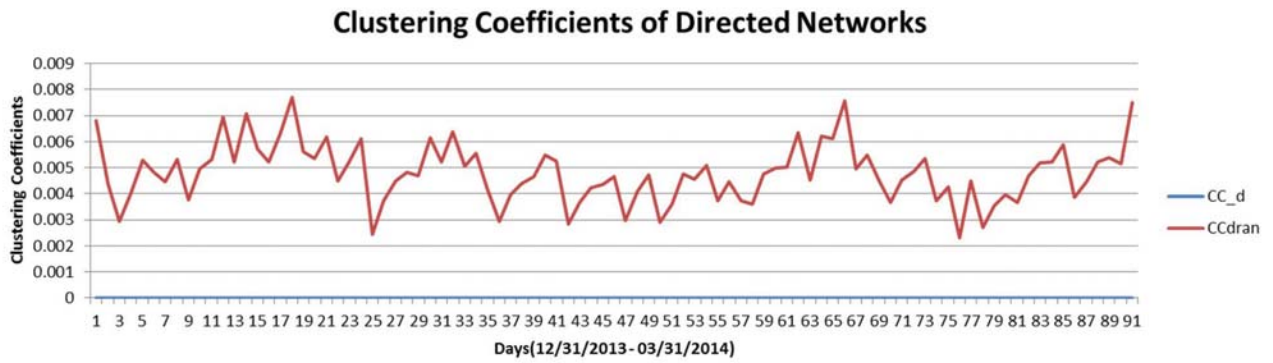


Fig. 5d. For the Twitter conversation about the Science product category.

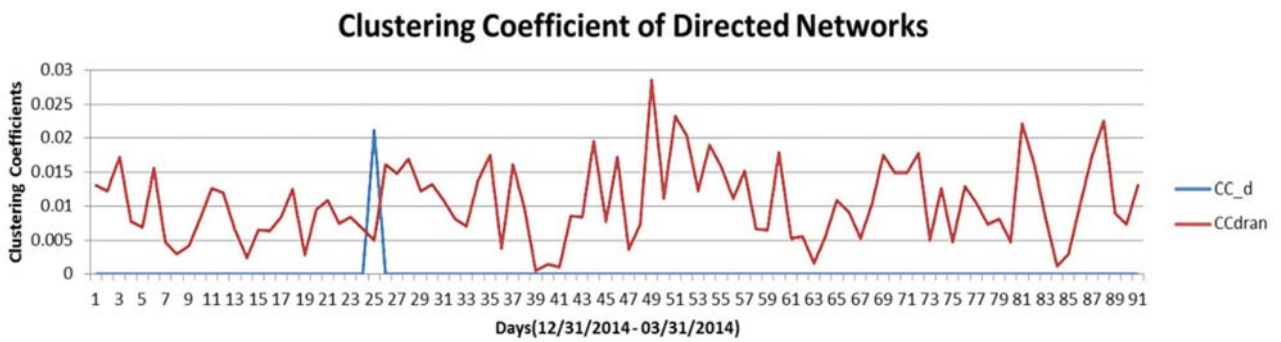


Fig. 5e. For the Twitter conversation about the Entertainment product category.

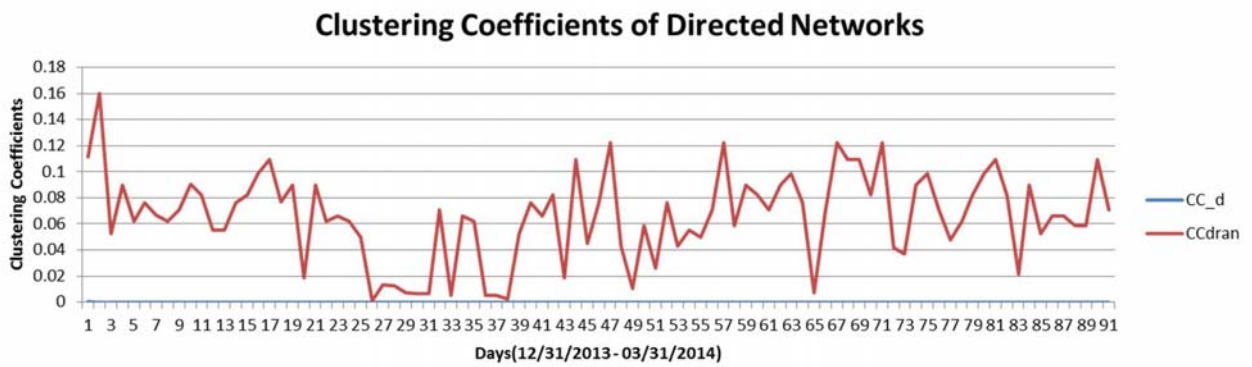


Fig. 5f. For the Twitter conversation about the Howto product category.