PewResearchCenter

NUMBERS, FACTS AND TRENDS SHAPING THE WORLD

In association with the Social Media Research Foundation

How we analyzed Twitter social media networks with NodeXL

Social media include all the ways people connect to people through computation. Mobile devices, social networks, email, texting, micro-blogging and location sharing are just a few of the many ways people engage in computermediated collective action. As people link, like, follow, friend, reply, retweet, comment, tag, rate, review, edit, update, and text one another (among other channels) they form collections of connections. These collections contain network structures that can be extracted, analyzed and visualized. The result can be insights into the structure, size, and key positions in these networks.

Social media networks form in Twitter around a wide range of terms. People talk about the news of the day, celebrities, companies, technology, entertainment, and more. As each person uses Twitter they form networks as they follow, reply and mention one another. These connections are visible in the text of each tweet or by requesting lists of the users that follow the author of each tweet from Twitter.

We collected, analyzed and visualized social media network data from Twitter using NodeXL, the free and open add-in for Excel 2007/2010/2013. <u>NodeXL</u> is a project from the <u>Social Media Research Foundation</u>¹, a not-for-profit organization dedicated to creating *open tools, open data*, and *open scholarship* related to social media.

NodeXL is a general purpose network analysis application that supports network overview, discovery and exploration². The tool enables the automation of a data flow that starts with the collection of network data and moves through multiple steps until final processed network visualizations and reports are generated (figure 1). NodeXL allows non-programmers to quickly generate useful network statistics; metrics and visualizations in the context of the familiar Excel spreadsheet (figure 2). Simple filtering and flexible display attributes can be used to highlight important structures in networks easily.

NodeXL supports the exploration of social media with import features that extract network data from a range of data sources like personal email indexes on the desktop, Twitter, Flickr, YouTube, Facebook, Wikis and WWW hyper-links (see figure 3). Other sources of data can be imported through text, CSV, or GraphML files.

¹ The Social Media Research Foundation: <u>http://www.smrfoundation.org</u>

² Technical questions about NodeXL can be asked on the project discussion boards at <u>http://nodexl.codeplex.com/Thread/List.aspx</u>. A book <u>Analyzing Social Media Networks with NodeXL: *Insights* <u>from a connected world</u> is available from <u>Morgan-Kaufmann</u>.</u>

NodeXL enables the automatic execution of a five step data work flow that starts with data collection from a variety of network data sources, through storage, analysis, visualization and finally publication.



Figure 1: Network analysis data flow

Each step is configured by the user and can then be executed in a batch. Users interact with the network via the familiar spreadsheet interface:

PEW RESEARCH CENTER



Figure 2: NodeXL

Users initiate the data analysis process by first collecting a data set using the NodeXL data importer. Multiple data sources and file formats contain network data structures. Notably, social media services generate many network data structures. NodeXL enables simple access to social media and other forms of network data through a menu of popular network services. Researchers and analysts can select and configure the data query that suits their needs.



Figure 3: Data import sources for NodeXL

We used the NodeXL Twitter data import feature (figure 4) to extract networks for six different types of topics. Several options are available to refine data requests using this network data importer. We configured our query using settings as pictured. We requested that NodeXL also add an "edge" which describes the connection between two twitter users that is formed when they follow, reply or mention one another. Data about each user along with the contents of their latest tweet were also selected to be added to the data set.

Import from Twitter Search Network
Why this might take a long time: Twitter rate limiting
Add a vertex for each person who tweeted this search term, or who was replied to or mentioned in those tweets:
Term1 OR Term2
Advanced search help
Add a <u>n</u> edge for each:
V "Replies-to" relationship in tweet
V Tweet that is not a "replies-to" or "mentions"
Follows relationship (very slow!)
Limi <u>t</u> to 18,000 (*) tweets
☑ Add a Tweet column to the Edges worksheet
Expand URLs in tweets (slower)
Add statistic columns to the Vertices worksheet
Your Twitter account
 I don't have a Twitter account. (Twitter discontinued this option in June 2013.)
I have a Twitter account, but I have not yet authorized NodeXL to use my account to import Twitter networks. Take me to Twitter's authorization Web page.
I have a Twitter account, and I have authorized NodeXL to use my account to import Twitter networks.
OK Cancel

Figure 4: NodeXL data importer for Twitter search networks

The NodeXL Twitter Search network data collector starts by performing a query against the Twitter Search service at <u>http://search.twitter.com</u>. Searches can be performed for any string of characters (including the use of operators like "OR"). This service returns up to 18,000 tweets that contain a requested search string. In practice, Twitter rarely returns 18,000 tweets. An age limit of about a week reduces the total set of messages Twitter will return. Studying longer time periods requires repeated data collections.

The resulting set of up to 18,000 tweets is then processed further by NodeXL. Data is assembled from the results of many queries to Twitter about the connections among the authors in the data set. The results are displayed in a NodeXL worksheet labeled "Edges" in the workbook (Figure 5).

	🚽 🔊 • 🖓 •	Ŧ		20	14-01-21 18-02-	31 NodeXL Gr	aph Serve	er pew interr	net.xlsx - N	Aicrosoft Excel	_		_		D X
F	ile Home	Insert Page Layo	ut Formulas	Data Review	View D	eveloper	NodeXL							۵ (1	- # ×
1	Import *	📩 Show Graph 🛯 🛼 Ty	pe: Directed		Color 👘 V	ertex Shape *		🔢 Graph	Metrics *		🐻 Import	Workbook Colum	nns 🔹 👩	🔘 Online 🔻	
🐺 Export * Σ Summary 🔅 Layout: Harel-Koren * 🏧 🔳 🕻		Opacity 🛛 🚍 V	ertex Size	u.	🛃 Subgr	aph Images	, <u> </u>	🛃 Export	Graph Elements	- 🕑	👔 Sample Wo	rkbook			
EF.	Prepare Data *	Automate		Autofill	/isibility - = E	dae Width	Dynami	Group	s *	Use Curren	t 🚺 Reset All	Charles Show Notification	About		
13.5	Data	Gra	nh	Columns	/isual Properties	age man	Filters	Analysi	- -	TOTINEW	ntions	Show/Hide		Help	
-		6 6	Pri -		in a second s			/ maij si	-		ptions	Show, mac		Theip	
	Al	▼ (= Jx											,		¥
	A	В	К	N	0	Р		Q	R	S	т	U	V	W	X 🔺
1		_!	Graph Metrics	Other Columns											
				Add Your Own		Relationship	Date		URLs in	Domains in	Hashtags in		Twitter Page	for	
2	Vertex 1	Vertex 2	Reciprocated?	Columns Here 🔽	Relationsh 💌	(UTC)	*	Tweet 💌	Tweet 💌	Tweet 💌	Tweet 💌	Tweet Date (UTC) 🖵	Tweet	Latitude	Longitude
3	thecreditwomar	peppermiller	No		Mentions	11/24/20	13 10:08	RT @peppe	ermiller: 15	5% of American	adults 18+ (sti	11/24/2013 10:08	https://twitte	er.com/#!/thecre	editwoman
4	danangtriprhnti	faraumainaina	No		Replies to	11/24/20	13 14:01	@farauma	inaina iye	nih pew interne	et gue lg masala	11/24/2013 14:01	https://twitte	er.com/#!/danar	ngtriprhntr
5	diyauthor	diyauthor	NO		Iweet	11/24/20	13 15:10	Tablet and	nttp://www	pewinternet.or	g	11/24/2013 15:10	https://twitte	er.com/#!/diyau	thor/statu
0	novelistscorner	diyautnor	NO		Mentions	11/24/20	13 15:13	RI @DIYa	nttp://www	pewinternet.or	g - 1 - 6 6 - 1 - 6 6 - 1 - 1	11/24/2013 15:13	https://twitte	er.com/#!/novel	istscorner,
-	alice_fromearth	yarawitngiasses	NO		Dealise to	11/24/20	13 20:23	@Nevnaya	@tarawi	thGlasses on b	olonbolonbol.	11/24/2013 20:23	https://twitte	r.com/#!/alice	fromearth
0	scimaccoc	revnaya	No		Turent	11/24/20	13 20.23	E sites too	w Tarawi	newinternet or		11/24/2013 20:23	https://twitte	r.com/#!/aiice	nomearth
10	nowmodianci	christinguiora	No		Montions	11/24/20	12 10-46	DT @Chrie	tipe//iero: I	F7% use their s	sil phone to ge	11/24/2013 23.32	https://twitte	r.com/#!/ssima	adianci/statt
11	apgeltorrer	angeltorrer	No		Tweet	11/25/20	12 15:22	Internet a	theviera.	pewinternet or	en priorie to go	11/25/2013 10:40	https://twitte	r.com/#!/newin	torrer (stat
12	Irainie	Irainie	No		Tweet	11/25/20	13 15:56	Our mant	http://pew	pewinternet.or	6 g newinternet o	11/25/2013 15:56	https://twitte	ar.com/#l/Iraini	e/status //
13	oceanshaman	Irainie	No		Mentions	11/25/20	13 15:57	PT @lrain	http://pew	pewinternet or	g pewinternet.c	11/25/2013 15:57	https://twitte	r.com/#l/ocean	chaman/c
14	baibulao	Irainie	No		Mentions	11/25/20	13 15:03	PT @lrain	http://pew	pewinternet.or	g pewinternet.c	11/25/2013 15:03	https://twitte	ar.com/#!/baibu	lao/status
15	newresearch	Irainie	No		Mentions	11/25/20	13 16:10	PT @lrain	http://pew	pewinternet or	g pewinternet.c	11/25/2013 16:10	https://twitte	r.com/#l/pewre	search/sta
16	pewinternet	Irainie	Vec		Mentions	11/25/20	13 16:10	RT @lrain	http://pew	newinternet or	g newinternet o	11/25/2013 16:10	https://twitte	r com/#l/pewin	ternet/sta
17	zaidbeniamin	Irainie	No		Mentions	11/25/20	13 16:10	RT @lrain	http://pew	newinternet or	g newinternet o	11/25/2013 16:10	https://twitte	er com/#l/zaidb	eniamin/st
18	msmatchgirl	Irainie	No		Mentions	11/25/20	13 16:10	RT @lrain	http://pew	pewinternet or	g pewinternet o	11/25/2013 16:10	https://twitte	ar com/#1/msma	tchgirl/sta
19	kimplicate	Irainie	No		Mentions	11/25/20	13 16-13	RT @lrain	http://pew	newinternet or	g newinternet o	11/25/2013 16:13	https://twitte	er com/#l/kimpl	icate/stati
20	thatchea	Irainie	No		Mentions	11/25/20	13 16:14	PT @lrain	http://pew	pewinternet or	g pewinternet.c	11/25/2013 16:14	https://twitte	ar com/#1/thats	hea/status
21	iavkavdee	Irainie	No		Mentions	11/25/20	13 16:15	RT @lrain	http://pew	pewinternet or	g pewinternet o	11/25/2013 16:15	https://twitte	r com/#!/iavka	vdee/statu
22	nebraskadan	Irainie	No		Mentions	11/25/20	13 16:20	RT @lrain	http://pew	newinternet or	g newinternet o	11/25/2013 16:20	https://twitte	er com/#l/nebra	skadan/st
23	tabitamoreno	Irainie	No		Mentions	11/25/20	13 16:27	RT @lrain	http://pew	pewinternet.or	g pewinternet.o	11/25/2013 16:27	https://twitte	er.com/#!/tabita	moreno/st
24	waepoint	Irainie	No		Mentions	11/25/20	13 16:54	Data = tar	http://t.co.	tcotco	mrx	11/25/2013 16:54	https://twitte	r com/#!/waep	oint/status
25	mikewaterman	Irainie	No		Mentions	11/25/20	13 16:56	RT @lrain	http://pew	pewinternet.or	g pewinternet.o	11/25/2013 16:56	https://twitte	er.com/#!/mikev	vaterman/
26	dmmikhail	Irainie	No		Mentions	11/25/20	13 17:08	RT @lrair	http://pew	pewinternet.or	g pewinternet.o	11/25/2013 17:08	https://twitte	er.com/#!/dmmi	khail/stati
27	barrywellman	Irainie	No		Mentions	11/25/20	13 18:31	RT @lrair	http://pew	pewinternet.or	g pewinternet.c	11/25/2013 18:31	https://twitte	er.com/#!/barry	wellman/s
28	tikidaisy	Irainie	No		Mentions	11/25/20	13 19:13	RT @lrair	http://pew	pewinternet.or	g pewinternet.o	11/25/2013 19:13	https://twitte	er.com/#!/tikida	isy/status
29	guada_wut_	pew_pew_peww	No		Mentions	11/25/20	13 19:50	@KevinKir	gdomKK @	jerryispro1 @	pew_pew_pewv	11/25/2013 19:50	https://twitte	er.com/#!/guada	a wut /sta
30	guada_wut_	jerryispro1	No		Mentions	11/25/20	13 19:50	@KevinKin	gdomKK @	jerryispro1 @	pew_pew_pewv	11/25/2013 19:50	https://twitte	er.com/#!/guada	a wut /sta
31	guada_wut_	kevinkingdomkk	No		Replies to	11/25/20	13 19:50	@KevinKin	gdomKK @	jerryispro1 @	pew_pew_pewv	11/25/2013 19:50	https://twitte	er.com/#!/guada	a wut /sta
32	upayr	Irainie	No		Mentions	11/25/20	13 20:17	RT @lrair	http://pew	pewinternet.or	g pewinternet.c	11/25/2013 20:17	https://twitte	er.com/#!/upayr	/status/40
33	pew_pew_peww	pew_pew_peww	No		Tweet	11/25/20	13 22:58	How much	http://ask.	ask.fm		11/25/2013 22:58	https://twitte	er.com/#!/pew_p	oew peww
34	omnimon	mikesilver	No		Replies to	11/26/2	013 0:59	@MikeSilv	er pew pev	w pew internet i	missiles!	11/26/2013 0:59	https://twitte	er.com/#!/omnir	mon/statu:
35	hipstina	Irainie	No		Mentions	11/26/2	013 4:17	RT @lrair	http://pew	pewinternet.or	g pewinternet.c	11/26/2013 4:17	https://twitte	er.com/#!/hipsti	ina/status/
36	rehab_euphobia	rehab_euphobia	No		Tweet	11/26/2	013 5:48	Pew!Finall	y^.^Long ti	ime no internet.		11/26/2013 5:48	https://twitte	er.com/#!/rehab	euphobia
37	yourbct	yourbct	No		Tweet	11/26/20	13 14:25	9% of non	http://9	9	print	11/26/2013 14:25	https://twitte	er.com/#!/yourb	ct/status/
38	angeltorres	angeltorres	No		Tweet	11/26/20	13 14:49	The Diagn	http://pew	pewinternet.or	g	11/26/2013 14:49	https://twitte	er.com/#!/angel	torres/stat
39	cathybazinet	cathybazinet	No		Tweet	11/26/20	13 15:08	Internaute	http://pew	pewinternet.or	health20fr	11/26/2013 15:08	https://twitte	er.com/#!/cathy	bazinet/sta
40	oiiq	cathybazinet	No		Mentions	11/26/20	13 15:13	RT @Cathy	Bazinet: In	nternautes et ma	health20fr	11/26/2013 15:13	https://twitte	er.com/#!/oiiq/s	tatus/405
41	paragvr	paragvr	No		Tweet	11/26/20	13 15:15	#infograp	http://www	scoop.it	infographic Pa	11/26/2013 15:15	https://twitte	er.com/#!/parag	vr/status/
42	getwellable	getwellable	No		Tweet	11/26/20	13 15:30	Chronic d	http://mob	mobihealthney	vs.com	11/26/2013 15:30	https://twitte	er.com/#!/getwe	llable/stat
43	appsmedical	appsmedical	No		Tweet	11/26/20	13 15:42	Just 72 pe	http://mob	mobihealthney	vs.com	11/26/2013 15:42	https://twitte	er.com/#!/appsr	medical/sta
44	angeltorres	angeltorres	No		Tweet	11/26/20	13 15:57	The Diagn	http://pew	pewinternet.or	2	11/26/2013 15:57	https://twitte	er.com/#!/angel	torres/stat
14	Edges	/ vertices / Groups ,	Group Vertices	Overall Metrics	C Group Edges	Words /	word P	ars / Tw	itter Seal	4					
Rei	ady 📶												HE L L 9	0% —	(+)

Figure 5: NodeXL displaying "Edge List" connections between Twitter users who posted a tweet containing the search term.

Each "edge" represents a connection event between two people who tweeted within the data sample period. Edges can represent the various kinds of relationships that can be created through Twitter. NodeXL constructs four different types of Twitter edges from the data it collects: follows, replies, mentions and tweet. A "follows" edge is created if one author follows another who also tweeted in the sample dataset (the time stamp for a follows edge is the date of the query rather than the time when one user followed another user, which is information that is not available from Twitter). A "mentions" edge is created when one user creates a tweet that contains the name of another user (indicated with a preceeding "@" character, ex: "*just spoke about social media with @marc_smith*"). A "reply" relationship is a special form of "mention" that occurs when the user's name is at the very start of a tweet (ex: *"@itaih just spoke about social media*"). A tweet is a message that does not contain a reply or mention.

Metrics and a range of other processing steps are then calculated and performed using the NodeXL "Automate" feature. Using "Automate" NodeXL executes a series of user configured operations on the network without direct user control. The Automate dialog provides a good summary of the steps and operations applied to each network graph (see figure 6).

www.pewresearch.org

Automate	×
Run these tasks:	
 Count and merge duplicate edges Group by cluster Graph metrics Autofill columns Subgraph images Show graph Save workbook to a new file if it has never been saved Save image to file Export graph to NodeXL Graph Gallery Export graph to email 	Select All Deselect All Options
On this workbook On every NodeXL workbook in this folder: Browse (Excludes open workbooks)	
Run	Cancel

Figure 6: NodeXL Automate dialog contains a list of the analysis steps that can be applied to each network

A raw network is often processed through a number of these steps before the final result is ready for analysis and display. These steps can be selected and configured through the Automate dialog.

Sometimes it is useful for multiple edges in a network edge list to be merged together, particularly if the goal is to create an aggregate picture of all connections between two participants. In our analysis we did not merge duplicate edges in order to retain details about the individual connections among users.

Many networks can be decomposed into smaller sub-groups or regions based on differences in the ways groups of vertices or users connect to one another. The Clauset-Newman-Moore clustering algorithm is one approach of three offered by NodeXL (along with the Wakita and Tsurumi and Newman-Girvan algorithms) out of many thousand potential algorithmic methods for creating sub-groups from the larger population. These approaches generally divide the network guided by the ways some people connect to one another more than to other groups. Often networks have several densely interconnected but separate groups of people who connect more to themselves than to others.

A range of measures of the graph can be calculated for each vertex in the network and for the network as a whole (figure 7).

PEW RESEARCH CENTER

Graph Metrics	and the second		x
Metrics to calculate and	insert into the workbook:		
🔽 Overall graph metri	cs		
Vertex degree (und	rected graphs only)	Select A	JI
Vertex in-degree (d	irected graphs only)		5
Vertex out-degree (directed graphs only)	Deselect	AII
Vertex eigenvector	centrality		
Vertex PageRank	oondany	Ontions	
Vertex clustering co	pefficient		
Vertex reciprocated	vertex pair ratio (directed graphs only)		
Edge reciprocation	(directed graphs only)		
Words and word pa	irs		
Top items			
V Twitter search netw	rork top items		
Overall graph metr	ics		Ξ
-			
The following overa	Il metrics get inserted into the Overall Metrics	worksheet:	
Graph Type	Directed or undirected.		
Vertices	The number of vertices in the graph.		
Unique Edges	The number of edges that do not have dupli	cates.	
Edges With	The number of edges that have duplicates.		Ŧ
About duplicate edges	Calculate Metrics	Cancel	

Figure 7: Network Graph Metrics that can be calculated by NodeXL

Each of the network metrics captures a different dimension of the size and shape of the graph as a whole and the location and connection properties of each person or entity in the network graph. We selected the creation all of the network metrics available through NodeXL. A good description of these measures can be found on Wikipedia (See: <u>http://en.wikipedia.org/wiki/Centrality</u> and <u>http://en.wikipedia.org/wiki/Social_network_analysis</u>).

Many of these metrics can be mapped to various network display attributes. For example, the size of a vertex representing a Twitter user can be scaled to represent the number of users who have chosen to Follow each user. NodeXL has a feature called "Autofill Columns" that makes it simple to pick attributes about each edge and vertex and map them to display attributes like the size, color, shape, or transparency of each vertex (figure 8).

9 PEW RESEARCH CENTER

When Autofill is clicked, fill in these worksheet columns	using the values in these source columns, if the columns exist	Options
Vertex Color		>
Vertex Shape	In-Degree 👻	>
Vertex Size:	Followers -	>
Vertex Opacity	Followers -	-
Vertex Visibility		-
Vertex Label	Vertex 👻	-
L Fill Color		-
Vertex Label Position		-
Vertex Tooltip		-
Vertex Layout Order	Followers -	-
Vertex X		-
Vertex Y		>
Vertex Polar R		>
Vertex Polar Angle		-

Figure 8: NodeXL Autofill Columns dialog maps User attributes to display attributes

Using Autofill columns, we set the size of each vertex to be proportionate to the number of Followers the user had attracted. We also set the "opacity" of the vertex (which controls the transparency of the image) to be inverse to the number of Followers. This setting is controlled by the options which are accessed by the arrow on each row of the Autofill columns dialog. An inverse mapping for opacity means that the largest objects are somewhat transparent, allowing the other, smaller users to be seen through the images of the larger, more popular users. The vertex label is set to the name of the user. The vertex label is drawn beneath the profile photo from Twitter that is used to represent each user. The order of the layout controls the way each vertex is placed on the screen. Setting it to equal the same value as the size value means that objects often line up in size order.

Small thumbnail network images of the local connections around each vertex are created through the Subgraph Images feature (figure 9). These images are inserted into the Vertices worksheet (see figure 12.). Each image offers a quick summary of the "local neighborhood" that surrounds each person.

10 PEW RESEARCH CENTER

Levels of adjace include in each	ent vertices to subgraph: aph image files in folde	15 •	
			Browse
Image size	(pixels)	Image format:	
Width:	200	PNG (.png) 👻	
Height:	200		
Insert subg Thumbnai	aph thumbnails into the size (pixels)	e Vertices worksheet (can be s	low)
Width:	76 🜩		
Height:	50 🚖		
Create cub	raph images for select	ted vertices only he vertex	
In each veri	ex's subgraph, select t	he vertex's incident edges	

Figure 9: NodeXL Subgraph Image Options

We configured the creation of these images so that a small thumbnail was inserted into the worksheet for each user. Optionally, these images could be written to the local file system. In addition, we selected to have the focal person (the "ego") be highlighted.

The network visualization has many other possible configurations that control, for example, the default size for labels, the color of the background behind the network visualization, the color of selected edges, and the size and opacity of edges and their arrow heads (figure 10).

11 PEW RESEARCH CENTER

3 Graph Options
Edges Vertices Other
Color:
Shape: Disk -
Size (simple 1.5 💭 shapes):
Size (images):
 Actual size
Opacity: 100 🚔 %
Effects
◎ None
Orop shadow (slow)
○ Glow (slow) Width: 3.0 →
Selected vertices
Color:
When a vertex is clicked, select its incident edges
Reset All OK Cancel

Figure 10: NodeXL Network Graph Visualization Options

NodeXL also allows the user to configure the images it can create (Figure 11). We set our images to render at a large 4096 by 3072 pixels, which provides a good balance between size and detail. Thumbnails of these images provide a reasonable amount of overview detail while fine details, like the labels on nodes, can be seen when the image is zoomed to full scale.

12 PEW RESEARCH CENTER

Image O	ptions	-		-		x
When	saving an	image to a fil	e, make	the image th	is size:	
01	The size o	f the graph pa	ane:			
	Width:	1,014	units			
	Height:	644	units			
۲	This size:					
	Width: Height:	4096 🔹 3072 🚖	units units	Preserv for best	ve aspect ra t fidelity	itio
V Inc	lude a hea I media net	ader: work connect	tions amo	ng Twitterus	sers	
V Inc	lude a foo	ter:				
Create	ed with No	deXL (http://r	nodexl.co	deplex.com)	from the So	cial Med
Hea	ader/Foote	er Font				
<u>What a</u>	are the uni	ts?	C	ОК	Can	cel

Figure 11: NodeXL Network Graph Image Options

In an "edge list" (Figure 5), each person can appear many times, once for each relationship event in which they participate. A unique list in which all of the participants in the network appear only once is displayed in a separate "Vertices" worksheet (figure 12) in the NodeXL workbook.

A "Vertex" is a node or entity that exists in a network graph. Each Twitter user is a vertex in this network. Each vertex can have a set of attributes and related network metrics that measure their position within the larger network. NodeXL combines network metrics it calculates about each person in the network with data extracted from Twitter that describe, for example, the number of people the user follows, the number of users following that user, the number of Tweets the user has created to date, the number of tweets that person made a Favorite, their self description test from the user profile, their Web URL (if any), the Time Zone name the user selected as their own, the same Time Zone in terms of seconds offset from UTC or GMT, Date the user Joined Twitter (UTC), the text of the Tweet captured in the search results, the list of any URLs included in the Tweet, the Date (UTC) of the Tweet, the Latitude and Longitude (if any) included with the tweet which may represent the location of the user at the time of the tweet.

X	📕 🖓 - 🕅 -	T	-		-		2014-01-2	1 18-02-31 Node)	KL Graph Server p	oew internet.xls	x - Microsoft Exe	cel		-	_	a x	
F	ile Home	Insert	Page L	ayout.	Formulas	Data	Review View	w Developer	NodeXL						۵ () — F	×
1	Import *	🖄 Refre	sh Graph	式 Туре:	Directed	*	Color	Vertex Sh	nape *	🔢 Graph Me	trics *	Import 🛃	Workbook Columns	· ?	Online 🗧		I
1	Export * 2: Summary 🔅 Layout: Harel-Koren * Docity 🖓 Vertex Size								ze Dynamic	Dynamic				ph Elements * 🔛 Sample Workbook			
	Prepare Data *	🧭 Autor	mate			Ċ	olumns 🐵 Visibili	ty 👻 🔣 Edge Wie	dth Filters	Groups *	for	New 🔮 Reset All	Show Notifications		🕕 About		
	Data			Graph			Visual	Properties		Analysis		Options	Show/Hide		Help		
	A1	- (f_{x}													~
	A		R	S	т	U	v	W	х	Y	Z	AA	AD	AE	AF	AG	
1				Graph Met	rics								Other Columns				
1	Vortov	- Sub	graph x	Dograa X		Out-	Betweenness	Closeness	Eigenvector	DageDapk x	Clustering	Reciprocated Vertex	Add Your Own				
2	Vertex	June Sub	graph	Degree	III-Degi ·	Degree	centranty 🕶	Centrality -	Centrality -	Fagekalik ·	Coefficient 1		columns here	Jwed -	ronowers ·	I WEELS •	
		2	466														
3	pewinternet				435		5 810452.133	0.000	0.044	171.649	0.000	0.007	·	126	50767	4595	
		1															
4	scoopit	1	10.5		52		391330.000	0.000	0.000	18.281	0.004	0.000)	702	51363	3750	
			1000														
-	10						453003.054	0.000	0.000	40.007	0.004	0.000			450000		
5	libraryjournal				58		1 15/08/.961	0.000	0.000	18.297	0.001	0.000		250	152808	8193	
1			\mathbf{i}														
6	frouellet		~		0	:	2 150096.000	0.000	0.000	0.712	0.000	0.000)	490	335	1160	
7	tinkuv	1	S		0		130832.000	0.000	0.000	0.751	0.000	0.000)	808	1987	4033	
			\sim														
			Y .														
8	stephane_ozil	-			4		1 129354.000	0.000	0.000	1.420	0.000	0.000		658	1648	10475	
			×.														
9	_jg_up_		×		0		5 126221.000	0.000	0.000	1.873	0.133	0.000)	1305	7120	1734	
		1															
10	pewresearch	12	and P		80		2 124337.533	0.000	0.005	22.877	0.012	0.012	2	28	134157	15740	
		-21	segn.														
11	viumfr	台	1000		27	2	117212.000	0.000	0.000	16 495	0.007	0.140		1041	1902	6166	
	VIGEN	1			57		11/512.000	0.000	0.000	10.405	0.007	0.140		1041	1055	0100	
12	queerspawn		4		0	1	3 116440.000	0.000	0.002	1.093	0.333	0.000)	1085	707	3984	
		C	\rightarrow														
13	cynthiajabar		Δ		1	:	116440.000	0.000	0.002	1.093	0.333	0.000)	2040	2144	8229	
			<u></u>														
14	n4ik3	0	<u>~</u>		0		2 116440.000	0.000	0.002	0 783	0.000	0.000		808	418	321	
		تغو	1000		Ŭ			2.500	2.002		2.000	0.000			.10		
15	2morrowknight		22.2.4		54		102699.067	0.000	0.003	20.031	0.003	0.000		4/1836	593961	91097	
			1.5														
16	susannahfox	1	all a start		83		69462.667	0.000	0.006	24.384	0.012	0.023	5	847	16120	19685	
		Ő															
17	Irainie	200	7.00		26		47058,867	0.000	0.003	11.255	0.013	0.032		933	8871	5371	Ţ
H	► ► Edges	Vertice	es / Group	os 🖉 Grou	p Vertices	/ Overa	I Metrics / Grou	p Edges / Word	ds / Word Pairs	Twitter !	Seal 4			-		Þ	
Rea	ady 🎦													90%	6 -	U(÷

Figure 12: NodeXL Vertices worksheet displaying a network subgraph image ("egonet") along with network and Twitter metrics for each vertex.

The Vertex worksheet (figure 12) displays attributes from Twitter along with a range of network metrics for each vertex in the network. Each row represents one Twitter user who appeared in the results of the initial search query. A set of network metrics is listed for each user. Network metrics capture a range of qualities about the location and connection pattern of each user within the larger network. Sorting users by each metric can bring different users to the top of the list, highlighting people who occupy various kinds of positions within the network graph, some of whom may be considered to be "influential" or strategically located. "Degree" is the count of all unique connections each person maintains, while in- and out-degree capture the number of link to and from each user. Betweenness, Closeness, Eigenvector Centrality and PageRank all capture various ways in which each user is in the "middle of things". Clustering coefficient measures how closely connected each user's connections are connected to one another. This measure is visualized (in column B) in the form of a "sub-graph" – a small network graph that captures just the selected user and the connections among their immediate connections. Sorting by "Betweenness Centrality," for example, sorts people who have the quality of most broadly connecting across the network to the top.

When sorted by betweenness centrality, this worksheet displays a ranked list of the most central and, arguably, influential, people in this network. These users have strategic locations within the network created by their pattern of connections (and the connection pattern of others).

Sorting by other metrics can bring to the top of the list other people who have little or no connections to the large connected set of people at the core of the network. These people are called "isolates" when they have no connection at all to anyone else seen in the data sample. The presence or absence of many isolates is an indication of how "public" or "popular" a topic is. Very popular topics are talked about by many people who otherwise have little or no connection. In contrast, topics and keywords specific to a group often have few or no isolates.

Using NodeXL's network visualization feature, we can display an overview visualization of the collection of connections observed (figure 13). This map is dominated ______.



Figure 13: NodeXL network visualization of the connections among people who tweeted "Pew Internet."

NodeXL offers a filtering tool that can remove selected vertices from the network. It is sometimes useful to exclude certain people, for example the isolated users, in order to focus on other features of the network.

15 PEW RESEARCH CENTER

V Dynamic Filters	X
Edge Filters	·
Width:	
3.00	10.00 🜩
Opacity	
Opacity.	
	22 4
	32 👳
	E
Relationship Date (UTC):	
and the state of the	
11/24/2013 10:08 AM 🗐 🗸 刘 1/22/2014 12:	00 AM 🔲 🔻
Ť Ť	
Tweet Date (UTC):	
a data	
11/24/2013 10:08 AM	00 AM 🔲 🚽
Edge Weight:	
	35 🌩
Vertex Eller	
Circo	
5126.	
	1 000 00 1
	1,000.00
Opacity:	
70	100 🌩
Lavout Order:	
100	9 999 00 📥
	3,333.00 👻
X:	
240.00	9,771.53 🚔
Y:	
And the second state of the second	
324.00 🔶	9,674.01 🚔 👻
Show edges and vertices it cells are empty	Close
Filter opacity: 0 😴 %	

Figure 14: NodeXL Dynamic Filters allow selected vertices and edges to be removed from the network visualization

Using these tools, it is possible to collect and analyze the social networks that form among people who tweet a common keyword, phrase, term, URL or hashtag. These networks have distinctive shapes that indicate the types of conversations they contain. Network measures identify the relatively few people who occupy strategic locations at various positions within the network. Some people have high betweenness or high degree which may indicate that they are the most popular or influential people in the community. These people can be visually identified using image like Figure 15 or quantified and sorted as seen in figure 10.

www.pewresearch.org

16 PEW RESEARCH CENTER



Figure 15: NodeXL filtered visualization of the connections among the people who tweeted a common term. Only users with at least one connection are displayed.

Key people in each topical community can be identified using this method. In addition, sub-groups or clusters can also be identified. NodeXL places each sub-group in its own region (using a "Group-in-a-box" feature) to improve the clarity of the network visualization.

Summary information about each network can be found on the "Overall Metrics" worksheet of the NodeXL Workbook.

	17	
PEW	RESEARCH CENTER	

🔀 🛃 🧐 🔹 🖓 😴 😓 2014-01-21 18-02-31 Nod	eXL Graph Server p	ew internet.x	lsx - Microso	ft Excel		Table Tools								• ×
File Home Insert Page Layout Formulas	Data Rev	iew Viev	v Develo	per No	deXL	Design							۵ 🕜	- @ ×
Import × ☆ Refresh Graph *→ Type: Directed Data Export × ∑ Summary ☆ Layout: Data Graph	aren • Autofil Columr	Golor vertex Shape Opacity a Vertex Size mns ≪ Visibility ~ ≣ Edge Width Visual Properties			Graph Metrics * Subgraph Images * Filters Groups * Analysis		Use Current for New Options		Workbook Columns *		e Online * Help About Help		Workbook	
🛛 🛛 🗸 🕐 Al 🗸 🗸 🖌 Graph Metr	ic													*
A A Graph Metric	B Value ▼	С	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ 🛓
2 Graph Type	Directed													
3														
4 Vertices	2814													=
5														
6 Unique Edges	2887													
7 Edges With Duplicates	550													
8 Total Edges	3437													
9														_
10 Self-Loops	1586													
11														_
12 Reciprocated Vertex Pair Ratio	0.014526438													_
13 Reciprocated Edge Ratio	0.028636884													_
14 15 Composited Composite	1070													
15 Connected Components	1370													_
16 Single-Vertex connected components	1120													_
17 Maximum Venices in a Connected Component	1255													_
10 Maximum Edges in a Connected Component	1555													
20 Maximum Geodesic Distance (Diameter)	14													
21 Average Geodesic Distance	4 665584													
22														
23 Graph Density	0.000220572													
24 Modularity	0.623941													
25														
26 NodeXL Version	1.0.1.251													
27														
28														
29 Readability Metric	Value 🔽													
30														
31														
32														
33														-
H + H Edges / Vertices / Groups / Group Vertice	es Overall Met	rics / Group	Edges /V	Vords 🖉 V	Vord Pairs	/ Twitter S	iel 4			IIII				
Ready 🛅												100% 😑		+ ,;;

Figure 16: NodeXL Overall Metrics worksheet contains information that describes the size and density of the network

These measures allow users to contrast networks to capture the ways networks in general and social media networks in particular compare to other networks and to themselves overtime.

https://www.nodexlgraphgallery.org/Pages/Graph.aspx?graphID=15414



Large Groups in pew internet Twitter NodeXL SNA Map and Report for 2014-01-21 18-02-31

The graph represents a network of Twitter users whose tweets in the requested date range contained "pew internet." The tweets in the network were tweeted over the 58-day, 10-hour, 47-minute period from Sunday, 24 November 2013 at 10:08 UTC to Tuesday, 21 January 2014 at 20:55 UTC.

There is an edge for each "replies-to" relationship in a tweet, an edge for each "mentions" relationship in a tweet, and a self-loop edge for each tweet that is not a "replies-to" or "mentions".

The graph's vertices were grouped by cluster using the Clauset-Newman-Moore cluster algorithm.

The graph was laid out using the Harel-Koren Fast Multiscale layout algorithm.

The edge colors are based on edge weight values. The edge widths are based on edge weight values. The edge opacities are based on edge weight values. The vertex sizes are based on followers values. The vertex opacities are based on followers values.

Overall Graph Metrics: Vertices: 2814 Unique Edges: 2887

www.pewresearch.org

Edges With Duplicates: 550 Total Edges: 3437 Self-Loops: 1586 Reciprocated Vertex Pair Ratio: 0.0145264381173736 Reciprocated Edge Ratio: 0.0286368843069874 Connected Components: 1370 Single-Vertex Connected Components: 1120 Maximum Vertices in a Connected Component: 959 Maximum Edges in a Connected Component: 1355 Maximum Geodesic Distance (Diameter): 14 Average Geodesic Distance: 4.665584 Graph Density: 0.000220572016763473 Modularity: 0.623941 NodeXL Version: 1.0.1.251

Top 10 Vertices, Ranked by Betweenness Centrality: @pewinternet @scoopit @libraryjournal @frouellet @tinkuy @stephane_ozil @_jg_up_ @pewresearch @viuzfr @n4ik3

🗶 🛛 😓 🔊 🔹 🖓 🕶 🖓 🖛 🖓 🚽 📿 2014-0)1-21 18-02-31 N	odeXL Graph Server pew intern	et.xlsx - Micro	soft Excel Table To	ols			_ 0 _ ×
File Home Insert Page Layout Formulas Data Review View Developer NodeXL Design								
Import * 👘 Refresh Graph	Type: Direc	ted 🔹 🔛 🖧 Col	or 🔊 Ve	rtex Shape -	aph Metrics *	Import	Workbook Columns * 🧼 🙆	Online -
Export T Summary	* Lavout: Harel	-Koren ·	acity 🗖 Ve	rtex Size	boraph Image	Export	Graph Elements	Sample Workbook
	to cayour march	Autofill		Dynamic	ograph mage	Use Current	Help	Abaud
Re Prepare Data * Automate	~ .	Columns and Visi		ge width Filters	oups *	for New O Reset Al	I u Snow Notifications	ADOUT
Data	Graph	Vis	ual Properties	Ana	ilysis	Options	Show/Hide	Help
A1 • (*) Je Top URLs in Tweet in Entire Graph								
A	В	С	D	E	F	G	H I	J
	Entire Graph							
 Top ORLS in Tweet in Entire Graph http://pewinternet.org/Commentary 	(20 38	4 http://feedproxy.google.com	27	http://pewinternet.org/Comm	271	http://pewinternet.org/Com	66 http://www.infodocket.com/2	40 http://
3 http://pewinternet.org/Commentary	/20 19	4 http://www.mediabistro.com	26	http://pewinternet.org/Comm	67	http://www.pewinternet.org	21 http://www.infodocket.com/2	30 http:/
4 http://www.infodocket.com/2013/12	/11 6:	1 http://libraries.pewinternet	. 24	http://www.pewresearch.org	9	http://pewinternet.org/Comr	10 http://www.infodocket.com/2	10 http:/
5 http://www.businessinsider.com/pe	<u>w-</u> 6	0 http://pewinternet.org/Repo	18	http://31.media.tumblr.com/k	3	http://pewinternet.org/Repo	2 http://www.infodocket.com/	3 <u>http:</u>
6 http://www.viuz.com/2014/01/06/usa	ige 5	7 http://www.businessinsider	18	http://libraries.pewinternet.	2	http://libraries.pewinternet	2 http://goodereader.com/blog	2 <u>http:</u>
7 http://www.intodocket.com/2014/01 8 http://www.mediabistro.com/alltwite	1 <u>10</u> 4	http://pewinternet.org/Report http://pews.google.com/news/ http://pews.google.com/news/ http://pewinternet.org/Report http:	16	http://pewinternet.org/Prese	2	http://www.pewinternet.org	2 http://www.pewinternet.org/ 2 http://goodereader.com/blog	2 <u>nttp:</u> 2 http:
9 http://libraries.pewinternet.org/201	3/1 3	6 http://www.mediabistro.com	13	http://www.pewinternet.org/	2	http://susannahfox.com/201	1 http://web.docuticker.com/go	2 http://
10 http://www.viuz.com/2014/01/06/usa	1 <u>Re</u> 31	6 http://internet.lintas.me/art	13	http://www.pewresearch.org	1	http://www.pewresearch.org	1 http://ow.ly/28Nidn	1 http:/
11 http://pewinternet.org/Reports/2013	34 34	4 http://www.businessinsider	13	http://pewinternet.org/Report	1	http://www.pewinternet.org	1 http://libraries.pewinternet.	1 http:/
12								=
26 Top Hachtags in Tweet in Entire Gra	Entire Graph	Top Hashtags in Tweet in	G1 Count	Top Hashtags in Tweet in	G2 Count	Top Hashtags in Tweet in	Top Hashtags in Tweet in	G4 Count T Top I
20 Top Hashtags in Tweet in Entire Gra	- count - g	3 socialmedia	36	1 libraries	32 COUNT -	mH	50 Libraries	40 résea
28 mH	5	2 SocialMedia	20	CES2014	27	mHealth13	3 ereaders	24 Pinte
29 Libraries	5	1 Twitter	11	mH	2	internet	3 ebooks	21 Twitt
30 SocialMedia	4	8 PewResearchCenter	11	SoMe	2	edtech	3 ebok	7 Facel
31 libraries	3	5 social	9	mobile	2	بتترست	3 tablets	3 Linke
32 CES2014	3.	2 twitter 1 internet	9	HigherEd	1	تریش ehinitaldivida	3 demographics	2 journ
34 mobile	3	0 Internet	7	HITsm	1	online	2	photo
35 ebooks	2	8 sm	7	DC	1	valeur	2	socia
36 Twitter	24	4 tech	6	DigitalPatient	1	bibltiothèque	2	Usag
37								
29 Top Words in Twent in Entire Crach	Entire Graph	Top Words in Tweet in	C1 Court X	Top Words in Tweet in	C2 Courts T	lop Words in Iweet in	C2 Count X C4	
39 internet	240	0 internet	1235	pewinternet	401	susannahfox	82 pew	99 usar
40 pew	238	2 pew	1216	use	219	pewresearch	77 internet	94 1
41 social	99	6 social	499	online	200	online	70 releases	78 des
42 american	83	1 american	487	adults	189	adults	68 libraryjournal	57 les
43 life	75	b project	476	18	170	use	68 americans	47 étudi
44 project 45 research	73	u lite 4 research	454	farebook	165	s 91	67 new 63 libraries	45 pew 45 inter
46 use	69	4 2013	256	71	164	u	62 value	44 5
47 adults	65	1 center's	229	17	163	cell	62 public	44 ense
48 online	61	3 media	194	instagram	163	phone	61 communities	44 résea
49	Factor Orach	The Mont Date to Toron in						
50 Graph	Count	G1	G1 Count	G2	G2 Count	G3	G3 Count C4	G4 Count 🔻 G5
51 pew.internet	139	2 pew.internet	657	online.adults	169	u.s	62 pew.internet	90 usag
52 internet,american	734	4 internet,american	457	adults,use	167	adults,cell	62 internet, releases	67 5,ens
53 american,life	71	3 american, life	448	18,twitter	164	91,u	61 libraryjournal,pew	56 pew,
54 life,project	65	0 life,project	415	71,online	163	s,adults	61 americans,value	44 les,5
55 pew,research	57	b pew,research	385	twitter,1/ 17 instagram	163	cerr,phone	61 value, public	44 ense
57 social.media	37	9 center's.internet	229	facebook.18	165	go.online	60 libraries.communities	44 étudi
58 research,center's	35	9 social,media	190	22,linkedin	162	susannahfox,91	59 releases,americans	39 résea
59 center's,internet	35	9 internet,users	161	use,facebook	161	use,phones	58 communities,via	35 des,r 🖕
H + + H / Group Vertices / Overall Metrics / Group Edges / Words / Word Pairs Twitter Search Ntwrk Top Items / Top I								
Ready 🛅 🔠 20% 🔿 🛡 🔶								

Social Media Network Research Related Publications

Hansen, D., Smith, M., Shneiderman, B., EventGraphs: charting collections of conference connections. Hawaii International Conference on System Sciences. Forty-Forth Annual Hawaii International Conference on System Sciences (HICSS). January 4-7, 2011. Kauai, Hawaii.

http://www.cs.umd.edu/localphp/hcil/tech-reports-search.php?number=2010-13

EventGraphs are social media network diagrams constructed from content selected by its association with timebounded events, such as conferences. Many conferences now communicate a common "hashtag" or keyword to identify messages related to the event. EventGraphs help make sense of the collections of connections that form when people follow, reply or mention one another and a keyword. This paper defines EventGraphs, characterizes different types, and shows how the social media network analysis add-in NodeXL supports their creation and analysis. The paper also identifies the structural and conversational patterns to look for and highlight in EventGraphs and provides design ideas for their improvement.

In the Journal of Social Structure: "Visualizing the Signatures of Social Roles in Online Discussion Groups" is available from: <u>http://www.cmu.edu/joss/content/articles/volume8/Welser/</u> It illustrates different patterns of network structures associated with different kinds of roles and behaviors.

Social roles in online discussion forums can be described by patterned characteristics of communication between network members which we conceive of as 'structural signatures.' This paper uses visualization methods to reveal these structural signatures and regression analysis to confirm the relationship between these signatures and their associated roles in Usenet newsgroups. Our analysis focuses on distinguishing the signatures of one role from others, the role of "answer people." Answer people are individuals whose dominant behavior is to respond to questions posed by other users. We found that answer people predominantly contribute one or a few messages to discussions initiated by others, are disproportionately tied to relative isolates, have few intense ties and have few triangles in their local networks. OLS regression shows that these signatures are strongly correlated with role behavior and, in combination, provide a strongly predictive model for identifying role behavior (R2=.72). To conclude, we consider strategies for further improving the identification of role behavior in online discussion settings and consider how the development of a taxonomy of author types could be extended to a taxonomy of newsgroups in particular and discussion systems in general.

"Discussion catalysts in online political discussions: Content importers and conversation starters" in the Journal of Computer-Mediated Communication (JCMC) http://jcmc.indiana.edu/ at http://ping.fm/7NF5T

This study addresses 3 research questions in the context of online political discussions: What is the distribution of successful topic starting practices, what characterizes the content of large thread-starting messages, and what is

the source of that content? A 6-month analysis of almost 40,000 authors in 20 political Usenet newsgroups identified authors who received a disproportionate number of replies. We labeled these authors "discussion catalysts." Content analysis revealed that 95 percent of discussion catalysts' messages contained content imported from elsewhere on the web, about 2/3 from traditional news organizations. We conclude that the flow of information from the content creators to the readers and writers continues to be mediated by a few individuals who act as filters and amplifiers.

Analyzing (Social Media) Networks with NodeXL

Smith, M., Shneiderman, B., Milic-Frayling, N., Rodrigues, E.M., Barash, V., Dunne, C., Capone, T., Perer, A. & Gleave, E. (2009), "Analyzing (Social Media) Networks with NodeXL", In C&T '09: Proceedings of the Fourth International Conference on Communities and Technologies. Springer.

Abstract: In this paper we present NodeXL, an extendible toolkit for network data analysis and visualization, implemented as an add-in to the Microsoft Excel 2007 spreadsheet software. We demonstrate NodeXL features through analysis of a data sample drawn from an enterprise intranet social network, discussion, and wiki. Through a sequence of steps we show how NodeXL leverages and extends the broadly used spreadsheet paradigm to support common operations in network analysis. This ranges from data import to computation of network statistics and refinement of network visualization through a selection of ready-to-use sorting, filtering, and clustering functions.

Whither the Experts

Howard Welser, Eric Gleave, Marc Smith, Vladimir Barash, Jessica Meckes. "Whither the Experts? Social affordances and the cultivation of experts in community Q&A systems", in SIN '09: Proc. international symposium on Social Intelligence and Networking. IEEE Computer Society Press.

Abstract: Community based Question and Answer systems have been promoted as web 2.0 solutions to the problem of finding expert knowledge. This promise depends on systems' capacity to attract and sustain experts capable of offering high quality, factual answers. Content analysis of dedicated contributors' messages in the Live QnA system found: (1) few contributors who focused on providing technical answers (2) a preponderance of attention paid to opinion and discussion, especially in non-technical threads. This paucity of experts raises an important general question: how do the social affordances of a site alter the ecology of roles found there? Using insights from recent research in online community, we generate a series of expectations about how social affordances are likely to alter the role ecology of online systems.

First Steps to Netviz Nirvana

Bonsignore, E.M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D.L. & Shneiderman, B. (2009), "First steps to NetViz Nirvana: evaluating social network analysis with NodeXL", In SIN '09: Proc. international symposium on Social Intelligence and Networking. IEEE Computer Society Press.

Abstract: Social Network Analysis (SNA) has evolved as a popular, standard method for modeling meaningful, often hidden structural relationships in communities. Existing SNA tools often involve extensive pre-processing or intensive programming skills that can challenge practitioners and students alike. NodeXL, an open-source template for Microsoft Excel, integrates a library of common network metrics and graph layout algorithms within the familiar spreadsheet format, offering a potentially low-barrier to-entry framework for teaching and learning SNA. We present the preliminary findings of 2 user studies of 21 graduate students who engaged in SNA using NodeXL. The majority of students, while information professionals, had little technical background or experience with SNA techniques. Six of the participants had more technical backgrounds and were chosen specifically for their experience with graph drawing and information visualization. Our primary objectives were (1) to evaluate NodeXL as an SNA tool for a broad base of users and (2) to explore methods for teaching SNA. Our complementary dual case-study format demonstrates the usability of NodeXL for a diverse set of users, and significantly, the power of a tightly integrated metrics/visualization tool to spark insight and facilitate sensemaking for students of SNA.

Do You Know the Way to SNA?

Hansen, D., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E., Smith, M., Shneiderman, B. (July 2009) Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Data University of Maryland Tech Report: HCIL-2009-17

Abstract: Voluminous online activity data from users of social media can shed light on individual behavior, social relationships, and community efficacy. However, tools and processes to analyze this data are just beginning to evolve. We studied 15 graduate students who were taught to use NodeXL to analyze social media data sets. Based on these observations, we present a process model of social network analysis (SNA) and visualization, then use it to identify stages where intervention from peers, experts, and computational aids are most useful. We offer implications for designers of SNA tools, educators, and community & organizational analysts.