# **Network Analysis For News**

CAR meeting, St Louis, Feb 23 2012

Peter Aldhous



peter@peteraldhous.com

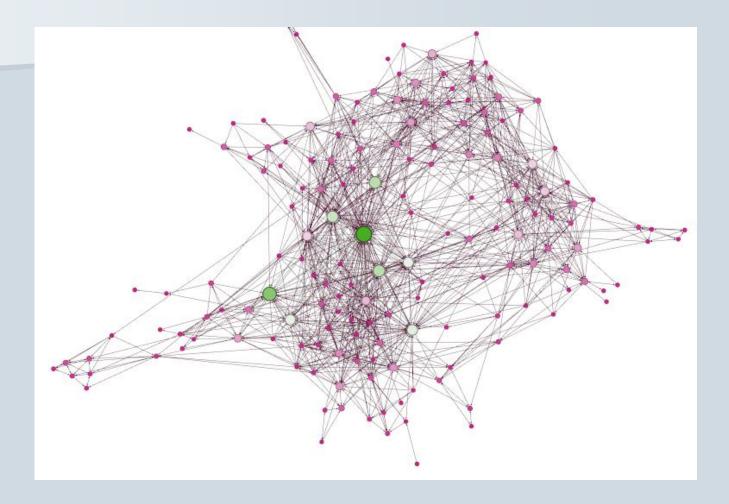
Peggy Heinkel-Wolfe

Denton Record-Chronicle

pheinkel-wolfe@dentonrc.com

# What is network analysis?

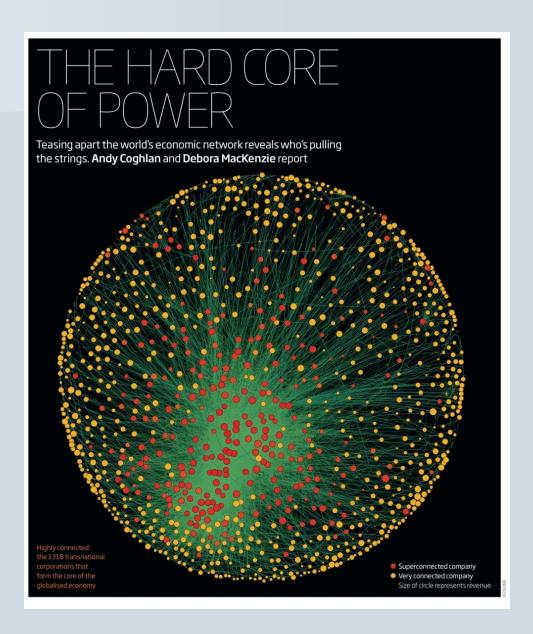
Visualizing and quantifying connections between people, organizations etc.



Graphs consist of edges (the connections) and nodes (the entities connected)

# Who uses network analysis?

Researchers in almost every discipline – it is being used to study everything from how networks of genes and proteins influence our health, to how connections between multinational companies affect the stability of the global economy:



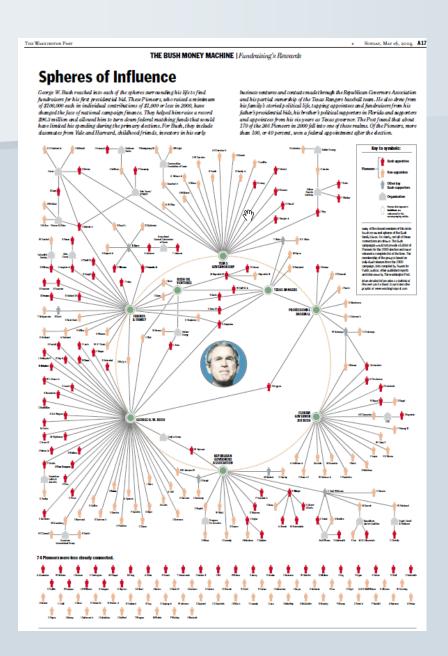
# OK, that sounds cool, but how can it help me?

### Do any of these phrases sound familiar?

- •It's not what you know, it's who you know
- •The problem with this city is that it's run by X's cronies
- Contracts in this town are all about kickbacks

Network analysis provides a means to display these and other connections in a way your readers (and your editors) can easily understand.

## A classic example



# The Washington Post

**Data:** Built database of George W. Bush "Pioneers" – those who raised more than \$100,000 to his 2000 presidential campaign.

Findings: Network showed who were the key Pioneers in the Bush campaign, who they were connected to, and what each Pioneer gained, if anything, from his or her association with Bush, such as ambassadorships and other federal appointments. Bush raised \$96.3 million, a record at that time. More than 100, about 40%, of the Pioneers received some federal appointment after the election.

### Is it difficult?

At its simplest level, network analysis is very intuitive.

#### Some basics:

Networks can be **undirected** (e.g. X and Y went to the same school) or **directed** (e.g. donor X gave money to candidate Y).

Network theorists have devised **layout algorithms** to draw network graphs; you can also lay out your network manually.

**Clustering algorithms** can be used to group similar nodes together, based on their pattern of connections – you can think of them as "clique finders."

### The science bit: network metrics

**Degree** is a simple count of the number of connections for each node. For directed networks, it is divided into **In-degree**, for the number of incoming connections, and **Out-degree**, for outgoing connections.

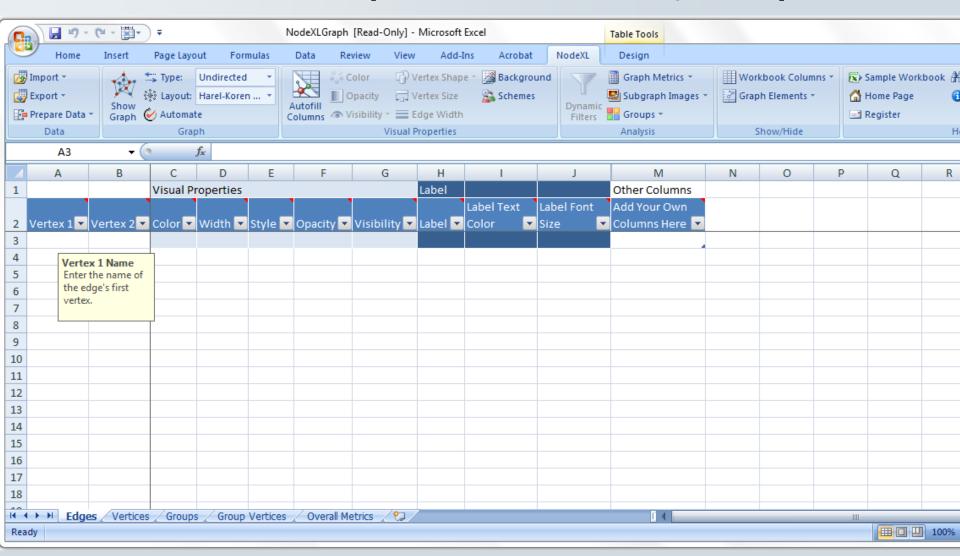
**Eigenvector centrality** accounts not only for the node's own degree, the also the degrees of the nodes to which it connects.

**Betweenness centrality** essentially reveals how important each node is in providing a "bridge" between different parts of the network. It highlights the nodes that, if removed, would cause a network to fall apart.

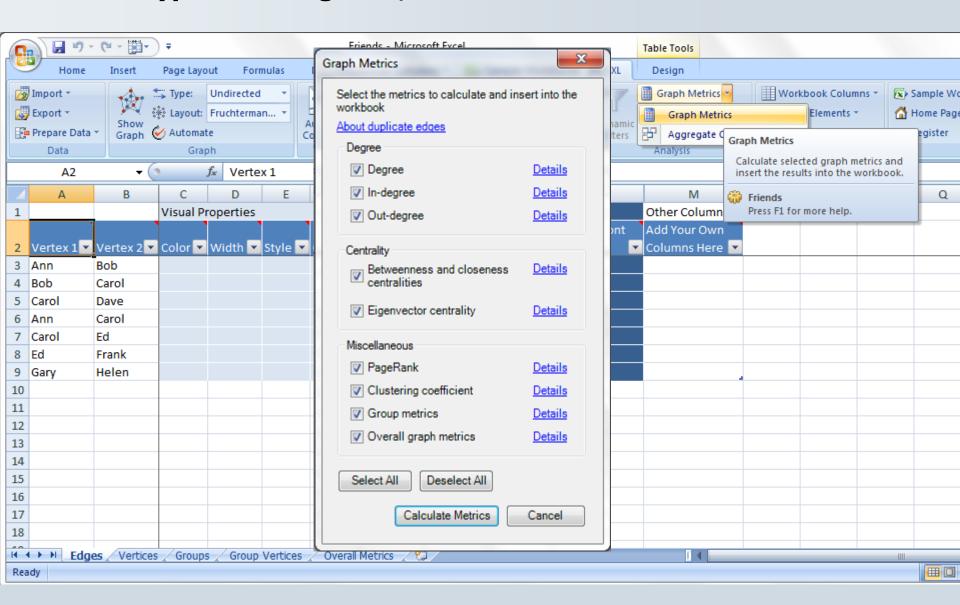
**Closeness centrality** is a measure of how close each node is, on average, to all of the other nodes in a network. It highlights the nodes that connect to the others through a lower number of edges – think <u>Kevin Bacon Game</u>.

## NodeXL: a simple tool for network analysis

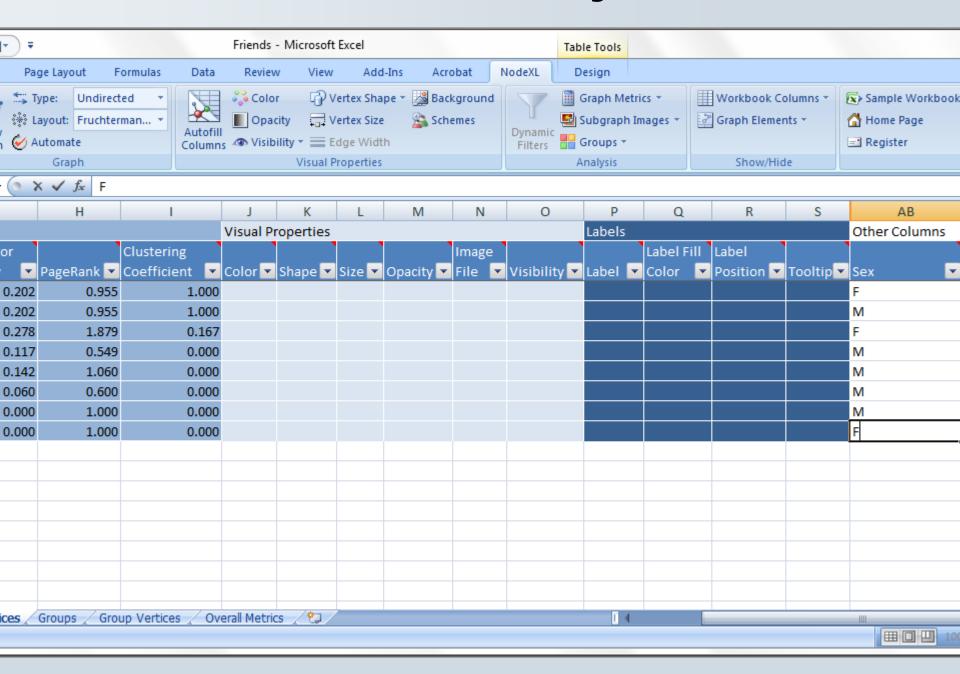
### Download the template from **NodeXL site**, then open:



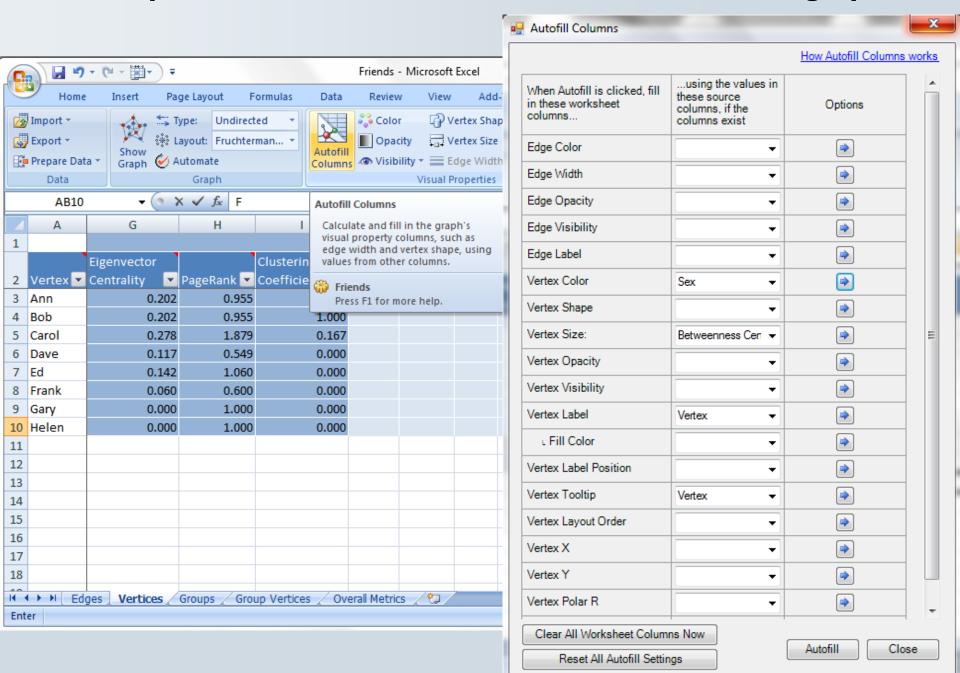
## Type in an edge list, and calculate network metrics



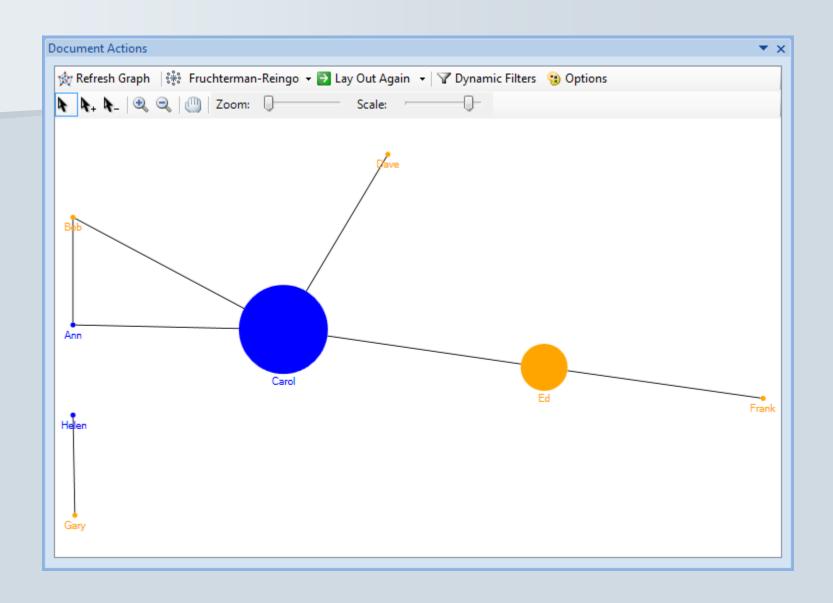
## Add more data: a new column for the gender of each friend



## Incorporate attributes from the data into the network graph



## The resulting friends network:



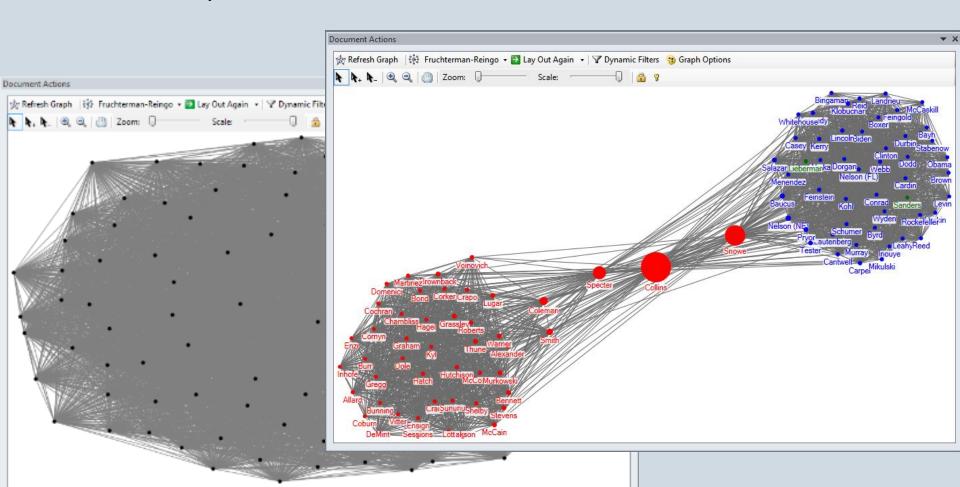
## **Learn NodeXL at NICAR 2012**

**DEMO: 9am Friday** 

PowerPivot and NodeXL: This is not your father's Excel

### **HANDS-ON: 11am Friday**

Network analysis with NodeXL







INSIDE

coupons & savings of \$197

Not in all areas

# Denton Record-Chronicle

An edition of The Ballas Morning News

DentonRC.com

Vol. 108, No. 75 / 42 pages, 4 sections

Sunday, October 16, 2011

Denton, Texas

One dollar

# Grand jury system questioned

'Pick-a-pal' selection process stacks group with political players

By Lowell Brown and Peggy Heinkel-Wolfe Staff Writers

The grand jury that indicted a man who often criticizes local government had multiple ties to City Hall and to the victim of the alleged crime, a Denton Record-Chronicle investigation found. Bob Clifton, a retired businessman, once branded a vexatious litigant by a visiting judge after he filed four lawsuits against the city in two years, stands accused of asking Mayor Mark Burroughs to pay him to abandon the 2010 mayor's race.

One grand jury member who indicted Clifton heads a city department.

One was a former mayor.

One served on Burroughs' re-election committee.

Four had donated to at least one of Burroughs' mayoral campaigns. And despite a state law meant to ensure that grand juries reflect the county's demographic and geographic makeup, most members of the grand jury that indicted Clifton and 600 other people last year were white males, residents of Denton, or both. More than 530,000 people live elsewhere in Denton County, yet 11 of the 12 grand jurors live in Denton, Corinth and Argyle — cities with a combined population of less than 137,000. Women and racial minorities were also underrepresented.

The Record-Chronicle investigated

the grand jury array after reporting that two members — former Mayor Perry McNeill and David Gosdin, husband of the current mayor pro tem, Pete Kamp — were asked by the district attorney to abstain because their names would come up as part of the evidence in the case.

The findings raise questions about the fairness of grand jury selection under the "key-man" system, a method used throughout much of Texas whereby an appointed commission —

See GRAND JURY on 4A

#### INSIDE

Graphic connecting grand jurors to local groups, politicians and campaigns/5A

Grand jurors' demographics/5A

History of the grand jury system/6A

#### ON DENTONRC.COM

Campaign finance reports for Mark Burroughs, Pete Kamp and James King

Documents relating to campaign committees for Burroughs and Kamp, potential grand jurors and grand jury commissioners



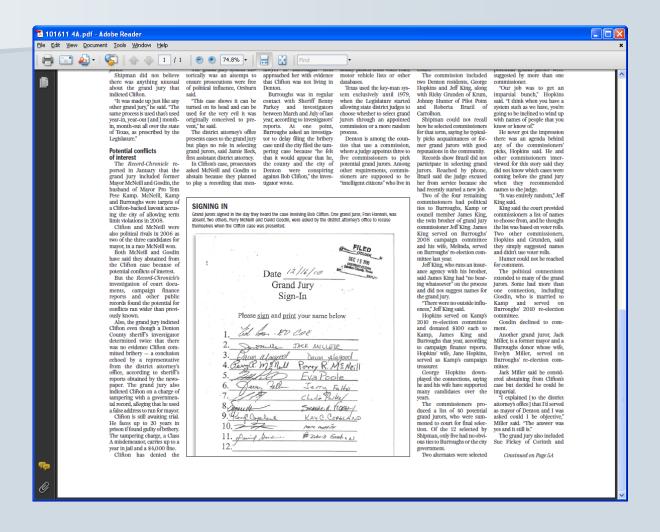
The next Bieber?

Vouna parformar

# Our story's main characters



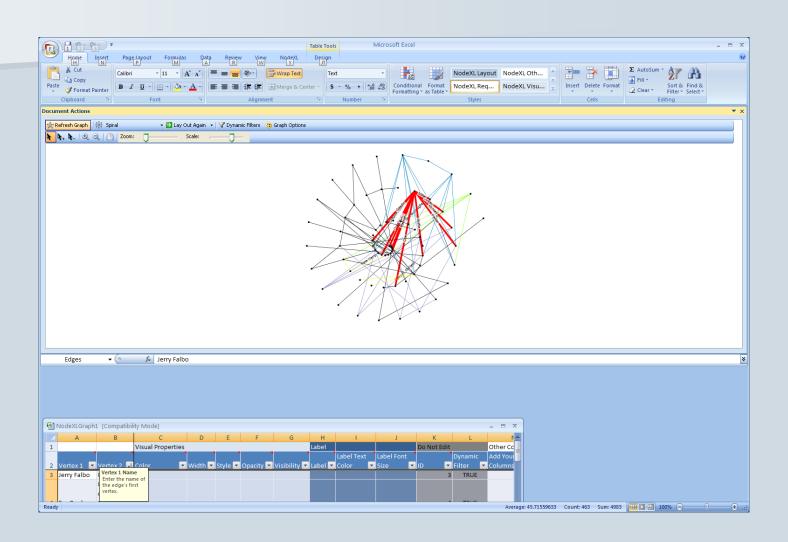
# The grand jury



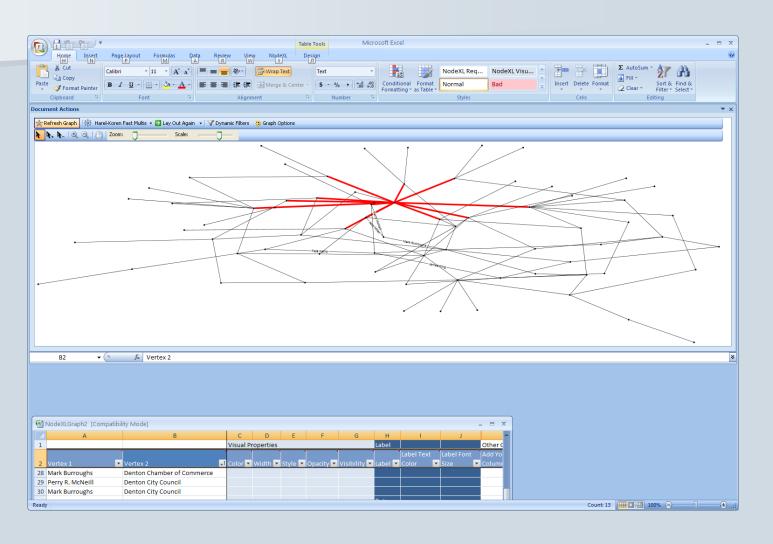
# **Basic demographics**



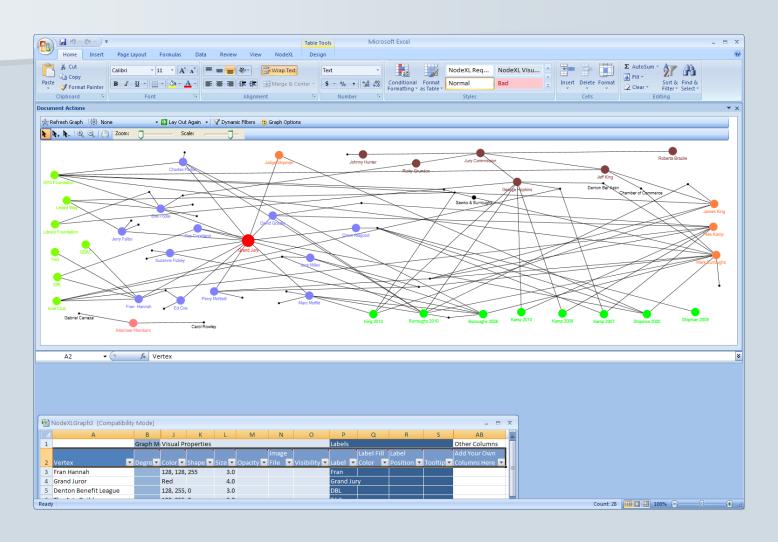
# The first network graph



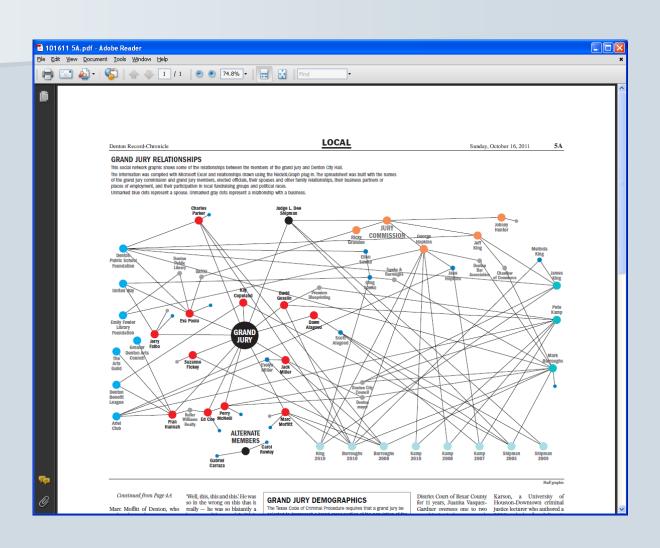
# The first network graph we showed to the editors



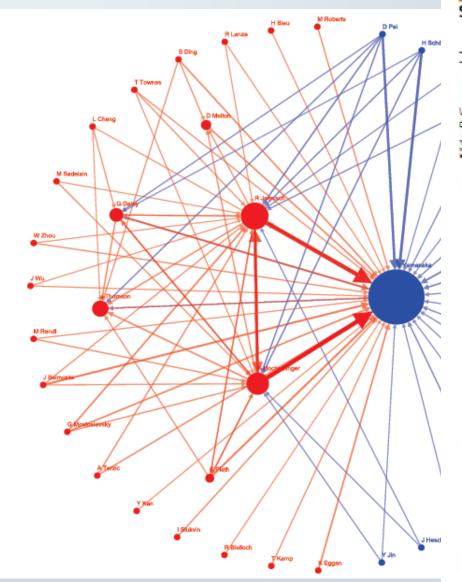
# The second network graph we showed to the editors



# The published network graph



# **Example: citations in stem cell research**

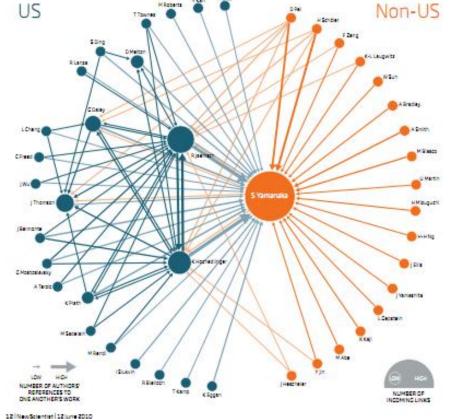


#### SPECIAL REPORT

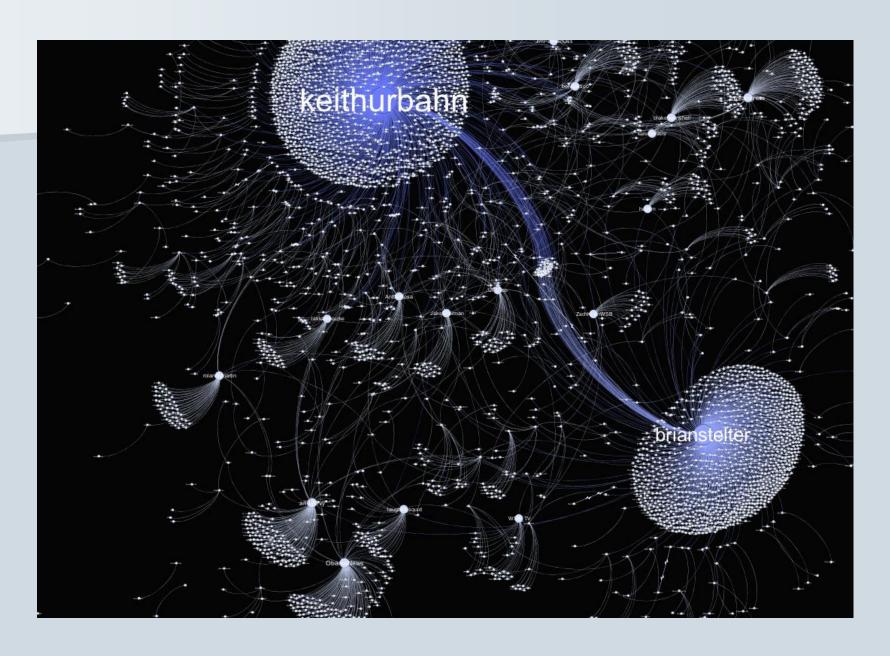
# THE STEM CELL WARS

When a Nobel prize is up for grabs, do scientists across the globe compete on a level playing field? Peter Aldhous investigates

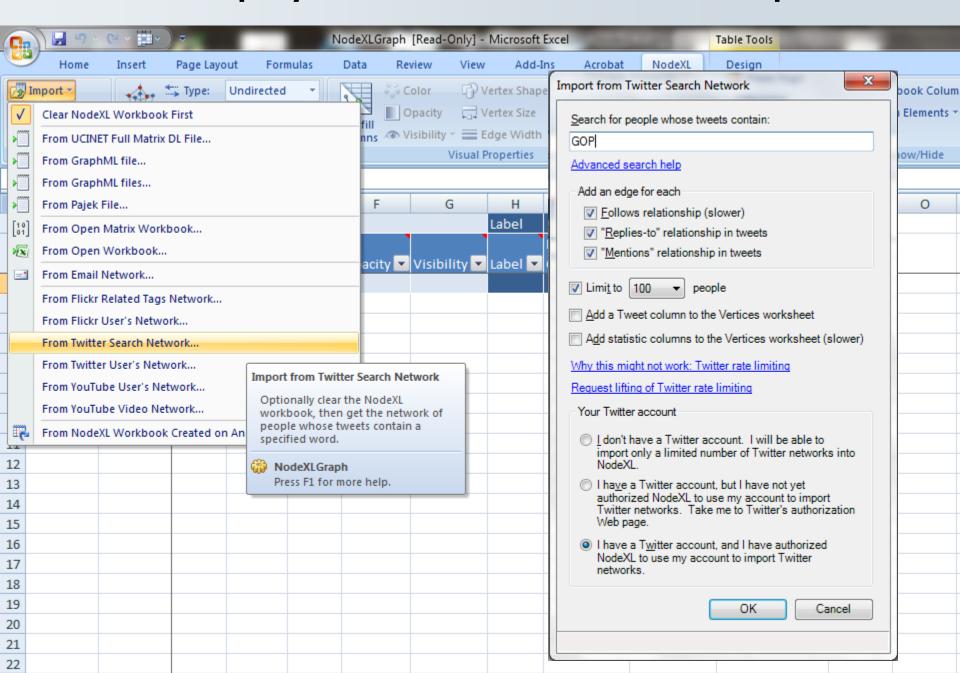
The most influential players in callular reprogramming are revealed by recording flow many times the oberdats have referred to according flow many times the oberdats have referred to according flow many as in pages; in leading journals (for many)s, see "The storograph ins', below right).



# **Network analysis of online social media**



## NodeXL can query APIs of online social media and import data



# **Example: Occupy vs Tea Party on Twitter**

# Occupy vs Tea Party: what their Twitter networks reveal

15:21 17 November 2011



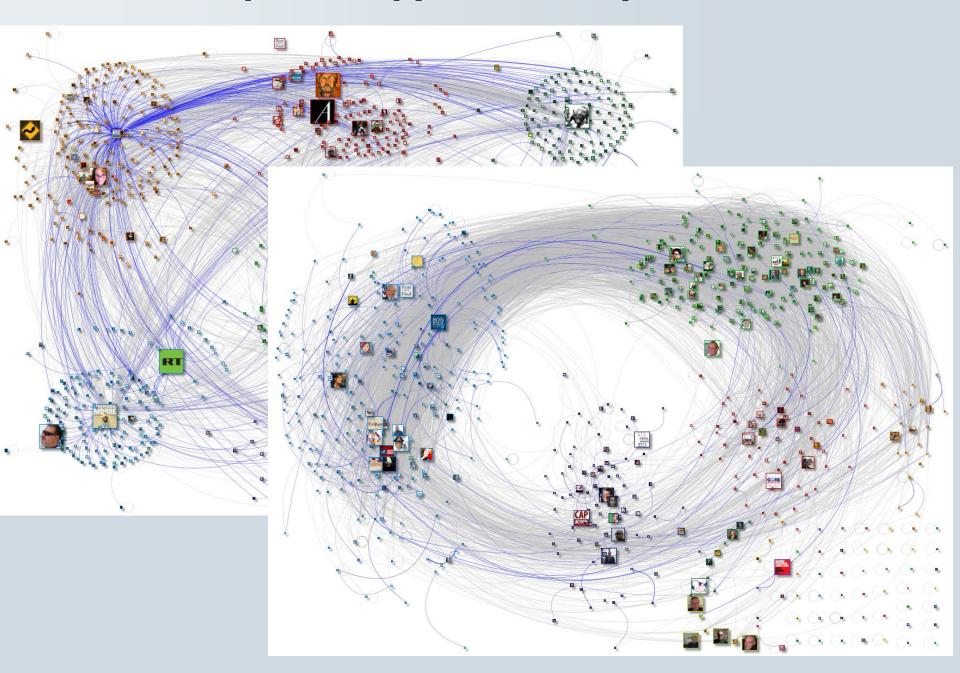
Peter Aldhous, San Francisco bureau chief

According to <u>some political commentators</u>, Occupy Wall Street is the left's answer to the Tea Party - driven by a similar anger towards elites. But the social networks of people tweeting about the two movements suggest that they have rather different dynamics.

Those tweeting about the Tea Party emerge as a tight-knit "in crowd", following one another's tweets. By contrast, the network of people tweeting about Occupy consists of a looser series of clusters, in which the output of a few key people is being vigorously retweeted.

The Occupy network also has many casual unconnected tweeters, shown to the bottom right of the diagram below. Whether Occupy takes off as a coherent movement may depend on its success in bringing these potential recruits into the fold.

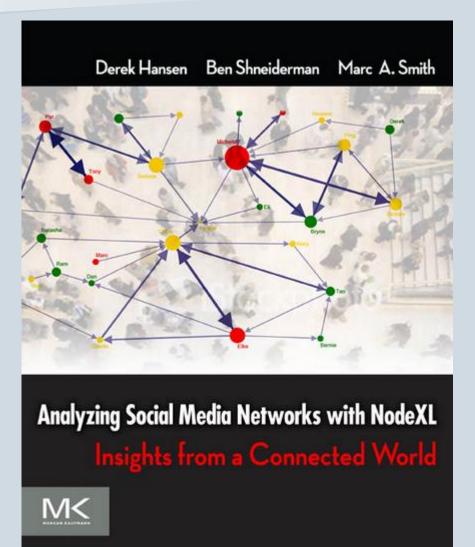
# **Example: Occupy vs Tea Party on Twitter**



## **Learn more**

Tutorial: Analyzing Social Media Networks: Learning By Doing With NodeXL

Book: Analyzing Social Media Networks with NodeXL: Insights From A Connected World



# **Network Analysis For News**

CAR meeting, St Louis, Feb 23 2012

Peter Aldhous



peter@peteraldhous.com

Peggy Heinkel-Wolfe

Denton Record-Chronicle

pheinkel-wolfe@dentonrc.com