



LINKING AUDIENCES TO NEWS

A NETWORK ANALYSIS OF CHICAGO WEBSITES

An **Advancing Chicago's**
Information Ecosystem Report

Funded by The Chicago Community Trust,
The John S. and James L. Knight Foundation and
The Richard H. Driehaus Foundation

Contents

Executive Summary.....	2
Background and Context.....	4
Introduction to Network Analysis.....	6
Research Objectives and Methodology.....	8
Network Analysis of the Chicago New News Ecosystem.....	11
Authorities and Hubs.....	22
Switchboards.....	26
Referrers and Resources.....	28
Using These Findings to Improve the Ecosystem.....	31
Opportunities for Future Research.....	32
Acknowledgements.....	34
Appendix A: Top Sites Ranked by Network Statistics.....	35
Appendix B: List of Seed Sites.....	43
Appendix C: About the Authors.....	52

Executive Summary

The mass media model, which sustained news and information in communities like Chicago for decades, is being replaced by a “new news ecosystem” consisting of hundreds of websites, podcasts, video streams and mobile applications. In 2009, The Chicago Community Trust set out to understand this ecosystem, assess its health and make investments in improving the flow of news and information in Chicagoland. The report you are reading is one of the products of the Trust’s local information initiative, Community News Matters.

“Linking Audiences to News: A Network Analysis of Chicago Websites” is one of the first – perhaps *the* first – research projects seeking to understand a local-news ecosystem by analyzing its **structure**: the way locally focused websites connect with one another through hyperlinks. With the help of webcrawler software, we have mapped the connections among local websites and used network analysis to identify the sites that are most important to the health of the ecosystem.

Our key findings:

- There are more than 400 websites that provide news and information relevant to Chicagoland residents.
- Of the sites identified in this study, almost eight in 10 received few if any links FROM other sites in the network. This means that no matter how good their content, these sites are unlikely to be found by users.
- More than four in 10 sites do not link TO other sites in the network. This means they are not helping their audiences find relevant content published by other Chicago area Web publishers.
- The sites that do give and receive links are clustered into content-oriented communities – for instance, sports, arts and restaurants – that link heavily among themselves but don’t get links from sites focusing on other content domains.
- Some sites are more important to the ecosystem than others. They link out more to other sites, are more likely to receive links, or both.
- Sites that receive a lot of links include major sources of original news reporting such as chicagotribune.com as well as online-only publications such as gapersblock.com and chicagoist.com.

- Organizations and institutions – which in the mass media era relied heavily on newspapers and TV to get their messages to local residents – increasingly are able to serve their audiences through their own websites. These institutions, such as the region’s mass transit systems and local museums, operate some of the most widely linked-to websites in the region.
- Websites operated by traditional media tend not to link out to other sites. The sites most likely to link out are organizations and institutions, as well as online-only publications, such as gapersblock.com.

Based on our research, we suggest that people and organizations seeking a better informed community should find ways to encourage and incentivize more linking among locally focused websites. We also offer some ideas for repeating and deepening this type of research on a regular basis.

Background and Context

During the mass media era, local newspapers and network broadcast affiliates held dominant positions in the media landscape because of significant barriers to entry (the cost of printing presses and scarcity of broadcast licenses). They also employed the largest staffs of journalists, paid for by their dominant shares of local advertising. Major media outlets paid journalists to do original reporting, and it was possible for their work – delivered to large, loyal audiences – to get noticed and have impact.

The modern news ecosystem – hundreds of small websites, not just a handful of major media companies – is fundamentally different. Some differences have a negative effect on the creation and distribution of local news and information. For instance, as research for The Chicago Community Trust has found, few of the news publishers in the Chicago area can afford to hire professional journalists. On the other hand, there is potential strength in the openness of digital platforms and the lowering of barriers to entry. Suddenly, anyone can publish news of any topic. The resulting explosion in digital news startups, in Chicago and nationwide, suggests a possible future in which publishers who deeply understand the needs of their niche audiences break news that would never have been discovered by mass-media journalists covering only stories that interest the largest number of people.

But even if these new publishers successfully generate news coverage of value to their communities, they cannot replicate the dominant distribution platforms that newspapers and television affiliates once had. If their original reporting is going to get noticed by large numbers of people – a requirement needed to spur civic action, hold governments accountable and generate momentum for change – our community will need to build alternate mechanisms for content visibility and distribution. On the Web, the key distribution mechanisms are rooted in networks: hyperlinks, search engine algorithms and social sharing via services such as “e-mail to a friend,” social networks and blogs. To put it simply, a great story without a built-in, mass audience risks becoming the equivalent of the proverbial tree falling in the forest. No one will hear about the story unless it is linked to or passed along via networks.

The Chicago Community Trust established the Community News Matters program in 2009 to improve the “new news ecosystem” in Chicago. The ecosystem metaphor is quite appropriate. To understand a natural ecosystem requires more than understanding the individual plants and animals that live in it. It’s also necessary to understand the relationships among those plants and animals: how they are connected, which ones depend on each other, which ones prey on each other. Similarly, to understand a news ecosystem on the Web, it’s not enough to look only at individual sites. It’s critical to understand how the sites relate to one another – which sites link to one another and which groups of sites are most tightly connected.

The research project described in this report was designed to help the Trust – and others concerned about the quality and quantity of news and information – understand the structure of the hyperlink network of news/information websites in the region. The foundation of this study was to use the tools of network analysis to quantify the degree to which these sites link among themselves and to traditional media. While researchers have previously examined hyperlink networks (for instance, focusing on the links among political and conservative political blogs) to the best of our knowledge this is the first research project focusing on links among sites in a locally defined market.

Websites or groups of sites that are highly interconnected by links can be said to form a community of sites – and to serve a community of users. Through links, users of these sites can become aware of news and information that matters to them. But in a world where anyone can publish, there is always the risk that these hyperlink communities will become “echo chambers,” serving people with only the most obvious information, reinforcing their assumptions and stereotypes, or failing to surface relevant information generated by unfamiliar sites. Our research sought to identify sites that connect Web communities to one another through links – thus increasing the formation of what sociologists call “social capital.” In his book *Bowling Alone: The Collapse and Revival of American Community*, Robert Putnam writes:

By analogy with notions of physical capital and human capital – tools and training that enhance individual productivity – the core idea of social capital theory is that social networks have value. Just as a screwdriver (physical capital) or a college education (human capital) can increase productivity (both individual and collective), so too social contacts affect the productivity of individuals and groups. Whereas physical capital refers to physical objects and human capital refers to properties of individuals, social capital refers to connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them.

Putnam and others have drawn a distinction between “bonding” social capital, which knits people of like interests together, and “bridging” social capital, which connects people who have diverse backgrounds, interests and social networks. In assessing the “information health” of a community, both kinds of social capital are important. Evaluating the topology (structure) of hyperlinks in a network of locally focused websites can point out sites that play important “bonding” and “bridging” roles – and, perhaps, help identify steps that can help build a healthier news/information ecosystem. It may also be useful in efforts to build financial support for the “new news ecosystem” – by yielding insights into which sites are most closely connected and, therefore, might appeal to the same advertisers or donors.

The Community News Matters program is dedicated to finding ways to help the “new news ecosystem” of digital publications become more successful – whether through better content, better audience development or better business practices. This report describes research that can prove important in helping Chicago area digital news sites build their audiences, impact and – potentially – revenue.

Introduction to Network Analysis

On a fundamental level, the world is full of “things” – such as people, organizations, communities and societies – that have relationships with one another. For example, people have friends, family and colleagues, while organizations have affiliates, competitors and stakeholders. The relationships between these “things” are often complex; their structure creates a web or network. Although people have many one-on-one interactions, in most cases, people have relationships with multiple people. These people then, in turn, have relationships among themselves. For example, I have a relationship with my mom and dad while the two of them also have a relationship; hence the three of us form a basic network, as shown in Fig. 1.

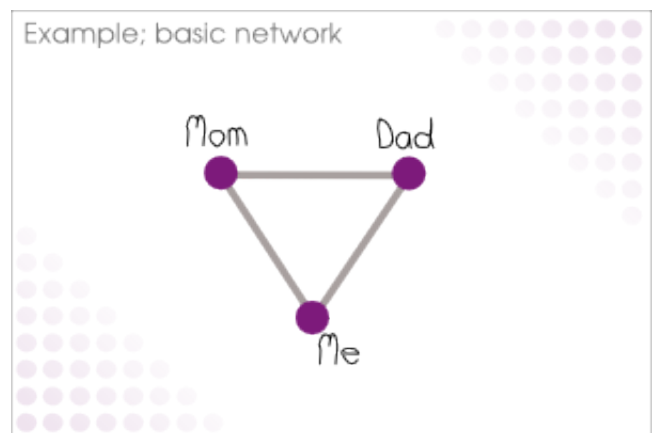


Fig. 1: A three-node network

As the number of nodes increases, so does the complexity of the network. Thankfully, these networks can be analyzed. We can assess who talks to whom, who is the central figure in a network, how close the “things” are that are represented in the network, and what the nature of the relationships in the network is. The art of calculating and analyzing network data is called Social Network Analysis (SNA). SNA can be used for many purposes across various fields. In communications science, SNA could, for example, be used to build and test theories of networked relationships such as “the Two Step Flow of Communication” (Katz & Lazarsfeld), “the Strength of Weak Ties” (Granovetter) and “Diffusion of Innovations” (Rogers). In practice, SNA can be used for a myriad of organizational purposes. For example, SNA allows us to identify the “opinion leaders” and “change agents” within companies, to analyze real rather than nominal organizational structures and to assess the cultural fit of two merging companies.

What Are Networks?

Networks are everywhere and can represent a variety of relationships. A few examples:

1. Friendship network: people are connected by being friends with each other;
2. E-mail communication network: e-mail addresses are connected by the act of sending e-mails to each other;
3. Document revision network: authors and editors are connected by virtue of contributing to a document.

Networks, although seemingly simple, can hold a wealth of information. While network science has existed for decades, only recently has a wealth of digital data, combined with modern analysis techniques, allowed us to formally study networks and network data in depth.

Network data comes from a variety of sources. Primary data can be collected directly from any sort of database. This includes, but is not limited to, company e-mail and instant messaging logs, phone logs and meeting calendars. Data generated by a webcrawler was used for the Chicago network study. A webcrawler is a type of software that finds and maps a group of websites that are connected through hyperlinks. The results help us map what websites are best positioned to drive traffic and what sites connect most closely to one another.

How Networks Are Analyzed

To study networks, we represent them as a set of nodes connected to one another by a set of edges. Nodes can be people, organizations, websites, or resources. Edges (which also might be called ties, links or connections) are the relationships that connect the nodes in a network.

For example, consider Fig. 2, a basic network consisting of six people (the nodes in the network). Some of these nodes are connected (through edges), but others are not. Node 1 is connected to three other nodes directly (nodes 2, 5 & 6), but node 3 is only connected to one other node (node 2). Indirectly all of the nodes are connected to one another. It takes node 1 just one step to reach nodes 2, 5 and 6 and two steps to reach the remaining two nodes (3 and 4). All of the nodes are connected to one another via no more than two steps.

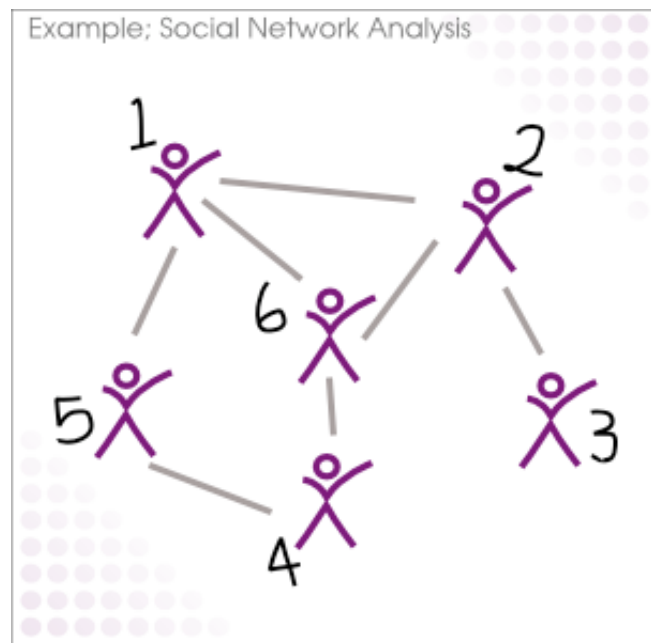


Fig. 2: A basic six-node network

From this simple setup, analysts can use a set of network metrics to quantify and compare how certain nodes and relationships contribute to the flow of information or resources through a network.

Research Objectives and Methodology

Our research project set out to:

- **Discover** the network of links connecting the news websites in Chicago's news ecosystem;
- **Build** a virtual representation of this network;
- **Diagnose** the overall health of the network;
- **Reveal** patterns and trends in the hyperlinks;
- **Identify** key sites in the ecosystem – sites that play the most important roles in the ecosystem by connecting people with information through links.

To identify the link structure among websites, researchers use webcrawler software. Webcrawlers (also known as spiders) start with a list of “seed sites” and then follow links to other sites. By following a set of rules, a webcrawler captures and stores link information for further analysis. We used Issue Crawler (<http://www.issuecrawler.net/>) to crawl Chicago news and information sites in August 2010. We analyzed the link data using software packages UCINET (a tool widely used by network scholars) and C-IKNOW (a tool developed at the Science of Networks in Communities laboratory at Northwestern University).

Seed Sites

To expose a hyperlink network, a webcrawler begins with a starting set of sites. It then crawls these seeds, following their hyperlinks outward to other sites. Seed sites are the foundation of the crawl and must be chosen carefully. A set of seeds that is too narrow can lead to an incomplete final network – if your set of seeds consisted mainly of Chicago sports news sites, the crawled hyperlinks would mainly be sports-related, and the final network would likely be missing other site genres. On the other hand, a set of seeds that is too broad can lead to a network that is too large; if the seeds contained sites geared to Chicago tourism, which rarely provide news about Chicago, the crawl would pick up mostly travel sites irrelevant to the Chicago new news network. Thus, seeds must be comprehensive so their links lead to as much of the network as possible while remaining relevant so the network remains issue-specific.

Our seed sites came from two sources:

- 118 sites that responded to the survey for CCT's 2010 report, *The New News 2010: Mapping Chicago's Online News Scene*.

- 240 Chicago websites found through fwix.com, a website that compiles news feeds by location – <http://fwix.com/Chicago/browse/sources>.

The surveyed sites were obvious starting points of interest. The fwix.com list was added because it was extensive and included both traditional news media sites and new news sites spanning a variety of topics. (See Appendix A for seed list.)

Issue Crawler

Issue Crawler is a network mapping software made available by the Govcom.org Foundation, Amsterdam. In technical terms, we conducted a co-link analysis by page, using three iterations of a crawl depth of three. The process worked like this:

- The list of 358 seed URLs was entered into Issue Crawler. Each seed URL led to one page of a website, usually the homepage (Depth One).
- Issue Crawler accessed these seed pages. Calling these the “front-pages,” Issue Crawler searched each page for links to other pages within the site. Accessing these front-page links gave a new set of pages located one level into the site (Depth Two).
- Issue Crawler then searched these one-level deep pages for links to deeper pages within the site. Following these links, the new set of pages found was two levels into the site (Depth Three).
- Issue Crawler then examined the external links, or hyperlinks, from these three sets of pages – the front pages, the pages one level in and the pages two levels in – to find a new set of sites.
- Sites which received external links from two or more sites formed the new network to be crawled. This completed the first iteration of the crawl.
- The entire process was repeated two more times, using the sites found after each iteration as the new set of seed sites.

Each subsequent iteration of the co-link analysis produced an even more specific set of sites: sites which had received links from sites, which in turn had received links in the previous crawl. The final network consisted of 439 sites – our best approximation of the core of the Chicago news ecosystem.

Site Categorizations

To aid in our analysis, we coded sites by “category” and “scope.” All sites were grouped into a category, while “scope” was attached only to sites providing news and/or information geared to local audiences. The categories were:

- **Legacy:** A Web publication related to or corresponding with a mainstream or traditional media brand. The key, defining characteristic is that the site carries the brand and/or content of this traditional media brand. (Examples: chicagotribune.com, cbs2chicago.com.)
- **Legacy-affiliated:** A Web publication owned by a mainstream or traditional media brand, but carrying a separate name. (Examples: chicagonow.com, vocalo.org.)
- **Micro-publisher:** A Web-only or Web-first publication focused on a topic, geographic area or audience segment. (Examples: gapersblock.com, evanstonnow.com, chicagoartmagazine.com.)
- **Organization/institution:** A website providing news or information from an organization, company, institution, government agency or nonprofit that historically would have relied on media intermediaries to get its information to relevant audiences. (Examples: fieldmuseum.org, cityofchicago.org.)
- **National brand:** A website with national scope that covers local news and information, usually with a structure of geographically segmented subsites. (Examples: citysearch.com, dailycandy.com, rivals.com, examiner.com, huffingtonpost.com, outside.in, sbnation.com.)
- **Web service:** A website that provides services to Web publishers. (Examples: addthis.com, wordpress.com, facebook.com, flickr.com, quantcast.com, twitter.com, vimeo.com.)

The “scope” categorization was intended – for sites providing locally focused news and information – to describe the type of information provided. We assigned a scope only to sites previously categorized as Legacy, Legacy-affiliated, Micro-publisher and, in some cases, Organization/Institution. For the Organization/Institution sites, we coded a scope only in cases where the site provided news or information about one or a few geographic areas within the Chicago region. The options for scope were:

- **Geo-publisher:** A Web publication focused on one or a few specific geographic areas (neighborhoods or municipalities) within the Chicago region. (Examples: dailyherald.com, evanstonnow.com, adentrodepilsen.org.)
- **Niche publisher:** A Web publication focusing on a topic or audience segment. (Examples: chicagobusiness.com, theexpiredmeter.com, bleedcubbieblue.com, newcity.com, catalyst-chicago.org.)
- **Mass media:** A website branded with a major mass media outlet. (Examples: chicagotribune.com, nbcchicago.com, suntimes.com.)

Network Analysis of the Chicago New News Ecosystem

Network scientists have a host of methods at their disposal to analyze a network; network diagrams and network analysis metrics are the two main methods.

Network diagrams are visual representations of how a network is arranged structurally. We will use different types of network diagrams to show how sites in the hyperlink network are interrelated, as well as to point out key insights that we have found.

Network analysts also use a set of network metrics unique to their field; these allow us to quantify and compare how certain nodes and relationships are vital to the flow of information or resources. For example, **centrality** is described by a series of measures that quantify the extent to which the rest of a node is (or isn't) at the center of its network, while **closeness** is the degree to which a node can reach other nodes in its network. Both are useful for measuring the ease by which information passes through a network.

In the sections below, we employ network diagrams and network metrics to display and explain our key findings from our research.

General Network Overview

One of the first things that network researchers do when examining a network is to look at it from a global perspective. By looking at the network as a whole, it is possible to notice some of the larger trends that may be indicative of the health and operation of the network.

On the next page, Fig. 3 depicts the Chicago network as a multi-colored social network. Each site is depicted as a little colored bubble with the site's name written inside. The lines that connect the sites represent the hyperlinks that connect each site on the web. During our crawl, we identified two different types of relationships that can exist between sites:

Binary: A binary relationship exists between two sites if there are any links that connect them. It signifies the presence of some type of relationship.

Continuous: A continuous relationship is a relationship that is weighted by the number of links present between two sites. This is useful because many sites have multiple links connecting them, which suggests stronger relationships.

In order to conduct our analysis, we used both continuous and binary relationships to help us learn different things about the network. For each measurement and graph we present, we will let you know which type of relationship we used and why.

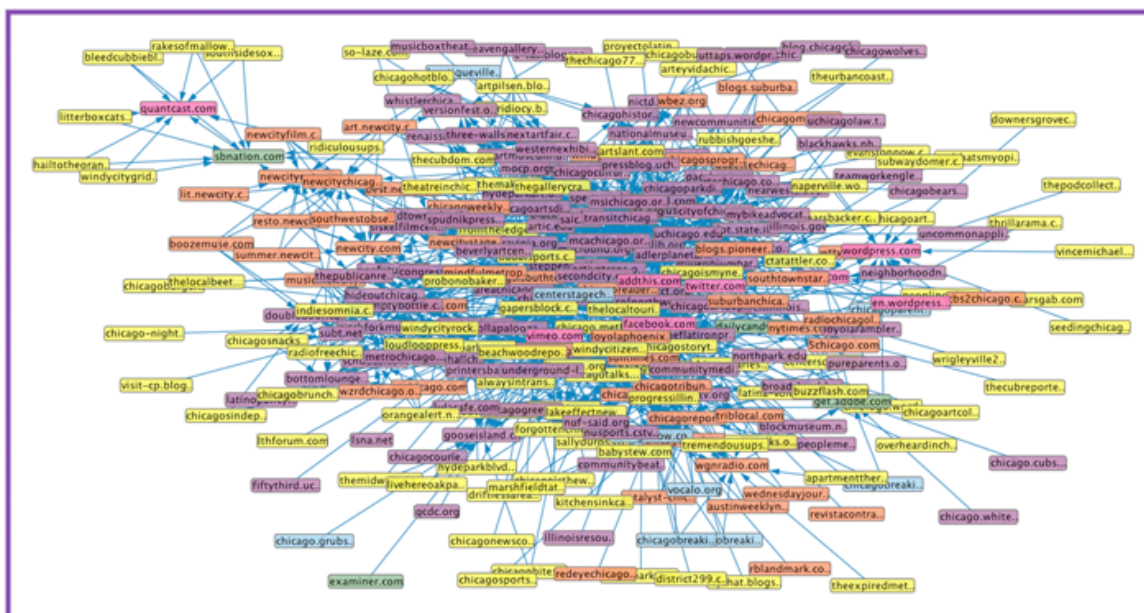


Fig. 3: The network of all the sites in the new news ecosystem, clustered based on degree centrality. Isolates are excluded.

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- ➡ Binary Edges

Network overview	
Number of sites	439
Number of isolates	162
Number of non-isolates	277
Density (without isolates)	0.0161
Density (with isolates)	0.0064
Total links	24,598
Total site relationships	1,232

At first glance, large network diagrams like Fig. 3 look like giant “hairballs” and cannot easily be deciphered. We use a combination of social network analytics and visual interpretation to help cut through the confusion.

We start first with the most basic information about the network, which is summarized in the table above. There are 439 sites in the network, of which 162 are “**isolates**,” or sites that are not connected to any other site in the network. Excluding isolates, we are left with one large network component of 277 nodes.

Looking more closely at the linking patterns, we find:

- 79 percent of the sites (345 of 439) receive no links FROM other sites in the network; and
- 44 percent of the sites (192 of 439) do not link TO any other sites in the network.

Because the crawler initially dropped sites with just one link in each iteration of the crawl, these numbers may slightly overstate the number of isolates and sites with zero inlinks or outlinks. But it is noteworthy that there are so many sites in the network that are fundamentally disconnected from the others. The data suggest that an opportunity exists to expose content to larger audiences through greater interlinking among sites in the ecosystem.

It is apparent from the data that sites have very different philosophies about linking out to other sites. For example, consider [chicagotribune.com](#), the Tribune Company's newspaper website, and [chicagonow.com](#), the same company's site aggregating contributions from local bloggers. [Chicagotribune.com](#) has zero links to other sites, while [chicagonow.com](#) has 63 links to eight other sites in the network.

It is of interest that all the non-isolate nodes form one giant component, or cluster, in which every site is directly or indirectly connected to every other site. To better understand this connectivity, we compute a metric called **density**.

Network Density

Density can be defined as the ratio of the number of actual links in the network, as opposed to the number of potential links.

The more links in the network, the higher the density becomes. The density has a number ranging from 0 to 1, where a 1 indicates that all of the potential links are present. In Fig. 4, the network on the left has a relatively

low density. There are 15 possible links in the network, five of which are present. Hence, the density is .33. In the network to the right, the density is much higher. Fourteen out of the possible 15 links are present, giving this network a density of .93.

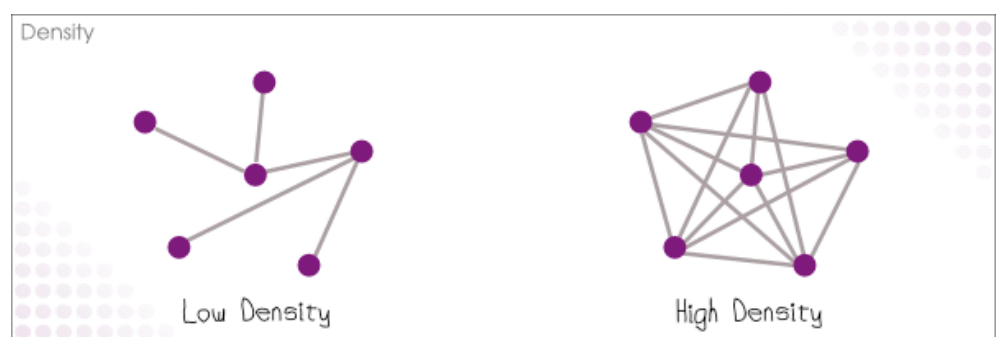


Fig. 4: Examples of low- and high-density networks

Density is an important measurement for exploring the network's health. In a nutshell, a network's density shows how many ties exist based on the range of possible ties.

The density of the network, not including isolates, is 0.0161, meaning that 1.61 percent of all possible site-to-site links exist. This is a low density, but this is not necessarily surprising for a large hyperlink network of sites that may have little in common but a Chicago orientation. Remember that this calculation excludes the 162 isolates and that the webcrawling rules excluded sites that failed to receive at least two links during each phase of the crawl. The density for the network including the isolates is .0064, meaning that less than 1 percent of the possible site-to-site links exist.

Clustering and Content Divisions

When we view the network at a macro level, it is possible to see some of the separate content-based clusters of websites that exist in Chicago – groups of sites that are more tightly connected by links. As one could imagine, it may make sense that sports sites link to sports sites, while arts sites tend to link to arts sites. By visualizing the network in different ways, it is possible to visualize some of these clusters of sites.

To identify the clusters, we view the network with continuous relationships (the number of links connecting the sites) as our edges. This allows us to impose a threshold so that a connection is displayed visually only if the number of links in each edge is above a certain amount. We set the threshold to 45 links. So, in Fig. 5 (on the next page), lines are shown connecting two sites only if there were at least that 45 links connecting them.

Immediately, we can see communities of sites that are differentiated by their content or affiliation. For example, the music sites link heavily to one another, as do a group of sites operated by the Tribune Company, including legacy.com, which provides obituaries and death notices for chicagotribune.com. Sites operated by NewCity Communications – a newspaper and web publications that focus on arts and entertainment – are tightly connected. What is interesting is that there is a core cluster of micro-publisher sites that tie together some of the otherwise disconnected components of this network. These sites link heavily to one another. Also interesting is that this cluster is completely separated from the Tribune Company sites. This illustrates that links between micro-publishers and the Tribune Company sites – in either direction – are relatively uncommon.

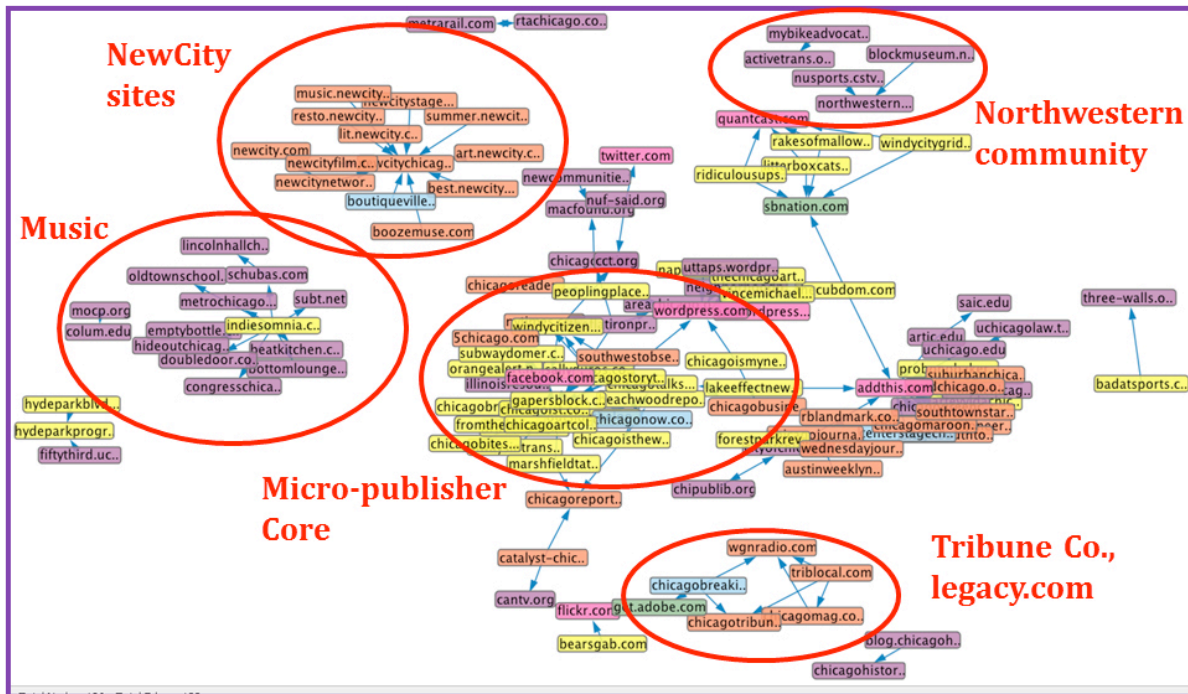


Fig. 5: Network with weighted edges; threshold for link occurrence is a 45-link minimum



We can also look at different types of sites in the network – for instance, the niche publishers. These sites are shown in Fig. 6.

The Value of the Crawl

As described in the methodology section, sites were identified for the crawl from three different locations: the CCT survey, fwix.com, and the crawl itself. In order to see how sites from each source are arranged in the network, we visualized the network with the nodes colored by site source. This can be seen in Fig. 7. From the diagram, we can clearly see a significant number of crawl-generated (blue) sites that are central to the network. This shows us that the crawl was effective in helping us identify key sites that would otherwise have been left out from the network.

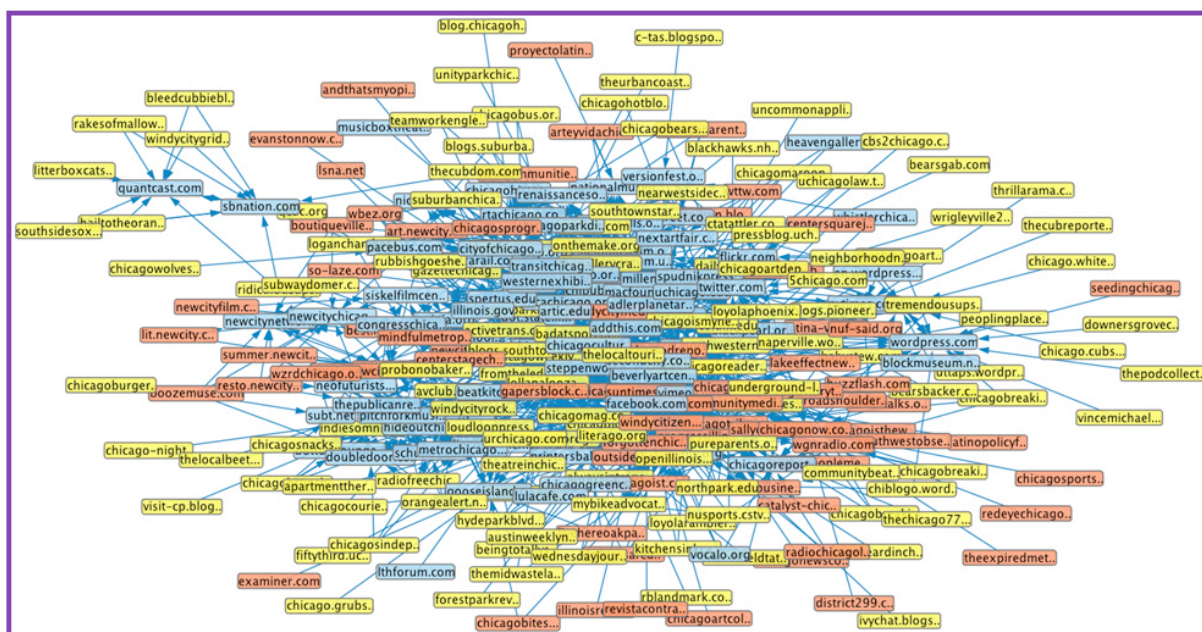


Fig. 7: Network with sites colored by origin (CCT survey, fwix.com, or crawl-identified). This shows how fwix.com and the crawl supplemented the CCT sites in key areas of the network.

Colored by site source

- CCT
- Crawl
- Fwix.com

Source	Number of sites
CCT survey	118
Fwix.com	240
Crawl-identified	81
Total	439

Category Links

As described in the methodology section, we coded the sites by category. When we observe the network globally, we can aggregate the links by category to see how they link to one another:

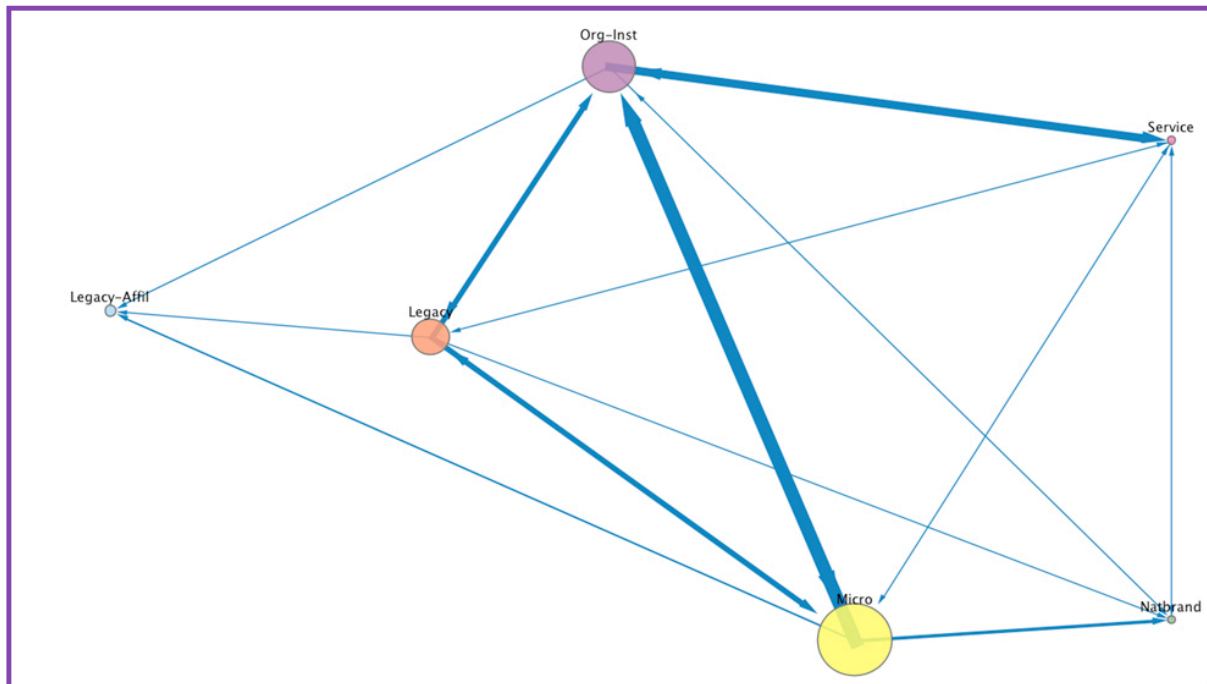


Fig. 8: Network of categories to categories, with nodes sized by number of sites in each category and edges sized by the number of binary links between categories

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- Binary Edges

Category	Number of Sites
Micro-Publishers	206
Organizations/Institutions	127
Legacy	77
Legacy Affiliated	12
Service	9
National Brand	8
Total	439

In Fig. 8, you can see the network aggregated into six “meta-nodes,” each representing a category of sites in the network. Each “meta-node” is sized based on the number of sites in that category; the width of the lines connecting them is sized by the number of binary connections (links between sites in each category). In this view, it is easy to see that the most relationships (299) exist between micro-publishers and organization/institution sites. The next highest frequency of relationships is between organization/institution and service sites, with 67 relationships.

A Note on Service Sites

In our analysis, service sites – especially Facebook, Twitter, AddThis and Vimeo – show up as relatively important in the Chicago network. This is understandable, but also a bit misleading. Facebook and Twitter show up because Web publishers frequently link to their own Facebook pages and Twitter feeds in order to encourage users to follow them for updates. While Facebook and Twitter are also important sources of inbound links – and traffic – web crawlers (including ours) do not typically capture most of them. There are several reasons for this, including the fact that most outbound links from Facebook are visible only to any individual user’s friends, and the fact that many links shared on social networks are concealed by URL shorteners such as bit.ly.

The situation with AddThis and Vimeo is a little different. AddThis offers an embeddable button that sites can easily install to offer users the ability to share content via social networks and other sites. The process of installing the AddThis button generates a link to addthis.com everywhere the button is published. Similarly, Vimeo allows site operators to embed video – and again, each embedded video shows up as a link to vimeo.com.

Because of these considerations, we have chosen to exclude services from the lists of top sites that appear in the remainder of this report. You may see them in network visualizations, however.

Network Roles

When analyzing a network, one of the key approaches is to determine the different “roles” that each node plays based on its position in the network structure. The most common example of a network role would be a “**connector**” in a network of individuals. In a network of people, the connectors are those with the most connections to others – the people who know everyone.

The same principle holds true for hyperlink networks. In hyperlink networks, the structure of links among sites in a community indicates the roles those sites play. The key metrics used for this type of analysis are different measures of **centrality**. As the term suggests, centrality refers to the extent to which a node is central to the network. Although the notion is fairly straightforward, there are quite a few forms of centrality that can be measured mathematically.

Three of the most useful centrality measures are:

- **Degree centrality;**
- **Betweenness centrality;**
- **Closeness centrality.**

We can use the “kite network,” first developed by David Krackhardt, a leading researcher in social networks, to show the distinctions between the first three measures.

Degree Centrality

Degree centrality simply counts the number of direct links with other nodes.

In Fig. 9, the node marked “Degree Centrality” has the highest degree centrality – the most direct links (six) to other nodes in the network.

In a hyperlink network, as well as many others, links are directional; site A can link to site B, even if site B does not link back to site A. In these directed networks, a node has both **in-degree centrality** and **out-degree centrality**, as indicated in Fig. 10 on the next page.

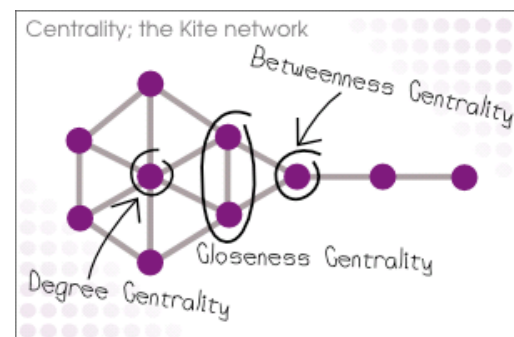


Fig. 9: Kite diagram

On the left side of Fig. 10, the node in the center has an in-degree centrality of five and an out-degree centrality of zero. On the right side, the measures are reversed: the node in the center has an in-degree centrality of zero and an out-degree centrality of five.

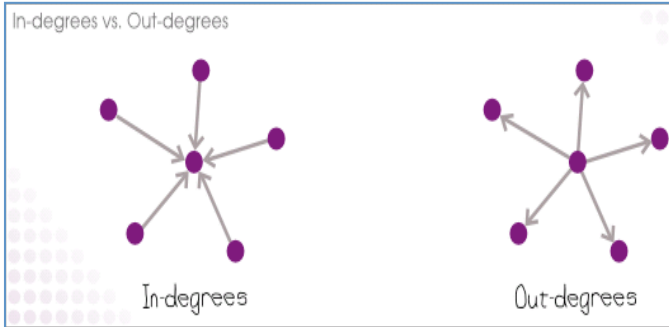


Fig. 10: In-degree vs. out-degree

Betweenness Centrality

Betweenness centrality refers to nodes that are central because they mediate between parts of the network.

In Fig. 11, the node marked “Betweenness Centrality” has just three direct connections – fewer than the average in the network – but has one of the best locations in the network. This node could be defined as playing a “broker” role in the network; it has potentially great influence over what flows through the network. It connects nodes that would otherwise not be connected. Without this “high betweenness” node, the two nodes on the right would not be connected to the larger group of nodes on the left. Nodes with high betweenness often connect otherwise disconnected communities.

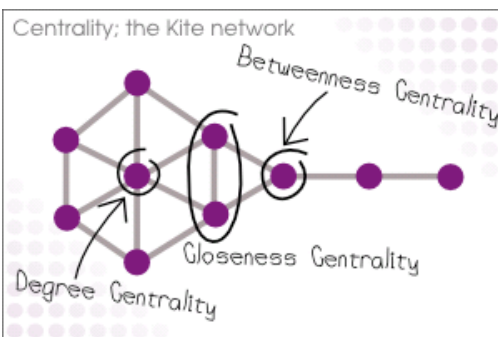


Fig. 11: Kite diagram

nodes can reach other nodes in the network in relatively few steps – two or fewer in most cases, and no more than three.

Closeness Centrality

Closeness centrality is the extent to which a node is close (in number of “hops”) to other nodes in the network.

In Fig. 11, the two nodes circled and marked “Closeness Centrality” have the shortest average paths to all other nodes in the network. These

Types of Sites

We calculated a variety of centrality measures for each website in the Chicago network. Based on these measures, we are able to identify sites that are:

- **Authorities** and **Hubs** (with high degree centrality);
- **Switchboards** (with high betweenness centrality);
- **Referrers** and **Resources** (with high closeness centrality).

Authorities and Hubs

The first two network roles are **Authorities** and **Hubs**.

Authorities are sites with high in-degree centrality: many other sites tend to link to them. Hubs are sites with high out-degree centrality: they link to many other sites. Fig. 12 illustrates the difference between Authorities and Hubs in a simple 'star' network.



Fig. 12: Authorities and hubs

In a hyperlink news network, Authorities are often first-source news sites; many other sites tend to link to them because they are a reputable or original source of knowledge.

Typically in a hyperlink network, an 80/20 principle applies, meaning that 20 percent of the sites receive 80 percent of the links. Many of these links come from Hubs, which tend to link out to many sites either because they cover a wide variety of content, provide a huge volume of content, or keep many active links on their front page.

Authorities

On the next page, Fig. 13 shows the network with circles representing each site scaled by the number of in-links. Thus, the largest circles are the most important Authorities.

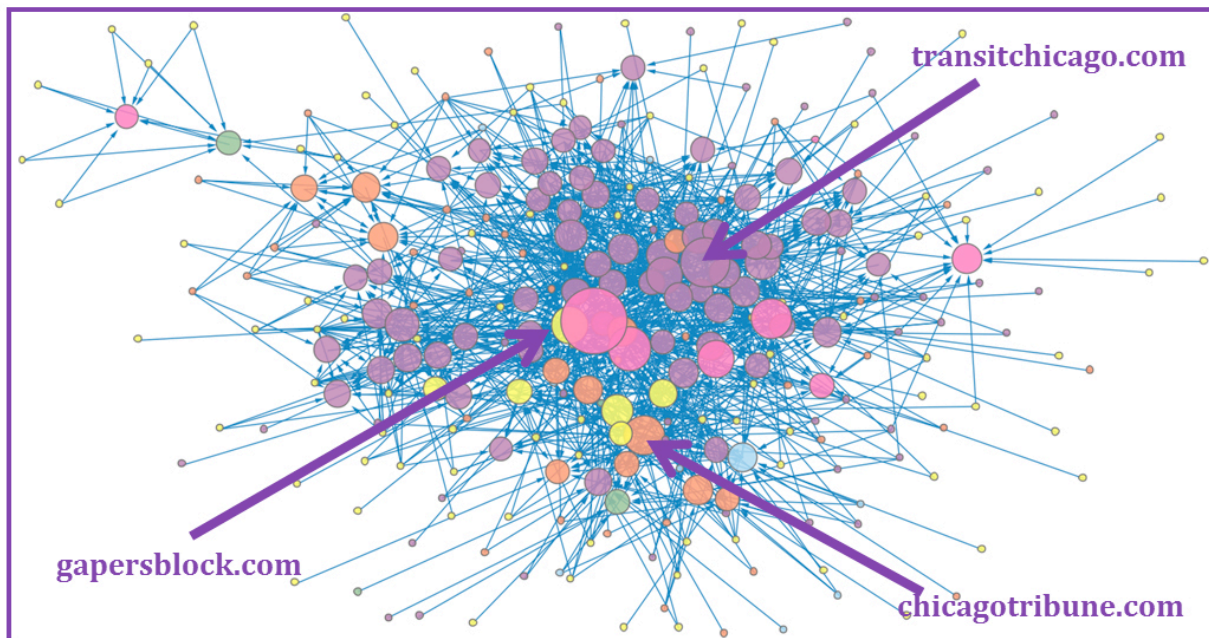


Fig. 13: Sites sized by in-degree centrality; larger nodes are more dominant authorities

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- Binary Edges

Top Authorities (excluding services)		
1	transitichicago.com	org/inst
2	chicagotribune.com	legacy
3	gapersblock.com	micro
4	mcachicago.org	org/inst
5	metrarail.com	org/inst
6	pitchforkmusicfestival.com	org/inst
7	chicagoist.com	micro
	msichicago.org	org/inst
9	hydeparkart.org	org/inst
	fieldmuseum.org	org/inst

Top Authorities include chicagotribune.com (a leading source of originally reported news in the region) as well as two online-only publications: gapersblock.com and chicagoist.com. The remainder of the Top 10 Authorities consists of seven organization/institution sites, including transitichicago.com (Chicago Transit Authority), three museums (Museum of Contemporary Art, the Museum of Science and Industry, and the Field Museum) and the site of the popular Pitchfork Music Festival. The prevalence of links to these sites demonstrates that organizations and institutions – which once relied on the mainstream media to communicate with the public – are now able to publish their own information online and have that information spread via links from other sites. (The large pink, unlabelled circle near the center of the graphic is facebook.com, which we excluded from our list of Top 10 Authorities for reasons explained earlier.)

Hubs

Hubs are sites that guide users to relevant information on other sites via linking. Fig. 14 is a visualization showing Hubs in the network:

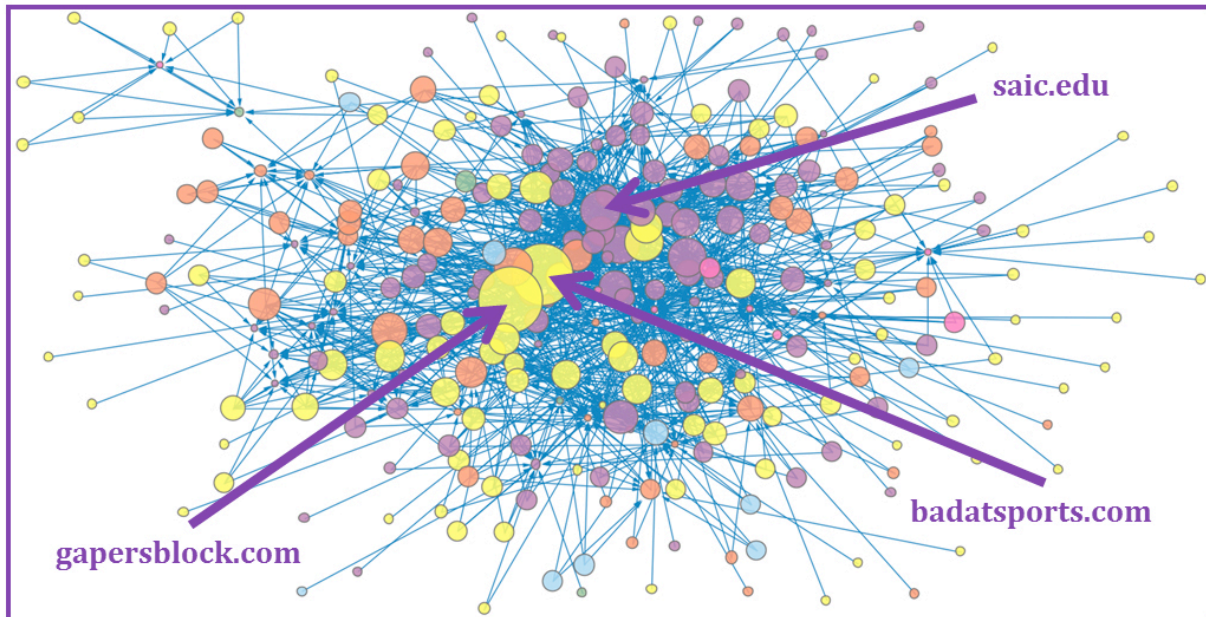


Fig. 14: Sites sized by out-degree centrality; larger nodes are more dominant hubs

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- ➡ Binary Edges

Top Hubs (excluding services)		
1	gapersblock.com	micro
2	badatsports.com	micro
3	saic.edu	org/inst
4	uchicago.edu	org/inst
	macfound.org	org/inst
	blogs.southtownstar.com	legacy
	thegallerycrawlandsomuchmore.blogspot.com	micro
8	avclub.com	legacy
9	communitymediaworkshop.org	org/inst
10	artslant.com	micro

Micro-publishers and organization/institution sites dominate the list of the top Hubs. It is noteworthy that Gapers Block, which was a top Authority (widely linked to), is also a top Hub (linking out frequently). This illustrates the important role Gapers Block plays in spreading content and connecting sites in the Chicago hyperlink network. Badatsports.com (a highly ranked Hub, not so highly ranked as an Authority) clearly plays an important role in providing links to arts-related sites and content, as does artslant.com. The organization/institution sites that show up as top Hubs are an interesting group: The School of the Art Institute, University of Chicago, John D. and Catherine T. MacArthur

Foundation, and the Community Media Workshop. These are sites that seem to be helping their users find relevant information by linking out to other sites in the Chicago ecosystem. The list of the Top 10 Hubs also shows that traditional-media sites tend not to link out very much. The only legacy sites that show up in the list of Top 10 Hubs are the Daily Southtown's blog site (blogs.southtownstar.com) and avclub.com, the Onion's entertainment section.

Switchboards

Switchboards are sites with high “betweenness,” a measure of the number of *shortest paths* that pass through a particular node. A path can be defined as the steps – or “hops” – needed to get from any node in the network to another node. A site with high betweenness lies on many of the shortest paths – those with the fewest “hops” between different pairs of sites. In practice, this means Switchboards connect nodes that would otherwise not be connected at all – or, if removed from the network, would require users to make many more clicks to travel from one site to another site. See Fig. 15 for an example of a Switchboard in a simple network.

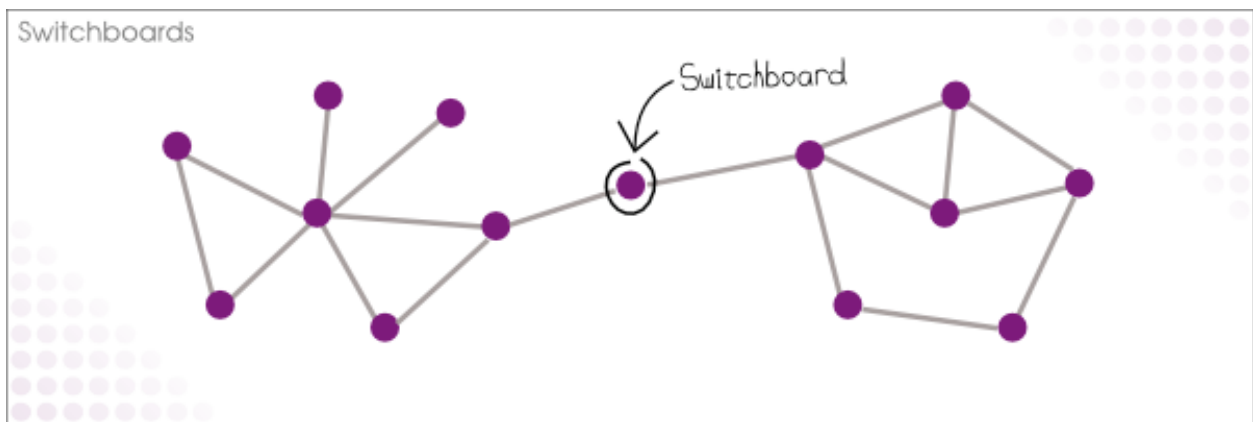


Fig. 15: Example of a Switchboard

In Fig. 16, we look at the Chicago website network based on betweenness. The size of the circles represents the betweenness score. One service (addthis.com) shows up prominently in pink near the center of the visualization. As explained earlier, we have excluded services from the list of the Top 10 Switchboards.

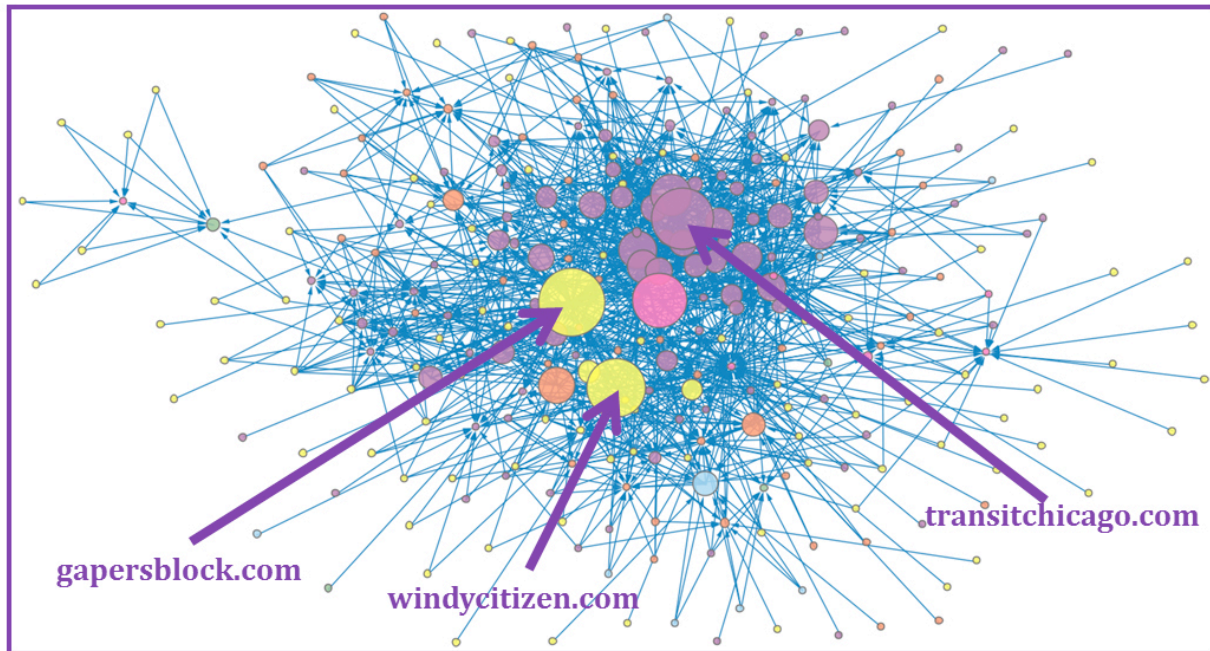


Fig. 16: Sites sized by betweenness centrality; larger nodes are more dominant switchboards

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- Binary Edges

Top Switchboards (excluding services)

1	gapersblock.com	micro
2	transitchicago.com	org/inst
3	windycitizen.com	micro
4	saic.edu	org/inst
5	chicagoartistsresource.org	org/inst
6	chicagomag.com	legacy
7	cityofchicago.org	org/inst
8	macfound.org	org/inst
9	uchicago.edu	org/inst
10	chipublib.org	org/inst

The top Switchboard sites include many of the same ones that showed up as Authorities and Hubs, including gapersblock.com, transitchicago.com, saic.edu and uchicago.edu. This is not really surprising. Hubs and Authorities are highly linked with other sites, so they are also likely to be positioned along many of the shortest site-to-site paths – the definition of a Switchboard. Two media sites rank high on the Switchboard list: windycitizen.com (a site whose users submit and vote on links to other Chicago news/information sites) and the site for Chicago magazine. Three sites affiliated with the Chicago city government – the municipal government (cityofchicago.org), public library system (chipublib.org) and the Chicago Department of Cultural Affairs (chicagoartistsresource.org) – also show up as strong Switchboards.

Referrers and Resources

Some nodes, by virtue of their structural position in the network, are ideally placed to quickly send or receive information in the network. Closeness centrality is the metric used to determine which sites fit these two roles.

The World Wide Web is a directed network, which means that one site can link to a second site without the necessity of a return link. This means there are two types of closeness centrality on the Web.

A site with high “**out-closeness**” has a position in the network that would enable a user following links to quickly reach other sites in the network. This means the site’s outbound links lead to sites with yet more outbound links to other sites. We call these sites **Referrers**.

A site with high “**in-closeness**” has a network position that would enable a user to reach it quickly by following links from other sites. This means the site’s inbound links come from other sites that are, in turn, linked from yet other sites in the network. We call these sites **Resources**.

In Fig. 18, the node with high in-closeness is a Resource while the node with high out-closeness is a Referrer.

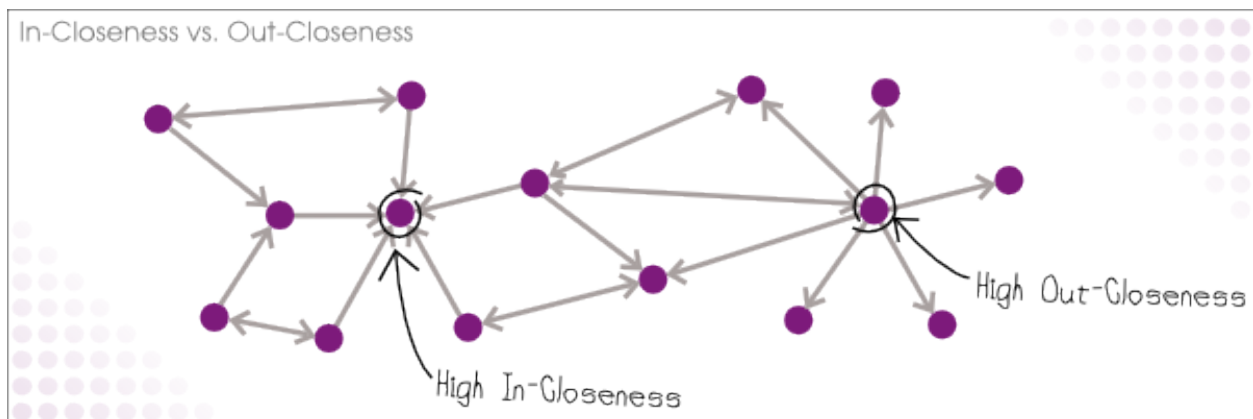


Fig. 17: In-closeness vs. out-closeness

Referrers

Fig. 18 shows the Chicago network based on out-closeness. The size of the circles represents the out-closeness score. The list of the Top 10 Referrers appears below the visualization.

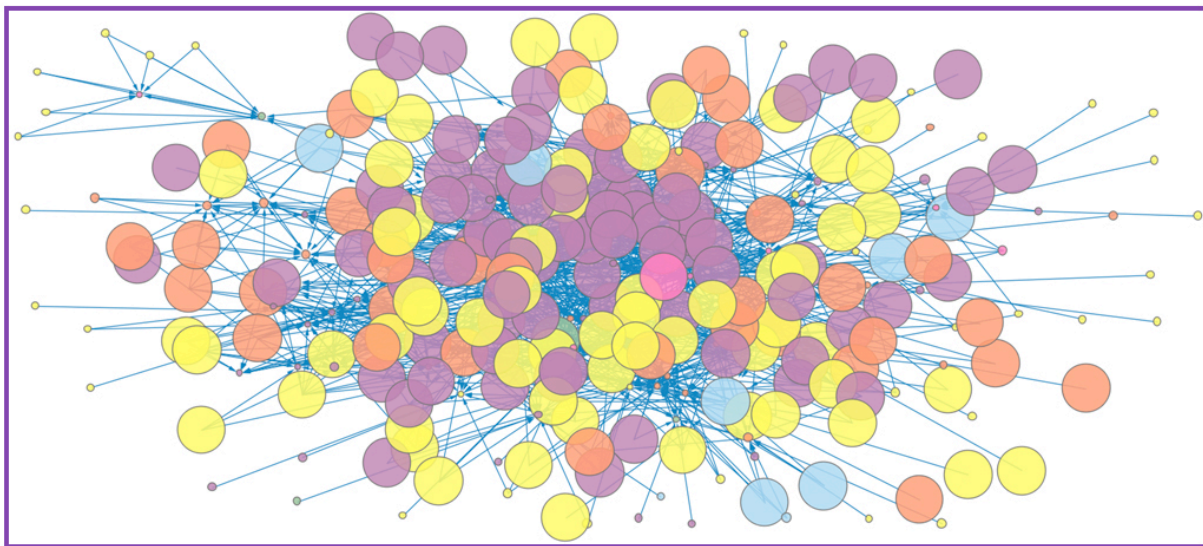


Fig. 18: Sites sized by out-closeness; note that there are many sites that are close to equal in size

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- ➡ Binary Edges

Top Referrers (excluding services)		
1	communitymediaworkshop.org	org/inst
	chicago.metblogs.com	micro
	chicagostorytelling.com	micro
	outsidetheloopradio.com	micro
	sallyduros.com	micro
	chicagocarless.com	micro
	austintalks.org	micro
	ctatattler.com	micro
9	123 sites are tied	

Among sites in this network, there is not a huge amount of variation in out-closeness. Since the network is one large component (every site can reach every other site from at least one path), and it is reasonably dense, it makes sense that most sites have similar ability to reach one another. There is, however, a clear list of the top eight Referrers, and they are predominantly micro-publisher sites, many of which have not appeared on other Top 10 lists: for instance, chicago.metblogs.com (local edition of a now-defunct group blog), chicagostorytelling.com (a collection of stories by DePaul journalism students), and outsidetheloopradio.com (site for a radio program airing on WLWU-FM).

One possible reason for this is that these micro-publisher sites lie on the periphery of many of the smaller micro-communities that make up the larger ecosystem. If you look at the network diagram in Fig. 18, you can see that the largest nodes often are outside the very center of the diagram. These nodes are much better positioned to reach both the periphery and the core of the network than many of the nodes in the core.

Resources

Fig. 19 shows the network based on in-closeness. The size of the circles represents the in-closeness score. These are the sites that are most easily reached by clicking from link to link in the Chicago news ecosystem.

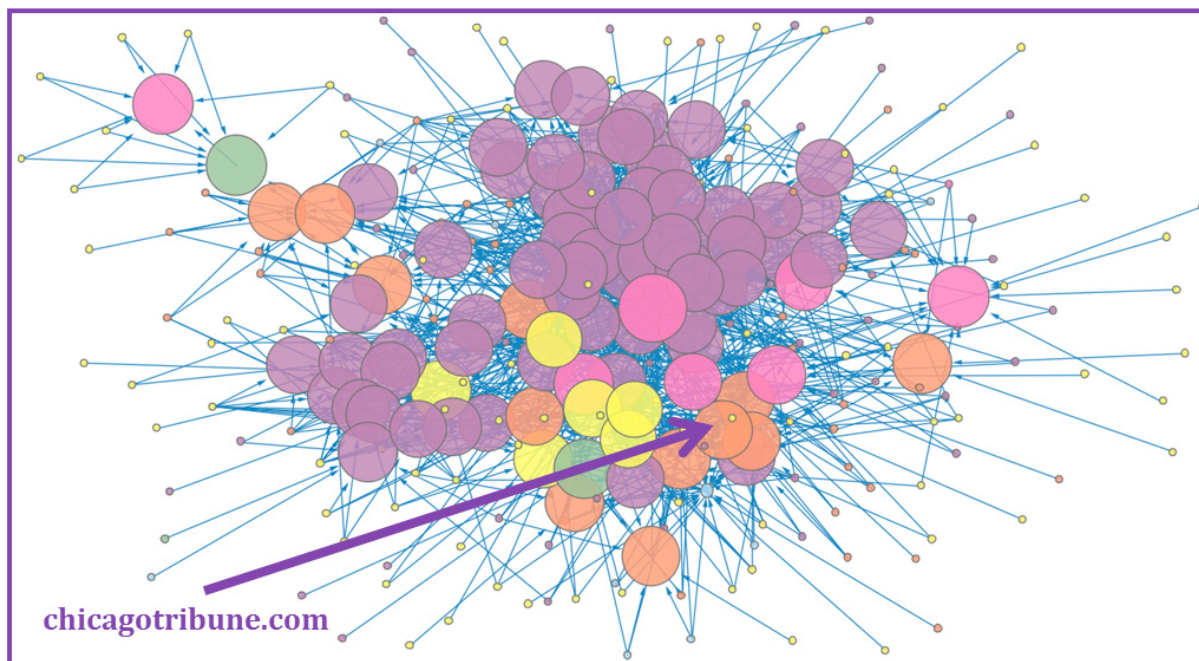


Fig. 19: Sites sized by in-closeness; note that there are many sites that are approximately equal in size

Colored by Category:

- Legacy
- Legacy Affiliated
- Micro-Publisher
- National Brand
- Organization/ Institution
- Web Services
- ➡ Binary Edges

Top Resources (excluding services)

1	chicagotribune.com	legacy
2	hydeparkprogress.blogspot.com	micro
	sbnation.com	natbrand
4	nytimes.com	legacy
	suntimes.com	legacy
6	chicagoreader.com	legacy
	gooseisland.com	org/inst
	newcity.com	legacy
	pitchforkmusicfestival.com	org/inst
	versionfest.org	org/inst

As with Referrers, there is not a huge difference among the top sites based on this network measure. The same reasons apply: every site can reach every other site from at least one path, and network density is reasonably high. The list of the Top 10 Resources includes quite a diverse mix of sites, including daily newspapers (chicagotribune.com and suntimes.com), weekly newspapers (chicagoreader.com and newcity.com), a neighborhood blog (Hyde Park Progress), a business (the Goose Island brewery) and two arts events (pitchforkmusicfestival.com and versionfest.org).

Using These Findings to Improve the Ecosystem

Our research suggests that links are important in a local news ecosystem and that increasing the amount of interlinking can strengthen it. Here are some ideas for ways organizations such as The Chicago Community Trust could accomplish this:

- **Encourage more linking through incentives.** The Trust has already used the preliminary findings in this study as input to its decisions on grants under the Community News Matters program. The more financial incentives are based on the volume of outbound links, the more publishers will want to link outward.
- **Promote link-sharing partnerships between traditional media and “new news” sites.** J-Lab, the Institute for Interactive Journalism, has created a “Networked Journalism Project” to nurture and support this kind of partnership in other markets.
- **Support sites that are both Hubs and Authorities.** Gapers Block is a good example of site that is both an Authority (getting links because of the quality of its content) and a Hub (linking outward to other sites frequently). These sites are critical to the flow of information among sites in the market, and it would be good to encourage other sites to follow Gapers Block’s lead.
- **Find ways to aggregate and distribute headlines across sites in the ecosystem.** For instance, a Chicago ecosystem widget with the most recent headlines from the region’s websites could be created and made available to all sites in Chicagoland. A headline widget could even include advertising space that could generate revenue to be shared among sites that use the widget.

Opportunities for Future Research

While our research has yielded significant insights into the network structure of Chicago's "new news ecosystem," it was also exploratory and experimental. There are many ways it could serve as a foundation for future research. Here are ways this research could be improved or extended – whether in Chicago or elsewhere:

(1) Conduct comparable research in multiple markets

Our research has clearly demonstrated the value of network analysis in understanding the structure of the ecosystem of news and information sites in a geographic region. Comparable studies in multiple markets would enable us to compare news ecosystems to one another and identify ways of improving these ecosystems.

(2) Create a systematic, replicable way to identify seed sites, especially those focusing on neighborhoods or municipalities

To do cross-market comparisons, it would be necessary to come up with a systematic approach to identifying seed sites in a community's news ecosystem. In Chicago, the list of seed sites generated from The Chicago Community Trust's survey and the fwix.com database gave us a strong foundation. But we know that the fwix.com database is incomplete (and the company has changed direction strategically, meaning that we cannot assume the database will be kept up to date). And few, if any, other communities have tried to survey the landscape of locally focused sites as the Trust has in Chicago. For fair cross-market comparisons, we will need a replicable approach for identifying seed sites. Some steps that should be included:

- **Identify the key traditional media sites:** daily and weekly newspapers, television and radio stations, locally focused magazines, alternative weeklies, specialized publications (business and sports), etc.
- **Identify high-traffic sites** by reviewing services such as comScore, Nielsen NetRatings, Compete and Quantcast.
- **Use search engines** to find relevant sites by entering search terms such as "[city name] news," "[city name] sports news" and "[city name] business news."
- **Use the best available list of neighborhood and town names** to search for websites about those communities.
- **Identify sites that link to the initial list** by searching Google using the "link:" query tool.

(3) Combine network analysis with other research about sites in the ecosystem.

The data gathered through network analysis could be made more valuable if it could be combined with other information about the sites in the ecosystem. For instance, if we had information about the age of different sites, the number of staff members, the site managers' attitudes toward linking, or the amount of traffic to different sites, we could explore correlations between link patterns and these other variables.

(4) Explore the relationship between links and user traffic

While the link structure of a network of Web pages can yield important insights into the relationships among those sites, the existence of a link doesn't necessarily mean that people click and follow that link. It would be quite desirable to understand the characteristics of a link that are most likely to cause a user to click. For instance, does the link appear in the body of a post, in a "blogroll" or in a list of links at the end of the post? Or, what language is used to describe the linked-to content ("check out this important post" versus "here's an example of a really stupid idea").

Unfortunately, there is no single data source that tracks traffic generated by users' clicks on hyperlinks. But in a specific community, it would be possible through our research approach to identify a group of sites that receive many inbound links. With the collaboration of those sites' publishers, researchers could gain access to the "referrer" logs showing what links were clicked on to drive their traffic. Referrer information could be analyzed against the number of inbound links – turning up sites, for instance, that link a lot but don't drive much traffic (or vice versa). Analysis of this data could yield actionable insights into the best ways of using links to drive audience attention.

(5) Conduct time-series research

With any or all of the research approaches we envision, there is great potential value in repeating the research over different time periods. Such research will reveal patterns and trends, and changes in the news ecosystem, which cannot be captured in a single snapshot at one moment in time. Research over time will enable deeper understanding of the dramatic changes that continue to take place in the "new news ecosystem," especially those driven by the rise of other platforms such as mobile phones and digital tablets.

(6) Extract link-sharing data from Twitter

Our webcrawling approach captured links from Chicago sites to Twitter, but not the links to Chicago sites from Twitter users. Through Twitter's application programming interface (API), it would be possible to extract links shared on Twitter. Then one could analyze the link data to understand the nature of link-sharing that happens on Twitter – and the prevalence of links from Twitter to Chicago news/information sites.

Acknowledgements

This report was commissioned by The Chicago Community Trust's Community News Matters project and was made possible by generous support from The Richard H. Driehaus Foundation.

It is one of a series of five Community News Matters reports designed to shed light on current conditions, opportunities and challenges in the Chicago area's news ecosystem. Other reports in the series are:

- *The NEW News: Journalism We Want and Need* and *NEW News 2010: Mapping Chicago's Online News Scene*, the first and second inventories of the area's online news sites;
- *Realizing Potential: What Chicago's Online Innovators Need*, which outlines the kinds of help needed by the area's online media innovators;
- *News That Matters: An Assessment of Chicago's Information Landscape*, which details the views of the public and of area leaders about the adequacy, strengths and weaknesses of Chicago's information landscape (to be released this summer).

Community News Matters is a funder collaborative created by The Chicago Community Trust in 2009 in response to the Knight Foundation's Community Information Challenge, a five-year, \$24 million initiative to inspire community foundations to find creative ways to use new media and technology to keep communities informed. Other funders of Community News Matters are the McCormick Foundation, the John D. and Catherine T. MacArthur Foundation and The Woods Fund of Chicago.

Links:

The New News: Journalism We Want and Need:

<http://communitymediaworkshop.org/newnews/2009report/>

NEW News 2010: Mapping Chicago's Online News Scene:

http://www.cct.org/sites/cct.org/files/CNM_TheNewNews2010.pdf

Realizing Potential: What Chicago's Online Innovators Need:

http://www.cct.org/sites/cct.org/files/CNM_RealizingPotential.pdf

Appendix A: Top Sites Ranked by Network Statistics

Top Authorities (ranked by binary in-degree [number of sites that link to these], excluding services)			
Rank	Site Name	Category	Number of Sites
1	transitchicago.com	org/inst	40
2	chicagotribune.com	legacy	28
3	gapersblock.com	micro	25
4	mcachicago.org	org/inst	24
5	metrarail.com	org/inst	22
6	pitchforkmusicfestival.com	org/inst	21
7	chicagoist.com	micro	19
	msichicago.org	org/inst	19
9	hydeparkart.org	org/inst	18
	fieldmuseum.org	org/inst	18
11	millenniumpark.org	org/inst	17
12	chicagoreporter.com	legacy	16
	artic.edu	org/inst	16
14	suntimes.com	legacy	15
	newcity.com	legacy	15
	chicagoreader.com	legacy	15
	newcitychicago.com	legacy	15
	windycitizen.com	micro	15
	colum.edu	org/inst	15
	emptybottle.com	org/inst	15
	cityofchicago.org	org/inst	15
	chipublib.org	org/inst	15
	chicagoparkdistrict.com	org/inst	15
24	chicagonow.com	legacy-affil	14
	hideoutchicago.com	org/inst	14
	uchicago.edu	org/inst	14
	activetrans.org	org/inst	14
	cct.org	org/inst	14
	rtachicago.com	org/inst	14
30	gooseisland.com	org/inst	13
	schubas.com	org/inst	13
	smartmuseum.uchicago.edu	org/inst	13
	pacebus.com	org/inst	13
34	lincolnhallchicago.com	org/inst	12
	chicagoartistsresource.org	org/inst	12

Top Authorities

(ranked by binary in-degree [number of sites that link to these], excluding services)

Rank	Site Name	Category	Number of Sites
	ravinia.org	org/inst	12
	http://www.doubledoor.com/	org/inst	12
38	newcitynetwork.com	legacy	11
	chicagomag.com	legacy	11
	beachwoodreporter.com	micro	11
	nationalmuseumofmexicanart.org	org/inst	11
	steppenwolf.org	org/inst	11
	westernexhibitions.com	org/inst	11
	dusablemuseum.org	org/inst	11
	lollapalooza.com	org/inst	11
	metrochicago.com	org/inst	11
	beatkitchen.com	org/inst	11
	subt.net	org/inst	11
49	chicagojournal.com	legacy	10
	get.adobe.com	natbrand	10
	spudnikpress.com	org/inst	10
	secondcity.com	org/inst	10
	siskelfilmcenter.org	org/inst	10
	chicagoculturalcenter.org	org/inst	10
	congresschicago.com	org/inst	10

Top Hubs

(ranked by binary out-degree [number of sites that these link to], excluding services)

Rank	Site Name	Category	No. of Sites
1	gapersblock.com	micro	47
2	badatsports.com	micro	44
3	saic.edu	org/inst	23
4	uchicago.edu	org/inst	21
	macfound.org	org/inst	21
	blogs.southtownstar.com	legacy	21
	thegallerycrawlandsomuchmore.blogspot.com	micro	21
8	avclub.com	legacy	19
9	communitymediaworkshop.org	org/inst	17
10	artslant.com	micro	16
	onthemake.org	micro	16
12	chicagoartistsresource.org	org/inst	15
	music.newcity.com	legacy	15
14	chicagomag.com	legacy	14
	literago.org	micro	14
	loudlooppres.com	micro	14
	windycitymediagroup.com	legacy	14
18	cityofchicago.org	org/inst	13
	chicagoismynewblog.wordpress.com	micro	13
	progressillinois.com	micro	13
	thelocaltourist.com	micro	13
	indiesomnia.com	micro	13
23	transitchicago.com	org/inst	12
	windycitizen.com	micro	12
	activetrans.org	org/inst	12
	chicago.metblogs.com	micro	12
	newcitystage.com	legacy	12
	windycityrock.net	micro	12
29	msichicago.org	org/inst	11
	best.newcity.com	legacy	11
	chicagoweekly.net	legacy	11
	jot.org	org/inst	11
33	radiofreechicago.typepad.com	micro	10
34	chicagoparkdistrict.com	org/inst	9
	smartmuseum.uchicago.edu	org/inst	9
	chicagojournal.com	legacy	9

Top Hubs

(ranked by binary out-degree [number of sites that these link to], excluding services)

Rank	Site Name	Category	No. of Sites
	adlerplanetarium.org	org/inst	9
	chicagostorytelling.com	micro	9
	loyolaphoenix.com	legacy	9
	outsidetheloopradio.com	micro	9
	sallyduros.com	micro	9
	art.newcity.com	legacy	9
	chicagosprogressivetalk.com	legacy	9
44	fieldmuseum.org	org/inst	8
	chicagonow.com	legacy-affil	8
	beatkitchen.com	org/inst	8
	courttheatre.org	org/inst	8
	probonobaker.typepad.com	micro	8
	chicagosnacks.org	micro	8
	fromtheledge.com	micro	8
	gazettechicago.com	legacy	8
	mindfulmetropolis.com	legacy	8
53	dot.state.il.us	org/inst	7
	chicagotalks.org	micro	7
	beverlyartcenter.org	org/inst	7
	southtownstar.com	legacy	7
	centerstagechicago.com	legacy-affil	7
	forgottenchicago.com	micro	7
	underground-library.org	org/inst	7
	urchicago.com	legacy	7

Top Switchboards

(ranked by betweenness, excluding services)

Rank	Site Name	Category	Betweenness
1	gapersblock.com	micro	4079.1
2	transitchicago.com	org/inst	3704.1
3	windycitizen.com	micro	3388.5
4	saic.edu	org/inst	2002.4
5	chicagoartistsresource.org	org/inst	1684.3
6	chicagomag.com	legacy	1567.6
7	cityofchicago.org	org/inst	1320.9
8	macfound.org	org/inst	1223.1
9	uchicago.edu	org/inst	1074.5
10	chipublib.org	org/inst	859.6
11	msichicago.org	org/inst	827.7
12	hydeparkart.org	org/inst	768.4
13	activetrans.org	org/inst	761.1
14	metrarail.com	org/inst	749.3
15	artic.edu	org/inst	731.2
16	ravinia.org	org/inst	721.1
17	chicagonow.com	legacy-affil	624.6
18	lollapalooza.com	org/inst	542.0
19	fieldmuseum.org	org/inst	512.7
20	metrochicago.com	org/inst	502.5
21	chicagoparkdistrict.com	org/inst	494.8
22	chicagojournal.com	legacy	475.2
23	beatkitchen.com	org/inst	462.2
24	mcachicago.org	org/inst	457.0
25	beachwoodreporter.com	micro	393.2
26	courttheatre.org	org/inst	373.5
27	smartmuseum.uchicago.edu	org/inst	369.5
28	newcity.com	legacy	295.4
29	nextartfair.com	org/inst	287.8
30	oldtownschool.org	org/inst	277.6
31	chicagotalks.org	micro	276.5
32	cct.org	org/inst	274.5
33	chicagohistory.org	org/inst	259.7
34	northwestern.edu	org/inst	249.1
35	steppenwolf.org	org/inst	237.5
36	millenniumpark.org	org/inst	237.5

Top Switchboards

(ranked by betweenness, excluding services)

Rank	Site Name	Category	Betweenness
37	pacebus.com	org/inst	233.8
38	rtachicago.com	org/inst	233.3
39	adlerplanetarium.org	org/inst	210.2
40	beverlyartcenter.org	org/inst	203.7
41	sbnation.com	natbrand	182.0
42	westernexhibitions.com	org/inst	168.1
43	spertus.edu	org/inst	152.7
44	cantv.org	org/inst	144.1
45	siskelfilmcenter.org	org/inst	125.2
46	areachicago.org	org/inst	118.9
47	mocp.org	org/inst	115.4
48	dot.state.il.us	org/inst	109.4
49	congresschicago.com	org/inst	93.4
50	three-walls.org	org/inst	92.0

Top Referrers

(ranked by out-closeness, excluding services)

Rank	Site Name	Category	Out-Closeness
1	communitymediaworkshop.org	org/inst	0.289
	chicago.metblogs.com	micro	0.289
	chicagostorytelling.com	micro	0.289
	outsidetheloopradio.com	micro	0.289
	sallyduros.com	micro	0.289
	chicagocarless.com	micro	0.289
	austintalks.org	micro	0.289
	ctatatattler.com	micro	0.289
9	123 sites are tied		

Top Resources

(ranked by in-closeness, excluding services)

Rank	Site Name	Category	In-Closeness
1	chicagotribune.com	legacy	0.396
2	hydeparkprogress.blogspot.com	micro	0.395
	sbnation.com	natbrand	0.395
4	nytimes.com	legacy	0.392
	suntimes.com	legacy	0.392
6	chicagoreader.com	legacy	0.391
	gooseisland.com	org/inst	0.391
	newcity.com	legacy	0.391
	pitchforkmusicfestival.com	org/inst	0.391
	versionfest.org	org/inst	0.391
11	newcitychicago.com	legacy	0.39
	newcitynetwork.com	legacy	0.39
13	chicagoist.com	micro	0.389
	chicagoreporter.com	legacy	0.389
	lillstreet.com	org/inst	0.389
	wgnradio.com	legacy	0.389
17	chicagobusiness.com	legacy	0.388
	colum.edu	org/inst	0.388
	hideoutchicago.com	org/inst	0.388
	nationalmuseumofmexicanart.org	org/inst	0.388
	neofuturists.org	org/inst	0.388
	printersball.org	org/inst	0.388
23	emptybottle.com	org/inst	0.387
	lincolnhallchicago.com	org/inst	0.387
	nictd.com	org/inst	0.387
	thepublicanrestaurant.com	org/inst	0.387
27	areachicago.org	org/inst	0.386
	blackpearl.org	org/inst	0.386
	illinois.gov	org/inst	0.386
	renaissancesociety.org	org/inst	0.386
	schubas.com	org/inst	0.386
	secondcity.com	org/inst	0.386
	spudnikpress.com	org/inst	0.386
34	activetrans.org	org/inst	0.385
	adlerplanetarium.org	org/inst	0.385
	artic.edu	org/inst	0.385
	bottomlounge.com	org/inst	0.385

Top Resources

(ranked by in-closeness, excluding services)

Rank	Site Name	Category	In-Closeness
	cantv.org	org/inst	0.385
	cct.org	org/inst	0.385
	chicagoartistsresource.org	org/inst	0.385
	chicagoartsdistrict.org	org/inst	0.385
	chicagoculturalcenter.org	org/inst	0.385
	chicagohistory.org	org/inst	0.385
	chicagomag.com	legacy	0.385
	chicagoparkdistrict.com	org/inst	0.385
	chipublib.org	org/inst	0.385
	cityofchicago.org	org/inst	0.385
	courttheatre.org	org/inst	0.385
	dot.state.il.us	org/inst	0.385
	doubledoor.com	org/inst	0.385
	dusablemuseum.org	org/inst	0.385
	fieldmuseum.org	org/inst	0.385
	hydeparkart.org	org/inst	0.385
	macfound.org	org/inst	0.385
	mcachicago.org	org/inst	0.385
	metrarail.com	org/inst	0.385
	millenniumpark.org	org/inst	0.385
	mocp.org	org/inst	0.385
	msichicago.org	org/inst	0.385
	northwestern.edu	org/inst	0.385
	oldtownschool.org	org/inst	0.385
	pacebus.com	org/inst	0.385
	ravinia.org	org/inst	0.385
	rtachicago.com	org/inst	0.385
	saic.edu	org/inst	0.385
	siskelfilmcenter.org	org/inst	0.385
	smartmuseum.uchicago.edu	org/inst	0.385
	spertus.edu	org/inst	0.385
	steppenwolf.org	org/inst	0.385
	three-walls.org	org/inst	0.385
	transitchicago.com	org/inst	0.385
	uchicago.edu	org/inst	0.385
	westernexhibitions.com	org/inst	0.385
	windycitizen.com	micro	0.385

Appendix B: List of Seed Sites

Seed Sites from The Chicago Community Trust Survey

600words.com
AdentroDePilsen.org
andthatsmyopinion.com
areachicago.org
art.newcity.com
arteyvidachicago.com
artoridiocy.blogspot.com/
artpilsen.blogspot.com/
austintalks.org
beachwoodreporter.com
best.newcity.com
blogs.vocalo.org
blogs.wttw.com/moreonthehistory/
blogs.wttw.com/unfiltered/
boozemuse.com
boutiqueville.com
broadshouldersupdate.com
bronzecomm.com
buzzflash.com
catalyst-chicago.org
centersquarejournal.com
centerstagechicago.com
chicago.everyblock.com/
chicagoargus.blogspot.com/
ChicagoArtCollector.com
ChicagoArtMagazine.com
chicagobites.com
Chicagoist.com
chicagoistheworld.org
chicagonewscoop.org
Chicagonista.com
chicagoparent.com
chicagoreader.com
chicagosportsinhaiku.com
chicagosprogressivetalk.com
chicagostorytelling.com
chicagotalks.org

chicagotribune.com
chicagov.org
communitymediaworkshop.org/npcommunicator
crabbygolightly.com
DaMadhouse.com
district299.com
driftlessarea.blogspot.com
edgewatercb.blogspot.com
evanstonnow.com
feastoffun.com
forgottenchicago.com
francesarcher.wordpress.com
gapersblock.com
greenparentchicago.com
HelloBeautifulBlog.com
hpherald.com
illinoisresource.net
insideonline.com
jot.org/journal_of_ordinary_thought.php
keepusconnected.org
lakeeffectnews.com
latina-voices.com/wp04/
latinopolicyforum.org
lit.newcity.com
LiveHereOakPark.com
lsna.net
marathonpundit.blogspot.com
metroplanning.org/news-events/blog-posts/
mindfulmetropolis.com
morsehellhole.blogspot.com
music.newcity.com
musicbox-online.com/
myfoxchicago.com
newcity.com
newcityfilm.com
newcitystage.com
newcommunities.org
nuf-said.org/
nwi.com
outsidetheloopradio.com
pilsenportal.org
playhotpotato.wordpress.com/
post-trib.com

progressillinois.com
proyectolatina.org
RadioChicagoland.com
resto.newcity.com
revistacontratiempo.com
RP1000.blogspot.com
sallyduros.com
secondcitycop.blogspot.com
seedingchicago.wordpress.com
skokienet.org
sloopin.com
so-LAZE.com
southlandbusinessnews.blogspot.com
southlandsavvy.com
southsnews.com
southwestobserver.com
sox35th.com
SuburbanChicagoland.com
summer.newcity.com
suntimes.com
theexpiredmeter.com
thehinsdalean.com
thelivingroom.blogspot.com/
thesixthward.blogspot.com
thisishell.com
thriftista.com
triblocal.com
uptownupdate.com
uscatholic.org
vivelohoy.com
wbez.org
wellesparkbulldog.com
wethepeoplemedia.org
wgnradio.com
windycitizen.com
windycitymediagroup.com
wttw.com
wzrdchicago.org

Seed Sites Identified from Fwix.com (<http://fwix.com/Chicago/browse/sources>)

312diningdiva.blogspot.com/
5chicago.com/index.html
60646blog.com/
abclocal.go.com/wls/index
achicagosojourn.blogspot.com/
activetrans.org/
alwaysintransit.typepad.com/hyde_park_urbanist/
anothercubsblog.net/
apartmenttherapy.com/chicago
arcchicago.blogspot.com/
artslant.com/chi/main
auditoriumtheatre.blogspot.com/
austinweeklynews.com/
avclub.com/chicago/
babystew.com/
badatsports.com/
bearsbacker.com/
bearsgab.com/
beingtotallysweetinchicago.blogspot.com/
belmontheights.blogspot.com/
berwynpoliceblotter.com
beverlymorganpark.net/
blackhawks.nhl.com/index.html
bleachernation.com/
bleedcubbieblue.com/
blog.chicagoarchitecture.info/
blog.chicagohistory.org/
blogdownchicagobears.com/
blogs.napervilleparks.org/default.aspx
blogs.pioneerlocal.com/
blogs.southtownstar.com/
blogs.suburbanchicagonews.com/
bluegraysky.blogspot.com/
brokehipster.com/
bullspodcasters.com/
cbs2chicago.com/
cdobs.com/
charliesnasties.blogspot.com/
chiblogo.wordpress.com/
chicago.citysearch.com/

chicago.cubs.mlb.com/news/headlines.jsp?c_id=chc
chicago.grubstreet.com/
chicago.metblogs.com/
chicago.metromix.com/
chicago.thedelimagazine.com/
chicago.timeout.com/
chicago.whitesox.mlb.com/news/headlines.jsp?c_id=cws
chicagoappraisers.blogspot.com/
chicagoartdepartment.org/
chicagobears.com/news/
chicagobreakingbusiness.com
chicagobreakingnews.com/blog/
chicagobreakingsports.com
chicagobrunchblog.com/
chicagobulldailynews.com/
chicagoburgerproject.blogspot.com/
chicagobus.org/
chicagobusiness.com
chicagocarless.com/
chicagocondosonline.blogspot.com/
chicagocouriersunion.org/
chicagocubs-baseball.com/
chicagocurrent.com/
chicagofedblogs.org/
chicagoflame.com/home/
chicagofoodies.com/
chicagogluttons.com/
chicagohotblog.com/
chicagoismynewblog.wordpress.com/
chicagojournal.com/
chicagolawbulletin.com/browser_pass.cfm
chicagomag.com/
chicagomaroon.com/
chicago-nightlife.blogspot.com/
chicagonow.com
chicago-outdoor-sculptures.blogspot.com/
chicagopizzaclub.com/
chicagosindependentmusicreview.com/
chicagosnacks.org/
chicago-sportsfans.com/
chicagosportstoday.com/
chicagoweathercenter.com/
chicagoweekendfun.com/

chicagoweekly.net/
chicagowolves.com/
chitownliving.com/
cltv.com
collegeofidiots.com/
colum.edu/News/index.php
columbiachronicle.com/
communitybeat.blogspot.com/
cribchatter.com/
csnchicago.com/
c-tas.blogspot.com/
ctatattler.com/
cubsstats.blogspot.com/
dabullseye.com/
dailycandy.com/chicago/
dailyherald.com/
dailyphotobutcherfortheworld.blogspot.com/
depaul.rivals.com/
depauliaonline.com/
dgnbands.blogspot.com/
dgreport.com/
dimbeautyofchicago.blogspot.com/
diningchicago.com/blog/
downersgrovechronicle.wordpress.com/
dspotblog.com/
dusablemuseum.org/
espn.go.com/chicago/
examiner.com/chicago/
eyeteeth.blogspot.com/
fiftythird.uchicago.edu/
fightingillini.com/
flowfeel.blogs.com/flowfeel/
forestparkreview.com/
fotoflow.blogspot.com/
foulballs.net/
frequentforeclosures.com/
fromtheledge.com/
gazettechicago.com/index/
globest.com/topics/chicago.html
greenbeanchicago.com/
groovychicago.com/
hailtotheorange.com/
herloyalsons.com/

hipsternascar.com/
huffingtonpost.com/chicago
hydeparkart.org/
hydeparkblvd.com/
hydeparkprogress.blogspot.com/
illinibballfansblog.blogspot.com/
illinoisloyalty.com/
indiesomnia.com/
irisheyespowerhour.blogspot.com/
irishroundtable.com/
irishsportsdaily.com/
ivychat.blogspot.com/
kcchronicle.com
kiss-chicago.blogspot.com/
kitchensinkcafe.blogspot.com/
lakeclaremontpress.blogspot.com/
lawyerjim.blogspot.com/
literago.org/
litterboxcats.com/
loganchamber.org/
logansquarefarmersmarket.org/
lollapalooza.com/
loudlooppres.com/
loyolaphoenix.com/
loyolaramblers.cstv.com/
marshfieldtattler.blogspot.com/
midwayillustrated.com/
mybikeadvocate.com/
naperville.wordpress.com/
nba.com/bulls/
nbcchicago.com/
nearwestsidecdc.org/home.aspx
neighborhoodnetworking.wordpress.com/
news.eastvillagechicago.org/
newshour.org
nhna.blogspot.com/
northpark.edu/News
northwestern.edu/newscenter/
nusports.cstv.com/
nwherald.com/
onthemake.org/
openillinois.org/
orangealert.net/blog

orlandparker.blogspot.com/
outside.in/chicago-il
overheardinchicago.blogspot.com/
peoplingplaces.wordpress.com/
pioneerlocal.com/
preservationpilsen.blogspot.com/
pressblog.uchicago.edu/
probonobaker.typepad.com/probonobaker/
pureparents.org/
qcfdc.org/home.aspx
radiofreechicago.typepad.com/
rakesofmallow.com/
rblandmark.com/
redeyechicago.com
ridiculousupside.com/
rogersparkbench.blogspot.com/
rubbishgoeshere.com/dump/
savechicagoculture.org/
seriouseats.com/
southchicagoan.blogspot.com/
southsidesox.com/
southtownstar.com
soxmachine.com/
suburbanchicagonews.com/index.html
subwaydomer.com/
swopchicago.org/home.aspx
swside.blogspot.com/
talkingchicagobaseball.com/
teamworkenglewood.org/home.aspx
theatreinchicago.com/index.php
theblogthatjordanbuilt.com/
thechicago77.com/
thechicagoartblog.wordpress.com/
thecubdom.com/
thecubreporter.com/
thecubsbrickyard.com/
theflatironproject.com/
thegallerycrawlandsomuchmore.blogspot.com/
thelocalbeet.com/
thelocaltourist.com/
themidwasteland.com/
thenapervillepost.com/
thenorthcoast.blogspot.com/

thepodcollective.wordpress.com/
theregionalnews.com/
theurbancoaster.com/
theweekbehind.com/
thrillarama.com/
tremendousupsidepotential.com/
truchicago.blogspot.com/
uchiblogo.uchicago.edu/
uchicagolaw.typepad.com/
uhnd.com/
uncommonapplication.blogspot.com/
underground-library.org/?cat=580
unityparkchicago.org/
urchicago.com/
uttaps.wordpress.com/
vincemichael.wordpress.com/
visit-cp.blogspot.com/
wednesdayjournalonline.com/
weisnd.blogspot.com/
westnorth.com/
westridgebungalowneighbors.blogspot.com/
whitesoxhomeplate.com/
whitesoxpride.mlblogs.com/
windycitygridiron.com/
windycityrock.net/
wrigleyville23.com/
yochicago.com/
yourseason.suntimes.com/index.html

Appendix C: About the Authors

Rich Gordon

Rich Gordon is professor and director of digital innovation at the Medill School of Journalism, Media, Integrated Marketing Communication at Northwestern University. His research and teaching focus on interactive publishing, including online audience development, social media and online communities. He co-developed, with Northwestern Prof. Noshir Contractor, an undergraduate course focusing on the application of network theory to media and journalism.

Before coming to Northwestern, Prof. Gordon was the first online director for the Miami Herald Publishing Co., overseeing editorial and business operations. Before that he worked as reporter, bureau chief and editor for newspapers in Virginia and Florida, where he was one of the early leaders in computer-assisted reporting.

Zachary Johnson

Zachary Johnson is the Co-Founder and CEO of Syndio Social. In January 2009, he initiated the expansion of the company by partnering with Professor Noshir Contractor to provide scalable network analysis solutions. He has secured and led projects for numerous clients, including Procter & Gamble.

Mr. Johnson studied network analysis at Northwestern University under the tutelage of Prof. Contractor. Prior to founding Syndio Social, he applied network analysis methodologies to numerous independent music marketing projects. His most successful collaboration was with pop-star Mike Posner. In 2009, Mr. Johnson advised the digital distribution of Posner's debut mix tape, "A Matter of Time," which has established him as an up-and-coming artist. Posner's hit single "Cooler Than Me," reached the tops of both Billboard charts and the iTunes music store.

Linking Audiences to News: A Network Analysis of Chicago Websites

By **Rich Gordon** (Northwestern University)
and **Zachary Johnson** (Syndio Social)
April 2011

© 2011 - Syndio Social

All rights reserved. Any redistribution or reproduction of part or all of the contents in any form is prohibited other than the following: a) you may print or download to a local hard disk extracts for your personal and non-commercial use only, b) you may copy the content to individual third parties for their personal use, but only if you acknowledge Syndio Social as the source of the material. You may not, except with our express written permission, distribute or commercially exploit the content.

THE CHICAGO COMMUNITY TRUST

AND AFFILIATES

111 East Wacker Drive
Suite 1400
Chicago, IL 60601
312.616.8000
www.cct.org

The Chicago Community Trust, our region's community foundation, partners with donors in ways that transform lives and communities. From mitigating foreclosures to feeding the hungry, the Trust grants more than \$100 million annually to nonprofit organizations that support arts and culture, basic human needs, community development, education and health. By connecting the generosity of donors with the needs of the community, the Trust ensures our region thrives today and for future generations.