

SNA Unit 4

Visualizing and Interpreting Facebook Networks

4.1.1 Introduction: The World's Social Graph

- **Publicly Articulated Networks:** Facebook represents a unique class of networks. They are not based on invisible behavior (like email traffic) but on the relationships we intentionally "show" to others to manage access to information.
 - **Scope:** At the time of writing, Facebook was the largest social network (~400 million members), with the **News Feed** acting as the primary stream of information shared between friends.
-

4.1.2 Historical Context

- **Growth Strategy:** Unlike MySpace (which allowed total customization), Facebook used **exclusivity**. It started at Harvard, moved through specific universities, and finally opened to the general public.
 - **Network Effects:** By targeting universities, Facebook benefited from pre-existing real-world connections.
 - **Contradiction:** Facebook provides granular privacy tools but pushes "information libertarianism" by making data as public and discoverable as possible (e.g., the introduction of the News Feed in 2006).
-

4.1.3 Why Map a Facebook Network?

1. **Privacy Management:** Identifying if a specific group (like "Family") is truly separate from other clusters.
 2. **Networking Style:** Identifying if you are a **Team Player** (closing triads/introducing people) or a **Broker** (keeping groups separate to maintain strategic value).
 3. **Strategic Planning:** Event planners use maps to see if their audience is one dense cluster or multiple disconnected groups.
 4. **Social Hygiene:** Finding "zombie" contacts added years ago but never engaged with.
-

4.1.4 What Kind of Network is Facebook?

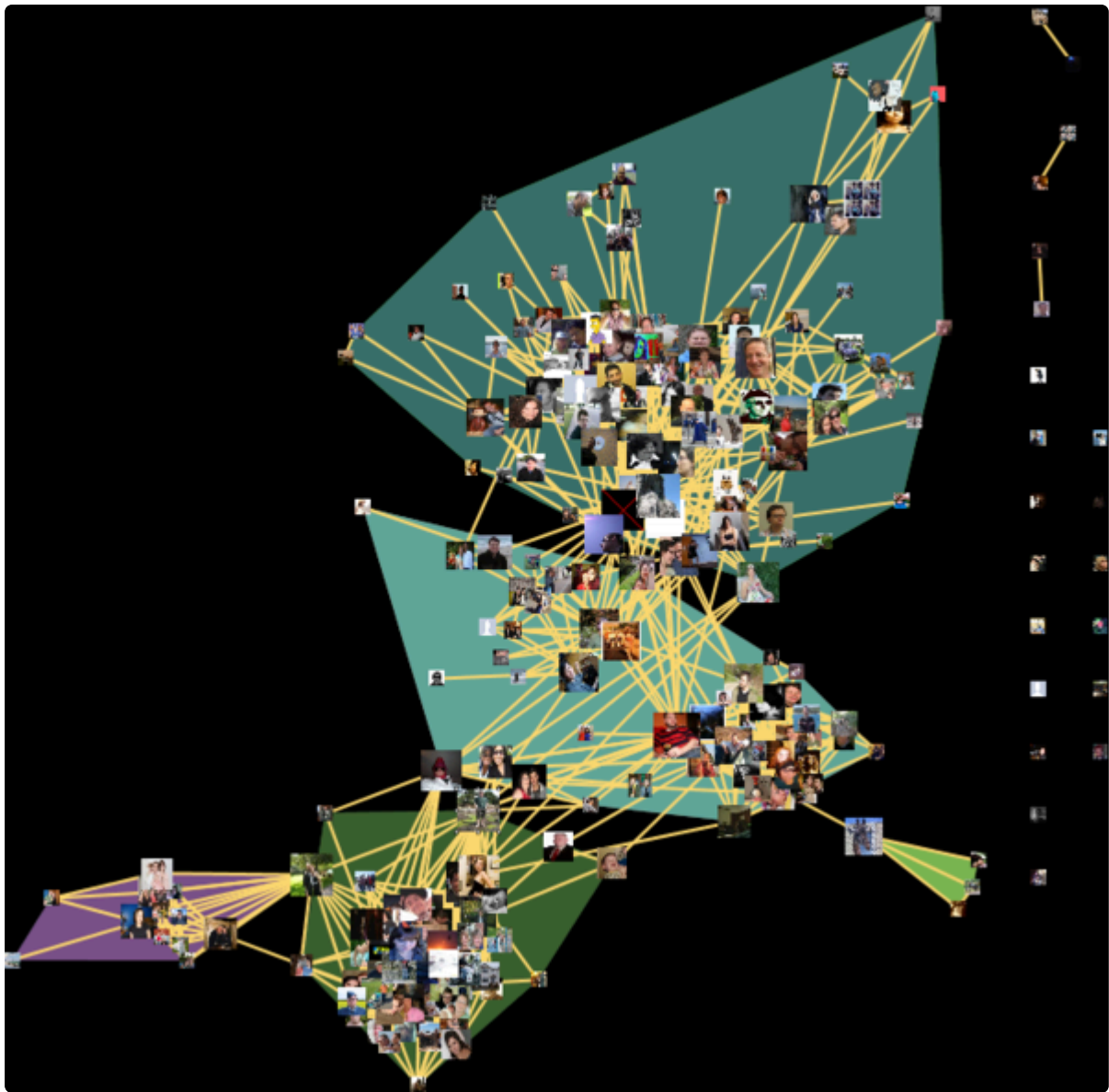
Facebook friendship networks are **Egocentric Networks**.

- **Ego:** The focal person (the owner of the network).

- **Alters:** The friends connected to the Ego.
 - **Network Levels:**
 - **1.0 Degree:** Just Ego and their alters (a simple star shape).
 - **1.5 Degree:** Ego, alters, and the connections *between* those alters. **This is what NodeXL maps.**
 - **2.0 Degree:** ego, alters, and *their* friends (including people Ego doesn't know). Facebook's API does not allow this.
 - **Properties:**
 - **Undirected:** Friendships must be mutual.
 - **Unweighted:** By default, all friend connections are treated as equal.
-

4.1.5 - 4.1.6 Basic Visualization in NodeXL

- **Data Source:** Traditionally imported via apps like `NameGenWeb` using the **GraphML** format.
- **Hiding Ego:** In an ego network analysis, it is standard practice to **exclude the owner (Ego)**. Since Ego is connected to everyone, including them clutters the graph and obscures the internal clustering of friends.
- **The "Networky" Look:**
 - **Layouts:** Fruchterman-Reingold and Harel-Koren are best.
 - **Iterations:** Default settings (10 iterations) are often too low for Facebook. Increasing to **80-100 iterations** helps resolve distinct clusters.



4.1.7 Data Types and Attributes

Categorical (Nonordered) Data

- **Examples:** Gender, Hometown, Cluster ID.
- **Visual Mapping:** Use **Color** or **Shape**.
- **VLOOKUP Strategy:** Advanced users create a "Categories" worksheet to manually assign colors/shapes to groups like "Male," "Female," or "Unknown."

Numerical (Ordered) Data

- **Examples:** Degree, Betweenness Centrality, Age.
- **Visual Mapping:** Use **Size** or **Opacity**.
- **Key Interpretations in Ego Networks:**
 - **Degree:** The number of mutual friends between you and that person.
 - **Betweenness:** Identifies **Bridges**. A person who links your high school friends to your work colleagues has high betweenness and likely knows

you very well.

- **Eigenvector Centrality:** Identifies those in the center of dense clusters. Mapping this to opacity can make dense groups "glow" or become transparent to see the structure inside.

4.1.8 From Friend Wheel to Pinwheel

- **Friend Wheel:** A popular radial layout where nodes are arranged in a circle. It is aesthetically pleasing and avoids node overlap, but can be hard to interpret.
- **Pinwheel Layout (The NodeXL Way):** A customized clustered radial layout.
 - **Logic:** Groups are arranged in "flames" or wedges around a circle.
 - **Mapping:** Radius is scaled to **Betweenness** (pulling brokers toward the center), while Size/Color are mapped to **Degree**.
 - **Benefit:** Reveals the internal density of a cluster while simultaneously showing how that cluster links to the rest of the network.

4.1.9 - 4.1.10 Practitioner and Researcher Summary

- **Practitioners:** Facebook maps reveal the "social context" of your life. Using Excel features like `VLOOKUP` and `LOG` formulas allows for much more sophisticated visuals than the standard "one-click" apps.
- **Researchers:** * **Clustering Limits:** Is it more useful to see "Work" as a group, or to find the "soft partition" of people who belong to both "Work" and "Sports"?
 - **Dunbar's Number:** Human brains have a cognitive ceiling of about **150 stable relationships**. Does Facebook act as a "cognitive enhancement" to break this limit, or just lead to information overload?

Exam Tip: Be prepared to explain why **Betweenness Centrality** is used to identify "Bridges" in Facebook and why the **Ego** should be removed before calculating metrics.

WWW Hyperlink Networks

4.2.1 Introduction to Web Networks

- **The World Wide Web (WWW)** is the largest machine-readable network graph on Earth.
- **Graph Components:**
 - **Vertices (Nodes):** Individual web pages.
 - **Edges (Ties):** URL hyperlinks connecting one page to another.
- **Organizational Web Presence:** While "Web 2.0" (social media) is vital, the "Web 1.0" or **static web presence** remains the primary medium for building corporate or institutional identity.

- **Key Insight:** Unlike other social media networks that link *people*, hyperlink networks primarily link **organizations and institutions**.
 - **Business Value:** Analyzing these networks reveals how an organization's online position matches its offline brand presence and provides ethical competitive intelligence.
-

4.2.2 Theory and Methodology of Hyperlinking

4.2.2.1 The Theory of Hyperlinking

Hyperlinks act as a form of "web currency." There is no single theory for why sites link to each other, but common motivations include:

- **Authority and Endorsement:** A link acts as a "vote" of confidence or credibility.
- **Trust:** Reflects a reliable relationship between two entities.
- **Alliance Building:** Creating a "critical mass" for a shared message or viewpoint.
- **Negative Affect:** Linking to a site specifically to criticize it.

Visibility vs. Retrievability

- **Retrievability:** An **absolute** concept. A site is retrievable if its server is operational.
- **Visibility:** A **relative** concept. Visibility is determined by the number of inbound links from other relevant, high-ranking sites.

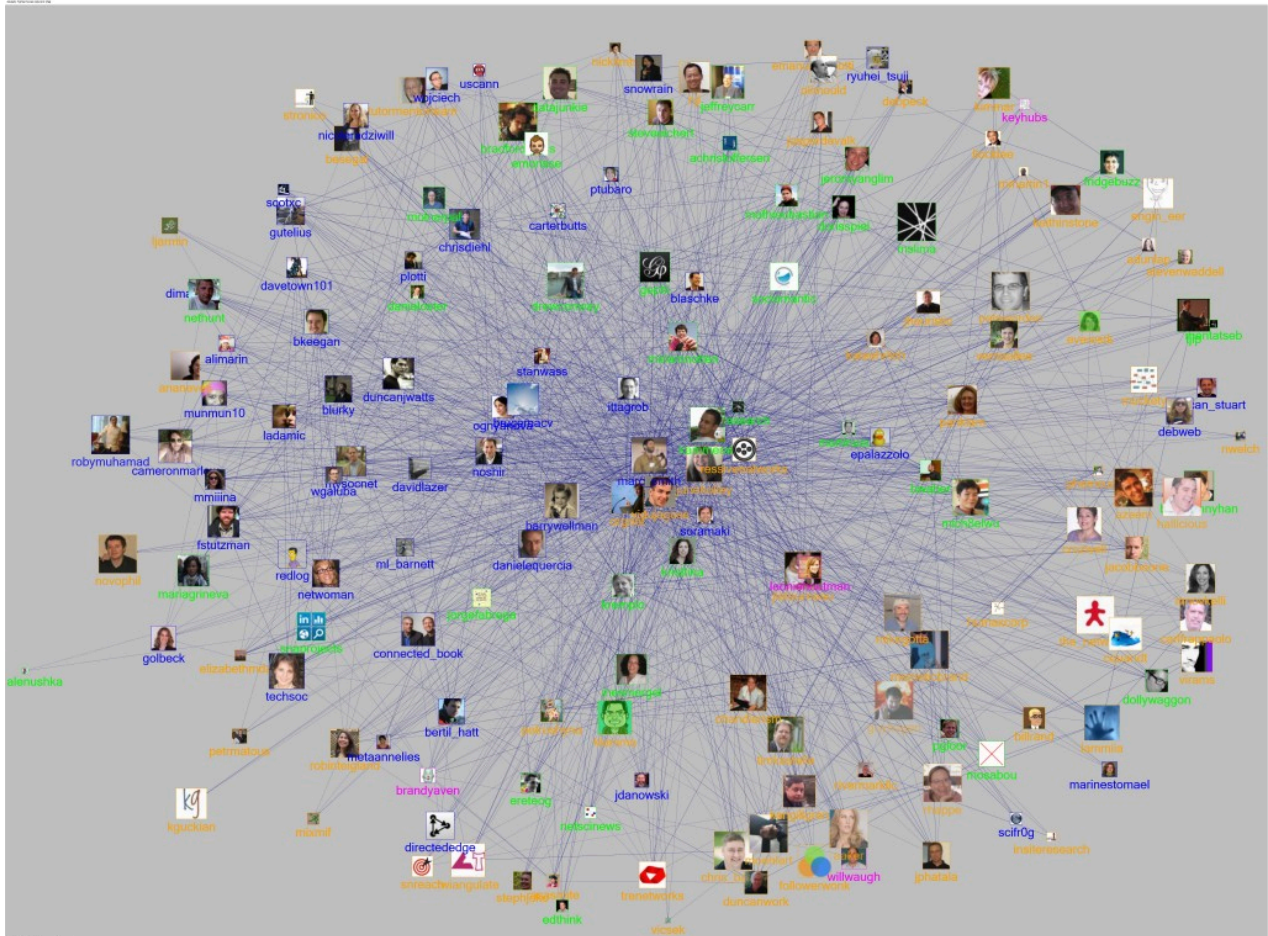
4.2.2.2 Methodological Issues

Analyzing hyperlink networks requires defining three parameters:

1. **Nodes:** Are they pages or entire websites?
 - **Meta-nodes:** Analysts often group pages from a single hostname or subdomain into one vertex to represent an entire organization.
 2. **Ties:** Are the edges directed or undirected?
 - **Weighting:** Can be based on the number of links between sites or the "depth" of the link within the site structure.
 3. **Boundaries:** The web is "borderless," making it hard to define where a network ends.
 - **Snowball Sampling:** Starting with a set of "seed sites" and crawling outward to discover the surrounding network.
-

4.2.3 The VOSON Data Provider

- **Definition:** A NodeXL plug-in (Virtual Observatory for the Study of Online Networks).
- **Function:** It provides a front-end for a web crawler that extracts hyperlinks and uses the Yahoo! API to find inbound links to specific sites.
- **Significance:** It allows non-programmers to conduct complex web-crawling and network analysis tasks within the familiar Excel interface.



4.2.4 Practical Example 1: The Ego Network

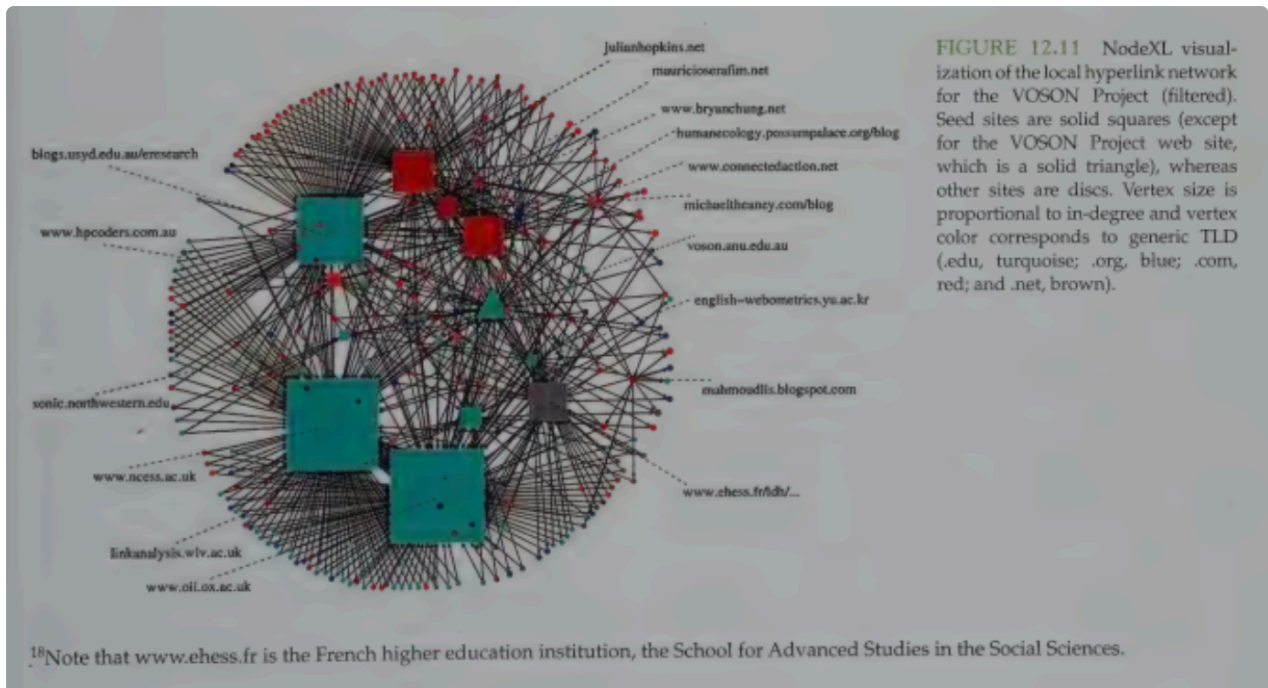
This explores who links to a specific organization (using the VOSON Project site as the "Ego").

The Process:

1. **Seed Sites:** Input the URL of the target organization.
2. **Crawl Parameters:**
 - * **Inbound Crawl:** Finding who links to you.
 - * **Outbound Crawl:** Finding who you link to.
3. **Analysis of TLDs (Top-Level Domains):** * Mapping vertex colors to TLDs (e.g., .edu, .com, .org) reveals the diversity of an organization's connections.

Managing "Topic Drift"

- **Topic Drift:** As you crawl deeper, the network can "blow up" with irrelevant sites (e.g., a relevant blog linking to a random cooking site).
- **The Solution:** Create a subgraph containing only "important sites"—defined here as sites with an **undirected degree** (meaning they are connected to at least two other sites in your target network).



4.2.5 Practical Example 2: Mapping a Field/Industry

Instead of an ego network, this maps an entire sector (e.g., "Social Network Analysis software").

Key Findings:

- **Central Actors:** Sites like INSNA or software providers (UCINET, Pajek) appear with high **in-degree** (authority) because many others link to them.
- **Hubs vs. Authorities (Kleinberg's Theory):**
 - **Authorities:** Provide specialized, high-value information (e.g., software distributors).
 - **Hubs:** Sites that provide organized lists of links to authorities (e.g., Wikipedia, Answers.com).
- **Actionable Insight:** If a top-ranked site in your industry does *not* link to you, submitting a request for a link can significantly increase your traffic and search engine ranking.



FIGURE 12.13 A filtered NodeXL visualization of seed sites focused on "social network analysis" plus important sites (sites with degree greater than or equal to two). The favicon (where available) is displayed for the seeds, whereas red discs are used to represent non-seed sites.

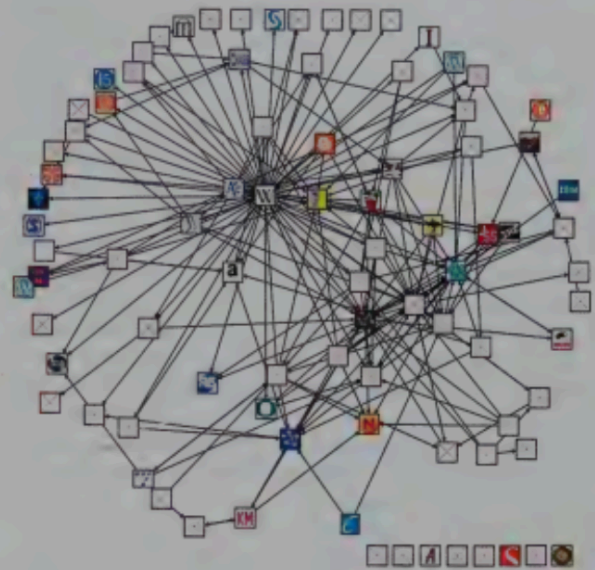


FIGURE 12.14 A filtered NodeXL visualization of seed sites focused on "social network analysis," using the favicon (where available) for the graph vertex.

4.2.6 Advanced Topics: The "Holy Grails"

1. **Blog Networks:** Much more complex than static sites because they require **longitudinal data** (links at specific time points) and the distinction between **permalinks** (in body text) and **blogrolls** (side lists).
2. **Dynamic Network Analysis:** Answering "How did the network get this way?" by looking at archived data (e.g., from the Internet Archive).
3. **Network Flow:** Measuring the actual volume of traffic moving along the "pipes" (hyperlinks), which usually requires access to private web logs or expensive third-party data (e.g., Alexa, Hitwise).

4.2.7 Practitioner's Summary

- **Measuring Success:** Use NodeXL to identify the leaders in your industry and emulate the structure of their hyperlink network.
- **Strategy:** * If you want to be a **Hub**, find the best **Authorities** to link to.
 - If you want to be an **Authority**, identify the top **Hubs** and request they link to you.
- **Caveat:** Always verify automated data with **domain knowledge**. "Garbage in, garbage out" applies heavily to web crawling.

4.2.8 Researcher's Agenda

- **E-Government:** Researching the "nodality" (central position) of government sites in social and informational networks.
 - **Network Ethnography:** A promising area combining quantitative hyperlink analysis with qualitative methods to understand the *meaning* behind the connections.
-

Links:

[SNA Unit 1](#)

[SNA Unit 2](#)

[SNA Unit 3](#)

[SNA Unit 4](#)

[SNA Unit 5](#)

[Communication Technologies](#)

[Conversational AI](#)