UNIT – IV
Case Study – II
Facebook, WWW Hyperlink

By
Ranjith Kumar M
Assistant Professor, CSE
SRITW

Content

- **Facebook**
  - Visualizing and interpreting Facebook Networks
- **WWW Hyperlink Networks**

Facebook

- Users interact with content on Facebook in a variety of ways, including a range of reactions to posts, posting comments, and sharing posts with their Facebook friends.
- Most of these interactions are private. They cannot be observed by users outside one's set of friends and typically cannot be captured by third party tools.
- In contrast, the networks that form when users engage with content within publicly available pages (fan pages) are visible.
- Engaging with content on a public Facebook page, users form networks of shared interests when they react, comment and share content.

Facebook

- **Facebook API limits**
- Facebook has limited the data it provides publicly for free, but some data is provided on relatively liberal terms.
- Facebook fan page data access is limited in terms of the speed data is delivered but no time frames are excluded, in contrast to platforms like Twitter that limit public access to the past 7 or 8 days at any time.
- Facebook provides as much historical data from fan pages as requested although the download may take quite a while, since the download is stopped and started to stay within the query per time budget allowed.

Organizational networks: Fan pages

- **Preparing the data collection**
- Facebook fan pages are a popular way for a group, company or organization to present itself.
- The fan page network captures inter-organizational connectivity on Facebook.
- When you search Facebook for fan pages, only use results that appear under the Pages tab.
- Using results from other tabs will cause an error when you try to collect data.

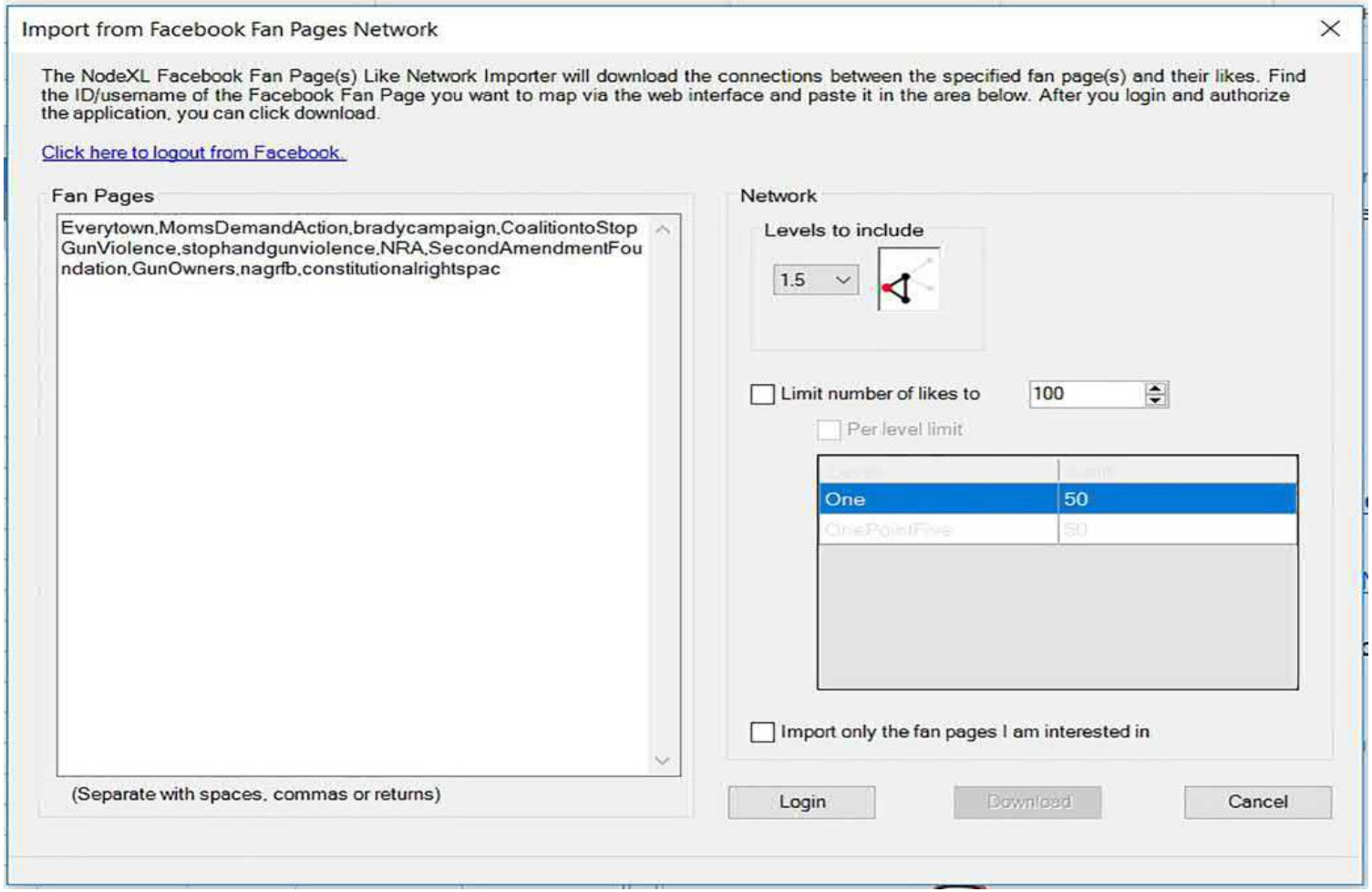
Organizational networks: Fan pages

- **Data Collection**
- NodeXL includes a built-in importer for Facebook Fan Pages.
- The importer does not require that you are an administrator of any fan page.
  - Select the Import drop-down menu from the NodeXL Ribbon and choose From Facebook Fan Pages Network.
  - In the Import from Facebook Fan Pages Network dialog type all the page names (or URLs) that you identified. Separate pages by a space or a comma. We will call each of these pages a Seed page.
  - Set the Levels to include to 1.5. This will include a complete network of each Seed Facebook fan page, other pages each Seed page liked or were liked by (i.e., alters), connections among these alters, and any connections with other selected Seed pages or their alters.

Organizational networks: Fan pages

- **Data Collection**
- NodeXL includes a built-in importer for Facebook Fan Pages.
  - The importer does not require that you are an administrator of any fan page.
  - Ideally, you do not want to Limit number of likes, unless a page has a very large number of alters.
  - Unselect the checkbox next to Import only the fan pages I am interested in. Checking this would only consider links between the Seed fan pages. In most cases, only including the Seed fan pages is too restrictive, so you will want to uncheck this box.
  - Click Login. If it is the first time, you will need to login to the Facebook page that will be opened.
  - Click Download.
  - Data collection may take time. Even an hour or more, depending on the number and popularity of the pages

Organizational networks: Fan pages

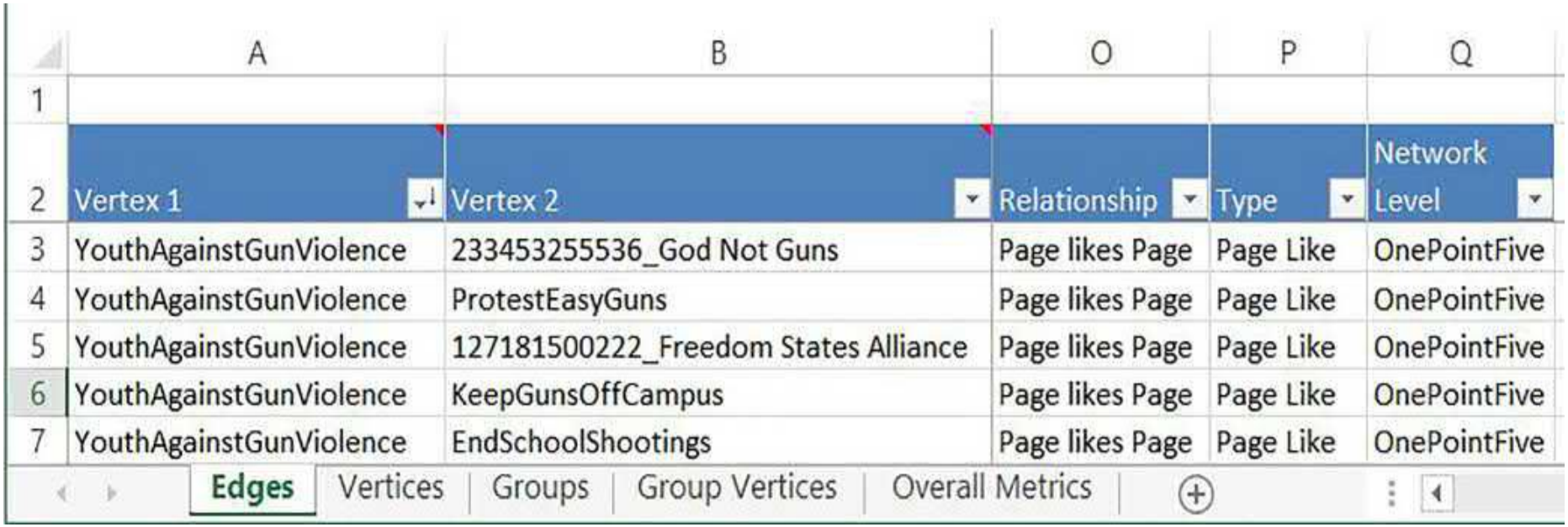


Organizational networks: Fan pages

- **Edges worksheet**
- Select the Edges tab in the NodeXL file, where a pair of connected fan pages is the unit of analysis.
- You will find the list of like-based relationships, with the fan page in Vertex 1 liking the fan page in Vertex 2
- Since only 1.5 levels of edges were selected for inclusion in the network, the edge list includes like-relationships between seed fan pages and their alters (Network Level equal to One) and between alters (Network Level equal to OnePointFive).
- As you analyze and visualize the network, all relationship-based metrics will be found on this worksheet.

Organizational networks: Fan pages

- **Edges worksheet**



	A	B	O	P	Q
1					
2	Vertex 1	Vertex 2	Relationship	Type	Network Level
3	YouthAgainstGunViolence	233453255536_God Not Guns	Page likes Page	Page Like	OnePointFive
4	YouthAgainstGunViolence	ProtestEasyGuns	Page likes Page	Page Like	OnePointFive
5	YouthAgainstGunViolence	127181500222_Freedom States Alliance	Page likes Page	Page Like	OnePointFive
6	YouthAgainstGunViolence	KeepGunsOffCampus	Page likes Page	Page Like	OnePointFive
7	YouthAgainstGunViolence	EndSchoolShootings	Page likes Page	Page Like	OnePointFive

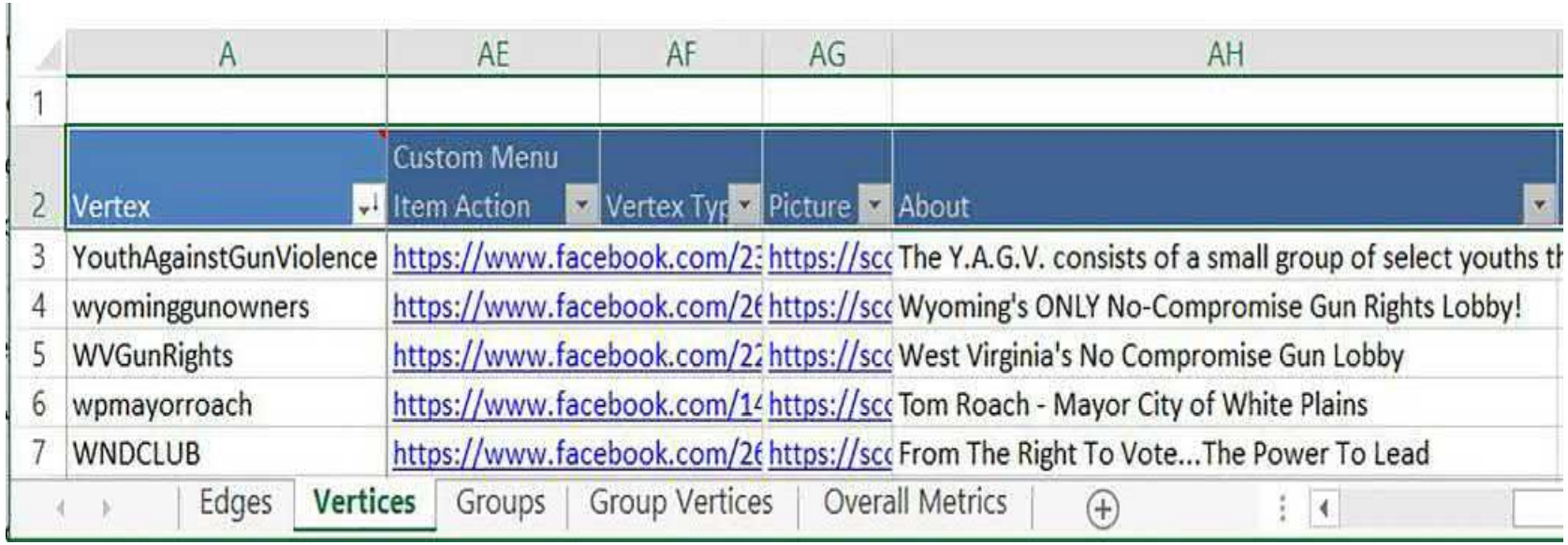
Edges | Vertices | Groups | Group Vertices | Overall Metrics

Organizational networks: Fan pages

- **Vertex worksheet**
- Vertices worksheet where a fan page (i.e., the Vertex column) is the unit of analysis.
- Some useful information includes a link to each fan page of Facebook, a direct link to the profile image that can be utilized for visualization (i.e., the Picture column), and a self description of the page for easy interpretation (i.e., the About column).
- As you analyze and visualize the network, all fan page-based metrics will be found here.

Organizational networks: Fan pages

- **Vertex worksheet**



	A	AE	AF	AG	AH
1					
2	Vertex	Custom Menu Item Action	Vertex Typ	Picture	About
3	YouthAgainstGunViolence	https://www.facebook.com/2:		https://scc	The Y.A.G.V. consists of a small group of select youths th
4	wyominggunowners	https://www.facebook.com/2(		https://scc	Wyoming's ONLY No-Compromise Gun Rights Lobby!
5	WVGunRights	https://www.facebook.com/2:		https://scc	West Virginia's No Compromise Gun Lobby
6	wpmayorroach	https://www.facebook.com/1(		https://scc	Tom Roach - Mayor City of White Plains
7	WNDCLUB	https://www.facebook.com/2(		https://scc	From The Right To Vote...The Power To Lead

Edges | Vertices | Groups | Group Vertices | Overall Metrics

Network Analysis

- Network as a topic or an issue-based community of organizations. The selection of the pages e.g., a political race, a social cause or an industry determines the domain of interest, interaction, and informing flow. Like Twitter-topic networks, all social network metrics must be evaluated within the boundaries of the network.
- **Vertex-level analysis**
- Ties between fan pages are directed, since "liking" a page is a unidirectional action.
- This is unlike friendship-ties among two Facebook users' personal pages, since "friending" only occurs when there is a mutual agreement.

Network Analysis

- Use the Graph Metrics tool to calculate the In-Degree, Out-Degree, Betweenness Centrality, Reciprocity, PageRank, and Overall Metrics.
- In-degree centrality measures the number of fan pages that "liked" a given fan page. It captures attention given to an organization or another public entity on Facebook.
- out-degree centrality measures the number of fan pages a given page "liked."
- This is an indication of the attention a fan page gives to other pages, and the channel that this page opens to its visitors to other related pages.

Network Analysis

- Sort on the In-Degree and Out-Degree columns and examine which fan pages fill these important network positions.
- Betweenness centrality helps identify bridge spanners who connect other fan pages that would not otherwise be as closely connected
- Reciprocity measures the mutuality of directed ties.
- In this network, the vertex-level reciprocity (i.e., Reciprocated Vertex Pair Ratio) measures the percent of cases where one fan page "liked" another fan page, which in turn "liked" the original fan page.

Network Analysis

- PageRank centrality is another useful metric that captures important fan pages that considers the number of other fan pages that "like" the page, as well as characteristics about the linking fan pages such as their centrality and link propensity.
- Fan pages that rank highly are ones that are "liked" by other popular fan pages.
- They are the ones who those "in the know" are pointing toward.
- PageRank considers the directed nature of the network, unlike Eigenvector Centrality, which shares many characteristics, but is designed primarily for undirected networks.

Network Analysis

- **Cluster-level analysis**
- A cluster, or subgroup of more tightly interconnected nodes, captures the naturally occurring and self-selected boundaries of information flow within a given network.
- Decades of literature highlights the tendency of individuals to connect with similar others, a phenomenon known as *homophily*, or "Birds of a Feather Flock Together".
- In Facebook's fan networks, clusters emerge when groups of pages "like" one another.

Network Analysis

- **Cluster-level analysis**
- Within a given domain of interest, clusters can capture, often hidden, sub-groups of pages that share common interest, common causes or even reflect factions within a large community devoted to a common cause.
- Calculate network clusters by choosing Group by Cluster from the Groups drop-down menu in the NodeXL Ribbon.
- Try out the three different clustering algorithms to see if one seems to create groups at the most meaningful level of granularity. Then make sure and use the Graph Metrics dialog to calculate Group metrics.

Network Analysis

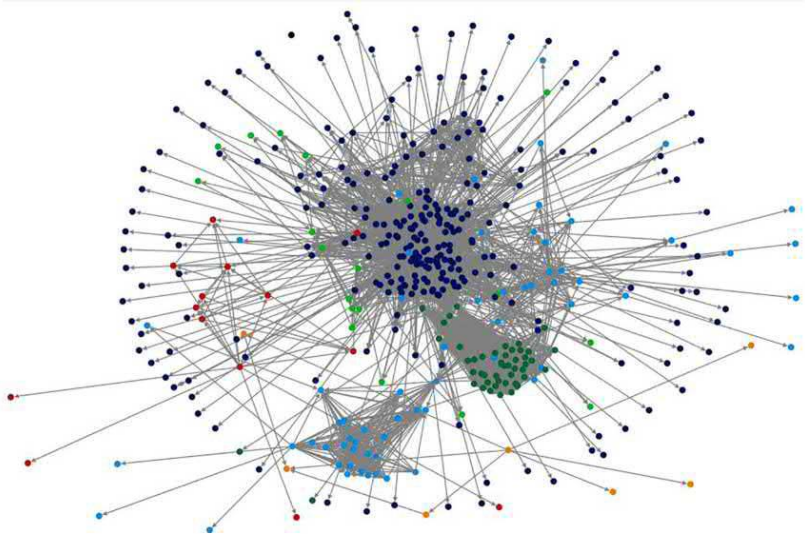
- **Cluster-level analysis**
- With the aim of understanding the cluster-structure of your network, examine cluster-level metrics and fan pages' characteristics.
- The Groups worksheet includes metrics about each cluster, including number of vertices (fan pages), number of edges (likes) and density (level of interconnectedness).
- It also provides visualization opportunities, which are the major clusters in the network? How are they different in terms of size, volume, reciprocity and density? Note that cluster size is inversely correlated with density, so larger clusters tend to have lower density.

Network Analysis

- **Visualization**
- Click Show Graph in NodeXL. The results may not initially look very meaningful.
- However, if the network is similar to the Gun Control network, you may see some clear clusters that could be visually separated to add clarity to the visual.
- In fact, it is not even clear from the current visualization that several of the clusters are not connected to each other at all (i.e., they are distinct connected components).

Network Analysis

- **Visualization**
- In order to layout the graph by its key clusters, choose Layout Options... from the layout drop-down menu and choose the Lay out each of the graph's groups in its own box option
- It include using the Size and Label options. You may even want to use Color or Shape to indicate which pages are Seed pages.



WWW Hyperlink Networks

- World Wide Web (www) hyperlink networks within the context of Social Network Analysis (SNA)
- **Network Structure:**
- Nodes and Edges: Nodes represent webpages, and edges denote hyperlinks between them.
- Directed Graph: The hyperlink network forms a directed graph, as hyperlinks have a source (the linking page) and a target (the linked page).

WWW Hyperlink Networks

- **Metrics:**
- Page Rank: Developed by Google, it evaluates the importance of webpages based on the quantity and quality of inbound links.
- HITS Algorithm (Hyperlink-Induced Topic Search): Identifies two types of nodes: hubs (many outgoing links) and authorities (linked to by hubs).
- Useful for understanding the role of pages as both information sources and information consumers.
- Link Analysis for Ranking: Search engines use link analysis algorithms to rank webpages.

WWW Hyperlink Networks

- **Metrics:**
- Nodes as Webpages, Edges as Hyperlinks:Apply SNA principles to hyperlink networks by considering webpages as nodes and hyperlinks as edges.
- Analyze centrality, degree, and clustering coefficients to understand the structure.
- User-Webpage Interaction: Extend SNA to study interactions in web-based social networks.
- Links can represent interactions, mentions, or shared content between users and webpages.
- Community Detection: Identify communities or groups of webpages that share similar content or are frequently linked together.

WWW Hyperlink Networks

- 1. Data Collection:
- Web Data Collection: Use NodeXL to import data from the web. It supports importing data from Twitter, YouTube, and other sources, but for www hyperlink networks, consider importing data from web pages directly.
- Web Page Data Import: Use the "Import" menu in NodeXL to pull data from web pages. This can include extracting hyperlinks and their relationships.

WWW Hyperlink Networks

- NodeXL Graph Visualization:
- NodeXL provides a visual representation of the hyperlink network. Nodes represent webpages, and edges represent hyperlinks.
- Color and Size Coding:
- Utilize color and size to encode additional information. For instance, node size can represent the degree of a webpage (number of hyperlinks).

WWW Hyperlink Networks

- **NodeXL Metrics:**
- Centrality Measures: NodeXL calculates various centrality measures, such as degree centrality, betweenness centrality, and eigenvector centrality, which provide insights into the importance of webpages in the network.
- Clustering Coefficient: Analyze the clustering coefficient to understand how tightly interconnected groups of webpages are within the hyperlink network.
- Filtering Nodes and Edges: Apply filters to focus on specific nodes or edges, enabling a more targeted analysis.

WWW Hyperlink Networks

- Grouping Nodes: NodeXL allows you to group nodes based on attributes. For example, you can group webpages by content category.
- Temporal Analysis: If your data includes timestamps, NodeXL can visualize the evolution of the hyperlink network over time.
- Dynamic Filters: Use dynamic filters to explore the network at different time points.
- NodeXL includes community detection algorithms to identify groups of webpages with dense connections. This helps in understanding thematic clusters within the hyperlink network.