

INDIANA UNIVERSITY - BLOOMINGTON

NETWORK SCIENCE

---

# Measuring Facebook Influence Using Centrality Measures and Community Detection

---

*Authors*

DANIEL HINDERS  
NHI TRAN

*Professor*

YONG-YEOL (YY) AHN

May 2, 2020

# Measuring Facebook Influence Using Centrality Measures and Community Detection

Daniel Hinders and Nhi Tran

May 2, 2020

## Abstract

Within the past couple decades, social media has become a large part of everybody’s life and our society. With the aid of social media, social influence can take place without physical human interaction and allow people to influence each other just by having social media accounts and the ability to access the Internet. After social media became popular, researchers and scientists have been studying the impact of social media using different methods such as surveys, trials, scientific experiments, and so on. Analyzing social media network graphs is another way of studying social media impact that this paper will cover. Given a public dataset of Page to Page network from Facebook in November 2017, multiple network graph measurements and a community detection method will help identify influential Facebook pages within a network. By understanding the differences between measurements and what information they provide, it will be beneficial to apply similar measurements and analysis methods on other social media network to identify influential nodes.

## 1 Introduction

With individuals having the ability to influence others anytime from any location anonymously, it is important to be able to identify their power and influential level within a social network to carefully assess the impact. Without the ability to identify influential nodes, a network can be too large to study and time-consuming for anyone to get useful information.

### 1.1 Background

Social media impact has drastically increased its effect on life in the last 10-15 years, the diverse and far reaching impacts cannot be overstated. Russian influence in the 2016 election, the Obama campaign utilizing social media to achieve a decisive victory in 2008 and social media fueling societal change and upheaval in the middle east and Hong Kong are some of the examples.

In another realm, corporations spend a massive 11.6 percent [?] average of their marketing through social media indicating vast impact on societies choices and buying habits through social media.

In the midst of these large organizations and vast movements in social media, there is a fasci-

nating and powerful impact propagated through lone individuals on social media. A single individual operating as a social media influencer is an existence that has become increasingly common in our societal vernacular. For many, a full time job as a social influencer has become a viable career.

Understanding all of these social media forces and impacts is important on many levels. Devastating negative ramifications: a nation state subverting democracy by impacting other countries' elections, hate groups spreading and inciting violence, and webs of misinformation leading to widespread distrust are all reasons alone to understand social media impact and influence.

## 1.2 Motivation and Objectives

The ultimate purpose of this paper is to analyze a public Facebook Page-to-Page network to identify important nodes and different communities.

The first objective is to answer the question that will observe the potential differences between multiple centrality measurements that help identify important nodes within a network - Can we simply use degree to identify powerful nodes in a network?

The second objective is to answer the question that will identify the different between a pre-determined classification of a network and how likely would the nodes form a community within their own classification - Can we use page type to detect communities?

## 1.3 Existing Work

**Existing Work 1** [?] Text here

**Existing Work 2** [?] Text here

**Existing Work 3** [?] Text here

# 2 Process and Approaches



Figure 1: High Level Process

**Figure 1** identifies the steps that are necessary to complete the objectives of this report. Data are collected from Stanford Network Analysis Project. Once the data are collected, the next would be to clean up the network graph, perform high level analysis, calculations, detection community method. Then, all findings will be analyzed, visualized, and compared to get final results.

## 2.1 Collecting Data

**Dataset [1]** The dataset is November 2017 Facebook Large Page-Page Network from Stanford Network Analysis Project [1]. The dataset is a compressed zip file that contains the following files :

- `musae_facebook_edges.csv` – File that contains all edges between pages
- `musae_facebook_features.json` – JSON file that contains features of the nodes - this file will not be used.

- `musae_facebook_target.csv` - File that contains list of node ids, page names, and page types.

Each edge represents a mutual like between pages and each node is a Facebook page. There are four categories classifies for the Facebook pages: governmental organizations, politicians, television shows and companies. The raw data contains 22,470 nodes and 171,002 edges.

## 2.2 Data Cleaning and High Level Information

**Tools** The following software and tools were used to perform the analysis:

- **Python and Python Packages** - Python packages include Networkx, Pandas, and Matplotlib
- **Jupyter Notebook**
- **Gephi** - visualization and analysis tool for network graph

## Cleanup

## High Level Information

## 2.3 Calculate Measurements

### 2.3.1 Betweenness Centrality

### 2.3.2 Closeness Centrality

### 2.3.3 Eigenvector Centrality

## 2.4 Community Detection - Modularity

## 2.5 Visualize, Analyze and Compare

## 3 Discussion

## 4 Conclusion

## 5 Acknowledgment

## 6 References

## References

- [1] S. N. A. Project, "Facebook Large Page-Page Network."  
<https://snap.stanford.edu/data/facebook-large-page-page-network.html>. Accessed Online on 2020-07-03.