Social Network Analysis on Food Category Facebook Pages

Lulu Maudhuna Alfani
School of Informatics
Telkom University
Bandung, Indonesia

Tiara Novri Zamroni
School of Informatics
Telkom University
Bandung, Indonesia

Laurentius Yudhistira Adhipratama
School of Informatics
Telkom University
Bandung, Indonesia

Tyas Nur Kumala
School of Informatics
Telkom University
Bandung, Indonesia

Ariiq Afrahtama
School of Informatics
Telkom University
Bandung, Indonesia

Abstract

Facebook is a widely used social media platform for sharing and community building, especially in the food category. Facebook food pages play a significant role in shaping culinary trends and influencing consumer preferences. This study aims to analyze the social network of Facebook food pages using centrality metrics such as degree centrality, betweenness centrality, closeness centrality, as well as PageRank, Epidemic Model, Contagion, and Community Detection. The data for this research comes from the fb-pages-food dataset from networkrepository.com, representing mutually liked relationships among food-themed pages. The research process involves data processing, centrality metrics analysis, and evaluating the impact of page interactions on consumer behavior. By understanding the network structure, the study provides insights for page managers to optimize content strategies and influence online culinary trends more effectively.

Keywords—component, Social Network, Facebook Pages, Food Category, Centrality Metrics, Culinary Trends

I. INTRODUCTION

Social media platforms like Facebook have become central to fostering interactions and community building across various fields, including the culinary sector. Food-focused Facebook pages now play a critical role in shaping culinary trends, influencing consumer preferences, and creating a space for interaction between users and food producers. Every day, thousands of users engage in discussions, share recipes, and provide food recommendations on these pages, generating a complex social network that reflects a dynamic ecosystem of food-related information and connections.

Analyzing the "Facebook Pages Food" network allows for an exploration of relationships between food-focused pages, uncovering influential pages within the community and understanding how these interactions shape online culinary trends. As social media significantly impacts consumer decision-making, analyzing the structure and interconnectedness of these Facebook pages becomes highly relevant, with recommendations from popular pages often guiding consumer choices in food and dining.

This study aims to offer insights into the dynamics of the online culinary community and its effects on consumer behavior. Additionally, understanding the network structure can aid page managers in developing more effective content strategies. To this end, social network analysis will employ centrality metrics, including degree centrality, betweenness centrality, closeness centrality, and PageRank, to identify key nodes, intermediary pages, and those with extensive reach and high-quality connections. This approach provides a structured framework for examining how food-related Facebook pages influence culinary trends and consumer preferences.

II. RELATED WORK

In a recent paper by Freeman introduced various measures of centrality, including degree centrality, closeness centrality, and betweenness centrality. These concepts are essential for understanding the position and role of individuals within social networks, which can also be applied in the context of wireless sensor networks to identify the most critical nodes for data collection [1].

Brandes developed a more efficient algorithm for calculating betweenness centrality, which measures how often a node appears on the shortest paths between pairs of other nodes. This algorithm is significant in network analysis as it helps identify nodes that function as connectors within the network, which can influence the design and operation of sensor networks [2].

Huang et al. discuss various localization schemes that do not rely on distance measurements, which are often challenging to perform in wireless sensor networks. A method that utilizes network topology and connectivity information to estimate the location of nodes [3].

Ilyas et al. [22] identified groups of nodes, social hubs, in the network which are at the center of influential neighborhoods. Tey then compared their results with the ones from the method using Eigenvector centrality (EVC). To further enhance the usage of α-centrality, Ghosh et al. introduced a normalized version of this centrality by generalizing a modularity maximization-based approach. Their method identified not just the local communities but also the global ones [23].

III. METHODOLOGY

In this study, we used the 'fb-pages-food' dataset obtained from the Network Repository. This dataset represents the relationship between Facebook pages in the food category. Each node in the dataset represents a Facebook page, and each edge represents the relationship between the pages. To maintain the anonymity of the pages, nodes are identified by unique numbers in this dataset. The structure of the dataset consists of two main files: fb-pages-food.edges, which contains a list of relationships between nodes in pair format (source and target), and fb-pages-food.nodes, which contains node data and attributes that may be related, such as page names or subcategories. These network characteristics make it possible to analyze larger social networks that show unique patterns of relationships within online communities, with the potential for central nodes or tight-knit community groups within the culinary industry. The methods used in this social network analysis focus on different types of centrality measures to identify the role and degree of influence of each node within the network, namely with degree centrality, closeness centrality, betweenness centrality, and PageRank centrality.

Measure Definitions

Degree Centrality. A graph theory-based approach used for whole-brain analysis at the voxel level, reflecting FC in brain networks, as well as exploring the degree of connection of each node in the network with other nodes [4]. In this study, pages that have a lot of centralities will be more frequently connected to other pages and can be considered as pages that are quite popular or frequently followed in the food category. For a graph G = (V, E) with n vertices, the degree centrality CD(v) for v is:

$$CD(v) = \frac{deg(v)}{n-1} \quad (1)$$

Closeness Centrality. A node is considered closest to another node if it is in the closest location in a particular network. In other words, it also indicates the shortest path connected between the nodes in use. As a result, the closest node in the network has the highest visibility. The calculation of Closeness Centrality (CC) is much easier than the calculation of Degree Centrality. A target node's CC value is found by finding the shortest distance between all nearby nodes in the WSN and then finding the reciprocal value of these distances to get the CC value [5]. Closeness centrality can be calculated by the equation:

$$CC(v) = \frac{1}{\sum a \in V / v dG(v,a)} \quad (2)$$

Betweenness Centrality. An important metric that refers to the number of times a node serves as a bridge for the shortest path between two other nodes. Betweenness centrality is measured as the number of paths that pass through an edge (node) as a proportion of the total number of shortest paths across all shortest paths. Thus, betweenness centrality reflects the role and influence of nodes and edges in the overall network. In this research, betweenness centrality is used to identify pages that act as links between other pages [6]. Pages with high betweenness centrality values have the potential to play a role as information mediators or connectors between communities in the network. A graph G = (V, E) has n vertices, the betweenness centrality can be calculated as follows [5].

- For each pair of nodes (*a*,*b*), find all possible shortest paths.
- For each pair of nodes (a,b), compute all fractional path that are in question.
- Sum the fraction over all possible nodes in pairs.

$$CB(v) = \sum a \backslash nev \backslash neb \frac{lab(v)}{lab} \quad (3)$$

PageRank. An algorithm that is designed to rank web pages on the Google search engine and display search results based on the highest ranking. The purpose of this algorithm is to calculate the rank of each web page on the network based on inbound and outbound links [7]. PageRank can be calculated using the equation as follows:

$$PR(v) = (1 - \alpha) + \alpha \sum \frac{PR(u)}{\mathrm{deg} deg\,(u)} \quad (4)$$

where α is the damping factor, generally 0.85. For networks where the influence of a particular node depends on many connections, PageRank centrality is very good. In this case, a food page with a high PageRank can be regarded as a popular page, which is often connected to other influential pages in the network.

Contagion. Contagion in social network analysis refers to the process by which certain behaviors, information or attributes spread through a network of interconnected individuals. This concept is critical to understanding how social interactions can influence individual actions and decisions within a broader social context. Contagion is divided into several types, local contagion and global contagion. Local contagion occurs when the spread of information or behavior is limited to direct connections between individuals in the network. On the other hand, global contagion involves wider spread and can occur through external sources, such as mass media or recommendation platforms [8].

One of the most often used models to explain how behavior or knowledge spreads in social networks is the Independent Cascade Model (ICM). The model assumes that every exposure to a naïve person raises their risk of contracting the infection on their own, independent of any further consequences from earlier exposures [9]. In the ICM model, everyone has a chance of becoming infected after being exposed to an infected individual. This probability is expressed as follow,

$$P\left(t;\ n_e, n_f\right) \ = \ \sum_{n=1}^{n_e} F(n)V_n\left(t, \left\{t_1, \ldots, t_{ne}\right\}; n_f\right)$$

which is the probability of information transmission from one person to another. The more exposures an individual experiences, the more likely they are to be infected [9].

Eigenvector Centrality is an extension of degree centrality. In the degree centrality, all node connections are credited of equal importance. But in real life, each node may have different importance. For example, a node connected to highly important nodes itself is an important node. Tus, Eigenvector centrality provides a relative score to each node depending on the type of nodes (high-scoring and low-scoring) it is connected to. For a given graph G = (V, E) containing n nodes, let A be the adjacency matrix of G and be the eigenvalue. Ten Eigenvector centrality is given by:

$$C_e(v_i) = \frac{1}{\lambda}\sum_{j=1}^{n} d_i a_{j,i} C_e(v_j)$$

IV. RESULTS AND DISCUSSION

A. Network Visualization

The network visualization highlights the relationships among Facebook food pages. Nodes represent individual pages, and edges depict mutual likes between these pages. Key hubs are visually identifiable, indicating their importance in the network.

B. Degree Centrality

Degree centrality measures how many direct connections (edges) a node in the graph has compared to other nodes. In the context of the fb-pages-food dataset, degree centrality indicates how connected an entity (such as a restaurant, chef, or brand) is to other entities.

From the calculation using degree centrality, it is found that Logan Junior Chef is the most influential among other entities with a degree centrality of 0.21647819063004847. This can mean that Logan Junior Chef has many direct connections to other nodes. Logan Junior Chef plays an important role in the network, for example connecting various chefs, restaurants, or food brands.

C. Closeness Centrality

Closeness centrality measures the average closeness of a node to all other nodes in the graph, based on the shortest path length. A high closeness value indicates that a node is strategically positioned to reach other nodes quickly. In the context of fb-pages-food, closeness centrality shows how efficiently an entity (such as a chef, restaurant, or brand) can reach other entities in the network.

Logan Junior Chef has the highest closeness centrality of 0.3314 which indicates that this node has a strategic position in the network. This node is at the “center” of the network and likely has direct and indirect relationships, and plays an important role in the dissemination of information or influence in the network such as in spreading culinary trends, brand promotion, or collaboration.

D. Betweenness Centrality

Betweenness centrality measures how often a node acts as a link between two other nodes in the network. In the context of the fb-pages-food network, betweenness centrality provides insight into which nodes connect different communities and play a role in the flow of information or influence within the network.

Logan Junior Chef has the highest betweenness centrality of 0.3499, indicating that this entity is very important in connecting different communities within the network. As the node with the highest betweenness, it is likely that Logan Junior Chef acts as a “key link” between different groups, for example between chef communities or large restaurants and food brands.

E. PageRank

Based on the calculations performed, Logan Junior Chef has the highest PageRank value in the network with a value of 0.025287. This value indicates that Logan Junior Chef is considered one of the most important nodes in the network, most likely because it has many high-quality connections with other nodes.

For example, if Logan Junior Chef has many connections with chefs, famous restaurants, or big culinary brands, this will increase its PageRank. As such, Logan Junior Chef may have a great influence when it comes to spreading information or trends within the culinary community.

F. Katz Centrality

Katz Centrality is one of the metrics used to measure the influence or importance of a node (in this case, a character or page) in the network, based on the number of direct and indirect connections it has. The higher the Katz Centrality value, the greater the influence or importance of the node in the network.

Logan Junior Chef has the highest Katz Centrality value (0.076826), which indicates that this page has considerable influence in the analyzed network. This could mean that Logan Junior Chef has many direct and indirect connections with other nodes, making it more central in the network structure. Some factors that can increase the Katz Centrality value include collaboration, social media influence, or interaction with other pages.

$$C_i = \sum_{j=1}^{n} \alpha^{d(i,j)} A_{ij}$$

G. Eigenvector Centrality

Eigenvector Centrality measures the strength or influence of a node based on its connection quality. This means that nodes that are connected to nodes that have a high centrality value will get a higher centrality value as well.

Logan Junior Chef has the highest Eigenvector Centrality value (0.325752). This shows that Logan Junior Chef not only has many connections in the network, but is also connected to nodes that have great influence.

H. Epidemic

The concept of epidemic in the context of fb-pages-food social network data refers to the spread of information, influence, or activity within the network similar to the spread of infectious diseases in the population. By using a model such as SIR (Susceptible-Infected-Recovered), the results are Susceptible: 94, Infected: 12, Recovered: 514. Susceptible (94) means that 94 nodes (food pages) are still vulnerable to information or trends spreading on the network. These nodes have not been exposed or have not spread the information. Infected (12) means that only 12 nodes are currently “active” in spreading information or trends. Recovered (514) means that a total of 514 nodes have “recovered,” i.e. they were exposed to and spread the information before, but are no longer actively spreading it.

We utilized models such as the SIR (Susceptible - Infected-Recovered)framework to analyze the spread of information within the network. In this context:

- ❖ **Susceptible (S)**: Nodes that have not yet been exposed to the information.
- ❖ **Infected (I)**: Nodes actively spreading the information.
- ❖ **Recovered (R)**: Nodes that have been exposed and spread the information but are no longer actively propagating it.

Simulation Process:
The information spread process within the network is simulated using the following parameters:

- ❖ **Transmission rate (β)**: The probability of information spreading from one node to another.
- ❖ **Recovery rate (γ)**: The time required for a node to stop spreading the information.

This approach models how information propagates through interconnected nodes, providing insights into the dynamics of influence and the overall reach of the information within the network.

With only 12 active nodes, the food trend does not seem to have spread to its full potential. There are still far more vulnerable nodes than infected ones. Furthermore, most of the network (514 out of a total of 620 nodes) has already gone through the spreading stage. This suggests a particular trend may have peaked and started to decline in popularity.

I. Contagion

The selected Seed Nodes are 'Logan Junior Chef', 'McDonald's', 'David Chang', 'Eric Ripert', and 'Scott Conant'. These are the nodes that have the highest degree in the graph, which indicates that they are highly connected actors or entities in this social network. They serve as the starting point for information dissemination.

The total Activated Nodes of 84 shows that through the ICM simulation, information was successfully spread to 84 different nodes from the total network. This means that even though there are only 5 seed nodes that initiate the spread, the spreading effect is quite extensive, covering almost the entire network. This illustrates how the spreading effect can be widespread in a social network when nodes have strong connections.

There are many activated nodes beyond the seed nodes. For example “McDonald's Paraguay”, “José Andrés”, “Bobby Flay”, and “Noma” appear as part of the activated nodes. This shows that the information successfully spreads not only to nodes that are directly connected to the seed nodes, but also to nodes that have connections to those nodes.

Based on the simulation results, nodes such as “McDonald's” and “Logan Junior Chef” appear several times in the list of activated nodes. This indicates that they have a great influence on the spread of information, and the spread occurs faster due to the presence of nodes that have many connections.

J. Community Detection

Community detection on data such as the fb-pages-food dataset means identifying groups or clusters of Facebook pages that have stronger connections to each other compared to pages outside those groups. By applying community detection, we can find groups of pages that focus on specific topics or pages with many connections to different communities. In performing community detection on this dataset, the Louvain algorithm is used. Louvain's algorithm is a modularity optimization-based algorithm for community detection. Modularity is a measure used to calculate how well-defined communities are in a graph.

By using Louvain's algorithm, we found that from the fb-pages-food data, there are 20 communities (0-19). For example, nodes 58 (Chef Jamie Gwen) and 603 (Elizabeth Karmel) are in the same community. This shows that these nodes are closely related. The detected communities may represent certain sub-groups within the network, for example, community 1 may involve culinary personalities or chefs who focus on a particular type of food or collaboration. These results can also help in understanding patterns of connection or influence within the network, such as which figures or brands act as "hubs" (central nodes) between communities.

References

L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc Networks*, vol. 1, no. 3, pp. 215–239, 1978, doi: 10.1016/0378-8733(78)90021-7.

U. Brandes, "A faster algorithm for betweenness centrality*," *Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001, doi: 10.1080/0022250X.2001.9990249.

T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-Free Localization Schemes for Large Scale Sensor Networks," *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, pp. 81–95, 2003, doi: 10.1145/938985.938995.

H. Liu *et al.*, "Alterations of degree centrality and functional connectivity in classic trigeminal neuralgia," *Front Neurosci*, vol. 16, p. 1090462, Jan. 2023, doi: 10.3389/FNINS.2022.1090462/FULL.

T. Ahmad, X. J. Li, B. C. Seet, and J. C. Cano, "Social Network Analysis Based Localization Technique with Clustered Closeness Centrality for 3D Wireless Sensor Networks," *Electronics 2020, Vol. 9, Page 738*, vol. 9, no. 5, p. 738, Apr. 2020, doi: 10.3390/ELECTRONICS9050738.

M. Zhou, H. Jin, Q. Wu, H. Xie, and Q. Han, "Betweenness centrality-based community adaptive network representation for link prediction," *Applied Intelligence*, vol. 52, no. 4, pp. 3545–3558, Mar. 2022, doi: 10.1007/S10489-021-02633-7/TABLES/3.

A. Boodaghian Asl, J. Raghothama, A. Darwich, and S. Meijer, "USING PAGERANK AND SOCIAL NETWORK ANALYSIS TO SPECIFY MENTAL HEALTH FACTORS," *Proceedings of the Design Society*, vol. 1, pp. 3379–3388, 2021, doi: 10.1017/PDS.2021.599.

A. Bartal, N. Pliskin, and O. Tsur, "Local/Global contagion of viral/non-viral information: Analysis of contagion spread in online social networks," *PLoS One*, vol. 15, no. 4, p. e0230811, Apr. 2020, doi: 10.1371/JOURNAL.PONE.0230811.

N. O. Hodas and K. Lerman, "The Simple Rules of Social Contagion," *Scientific Reports 2014 4:1*, vol. 4, no. 1, pp. 1–7, Mar. 2014, doi: 10.1038/srep04343.

Ilyas MU, Radha H. Identifying influential nodes in online social networks using principal component centrality. In:
IEEE international conference on communications (ICC); 2011.
Ghosh R, Lerman K. Parameterized centrality metric for network analysis. Phys Rev E. 2011;83(6):066–118.
J. Zhan, S. Gurung, and S. P. K. Parsa, "Identification of top-K nodes in large networks using Katz centrality," *Journal of Big Data*, vol. 4, no. 16, pp. 1–19, 2017. DOI: 10.1186/s40537-017-0076-5.
S. Yin, J. Wu, and P. Song, “Optimal control by deep learning techniques and its applications on epidemic models,” *Journal of Mathematical Biology*, vol. 86, no. 36, pp. 1–26, Jan. 2023, doi: 10.1007/s00285-023-01873-0.