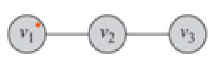
CMR INSTITUTE OF TECHNOLOGY

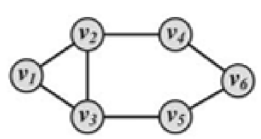
USN | | | | | | | | | | |

Internal Assessment Test II –December 2025

Sub:	**Social Media Analytics**							**Sub Code:**	**MMCA311F**
Date:	**18/12/25**	**Duration:**	**90 min's**	**Max Marks:**	**50**	**Sem:**	**III**	**Branch:**	**MCA**

Note : Answer FIVE FULL Question, choosing ONE full question from each Module

		MARKS	OBE	
	PART I		CO	RBT
1	Discuss the steps of dashboard creation in Tableau? **OR**	[10]	CO5	L2
2	Calculate the Eigen vector for the following graph. 	[10]	CO3	L4
3	**PART II** Explain Brute force clique identification algorithm **OR**	[10]	CO3	L2
4	Discuss the importance of visualizing data using Heat Maps with an example.	[10]	CO5	L2

5	**PART III** Discuss the group centrality methods in detail. **OR**	[10]	CO3	L2
6	Calculate the similarity values between node v2 and node v5 of the following graph 	[10]	CO3	L4
7	**PART IV** What are word clouds? Demonstrate it with a code snippet. **OR**	[10]	CO5	L2
8	Illustrate topic modeling using LDA.	[10]	CO3	L2
9	**PART V** Explain the ethical Considerations in Social Media Analytics? **OR**	[10]	CO5	L2
10	Explain the role of social media in Ad performance and customer engagement.	[10]	CO5	L2

1. Discuss How do you create dashboard using Tableau

1. Create a New Dashboard

A dashboard is created in Tableau in a manner similar to creating a new worksheet. At the bottom of the Tableau workbook, the user clicks on the New Dashboard icon. This opens a blank dashboard canvas where different visualizations (worksheets) and objects can be placed and arranged

2. Add Sheets to the Dashboard

Once the dashboard is created, existing worksheets (views) are added to it. From the Sheets list on the left side, the required views are dragged and dropped onto the dashboard area.

If a visualization needs to be replaced, the existing sheet in the dashboard can be selected, and another sheet can be swapped in using the Swap Sheets option

3. Add Interactivity to the Dashboard

Interactivity is a key strength of Tableau dashboards. A sheet can be enabled as a filter by selecting the Use as Filter option. This allows user selections in one visualization to dynamically filter data in other views within the dashboard.

Additionally, Tableau supports dashboard actions, such as filter actions, highlight actions, navigation actions, and URL actions, enabling richer user interaction and exploratory analysis

4. Add Dashboard Objects

In addition to worksheets, Tableau allows the inclusion of various dashboard objects to enhance layout, usability, and visual appeal. These objects are dragged from the Objects pane into the dashboard. Common objects include:

- Horizontal and Vertical Containers for structured layout and responsive resizing
- Text Objects for titles, headings, and descriptions
- Image Objects to add logos or visuals, with optional URL links
- Web Page Objects to embed web content
- Blank Objects to manage spacing and alignment
- Navigation Objects to move between dashboards or sheets
- Download Objects to export dashboards as PDF, PowerPoint, or images
- Extension Objects and Pulse Metric Objects to extend dashboard functionality

5. Set Options for Dashboard Objects

Each object added to the dashboard can be customized. By selecting an object and opening its shortcut menu, users can modify properties such as size, formatting, background, borders, and tooltips. These options help improve clarity, accessibility, and user experience

6. Work with Images, Navigation, and Download Objects

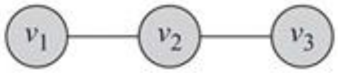
Tableau provides detailed configuration options for specific objects:

- Image Objects can either embed image files or link to web-based images. Linked images are preferred for large or animated images to improve performance. Images can also include clickable URLs and alternative text for accessibility.
- Navigation and Download Objects can be customized with text or images as buttons, tooltip descriptions, and formatting styles to clearly indicate their purpose, such as navigating to another dashboard or exporting data

7. Copy and Reuse Dashboard Objects

Tableau allows copying and pasting of dashboard objects within the same dashboard or across different dashboards and workbooks. However, certain elements such as sheets within dashboards, filters tied to specific sheets, and dashboard titles cannot be copied directly.

2. Compute eigen vector centrality for the given below graph



$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The adjacency matrix is

$$(A - \lambda I)\mathbf{C}_e = 0.$$

Assuming $\mathbf{C}_e = [u_1\ u_2\ u_3]^T$,

$$\begin{bmatrix} 0-\lambda & 1 & 0 \\ 1 & 0-\lambda & 1 \\ 0 & 1 & 0-\lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since $\mathbf{C}_e \neq [0\ 0\ 0]^T$, *the characteristic equation is*

$$det(A - \lambda I) = \begin{vmatrix} 0-\lambda & 1 & 0 \\ 1 & 0-\lambda & 1 \\ 0 & 1 & 0-\lambda \end{vmatrix} = 0,$$

or equivalently,

$$(-\lambda)(\lambda^2 - 1) - 1(-\lambda) = 2\lambda - \lambda^3 = \lambda(2 - \lambda^2) = 0.$$

So the eigenvalues are $(-\sqrt{2}, 0, +\sqrt{2})$. *We select the largest eigenvalue*
We compute the corresponding eigenvector:

$$\begin{bmatrix} 0-\sqrt{2} & 1 & 0 \\ 1 & 0-\sqrt{2} & 1 \\ 0 & 1 & 0-\sqrt{2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Assuming $\mathbf{C}_e$ *vector has norm 1, its solution is*

$$\mathbf{C}_e = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ \sqrt{2}/2 \\ 1/2 \end{bmatrix},$$

Node 2 is the most central node

3. Explain Brute force clique algorithm

Brute-force clique identification. The brute force method can find all maximal cliques in a graph. For each vertex vx, we try to find the maximal clique that contains node vx.

The algorithm starts with an empty stack of cliques. This stack is initialized with the node vx that is being analyzed (aclique of size 1). Then, from the stack, a clique is popped (C). The last node added to clique C. is selected (vlast). All the neighbors of vlast are added to the popped clique C sequentially, and if the new set of nodes creates a larger clique (i.e., the newly added nodeis connected to all of the other members), then the new clique is pushed back into the stack.

This procedure is followed until nodes can no longer be added. The brute-force algorithm becomes impractical for large networks.

For instance, for a complete graph of only 100 nodes, the algorithm will gen erate at least 299 − 1 different cliques starting from any node in the graph. The performance of the brute-force algorithm can be enhanced by prun ing specific nodes and edges. If the cliques being searched for are of size k or larger, we can simply assume that the clique, if found, should contain nodes that have degrees equal to or more than k−1. We can first prune all nodes (and edges connected to them) with degrees less than k − 1. Due to the power-law distribution of node degrees, many nodes exist with small degrees (1, 2, etc.). Hence, for a large enough k many nodes and edges will be pruned, which will reduce the computation drastically. This pruning works for both directed and undirected graphs

Require: Adjacency Matrix A, Vertex v_x
return Maximal Clique C containing v_x
CliqueStack = $\{\{v_x\}\}$, Processed = $\{\}$;
while CliqueStack not empty **do**
  C=pop(CliqueStack); push(Processed,C);
  v_{last} = Last node added to C;
  $N(v_{last}) = \{v_i | A_{v_{last}, v_i} = 1\}$.
  for all $v_{temp} \in N(v_{last})$ **do**
    if C $\bigcup \{v_{temp}\}$ is a clique **then**
      push(CliqueStack, C $\bigcup \{v_{temp}\}$);
    end if
  end for
end while
Return the largest clique from Processed

4. Discuss the importance of visualizing data using Heat maps with an example.

A heat map is a data visualization technique that uses color intensity to represent the magnitude of values in a dataset.

Importance: Quick Pattern Recognition, Easy Comparison Across Dimensions, Efficient Analysis of Large Datasets, Identification of Trends and Anomalies, Better Decision-Making Support.

Ex: import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

Generate random data

data = np.random.randint(1, 100, (10, 10))

Create a heatmap

sns.heatmap(data)

plt.show()

5. Discuss the group centrality methods in details.

Group Degree Centrality

Group degree centrality is defined as the number of nodes from outside

the group that are connected to group members. Formally,

$$C_d^{\text{group}}(S) = |\{v_i \in V - S | v_i \text{ is connected to } v_j \in S\}|.$$

Similar to degree centrality, we can define connections in terms of out-degrees or in-degrees in directed graphs. We can also normalize this value. In the best case, group members are connected to all other nonmembers.

Group Betweenness Centrality

Similar to betweeness centrality, we can define group betweenness centrality as

$$C_b^{\text{group}}(S) = \sum_{s \neq t, s \notin S, t \notin S} \frac{\sigma_{st}(S)}{\sigma_{st}},$$

where σst(S) denotes the number of shortest paths between s and t that pass through members of S.

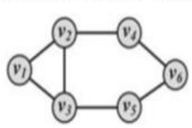
Group Closeness Centrality

Closeness centrality for groups can be defined as

$$C_c^{\text{group}}(S) = \frac{1}{\bar{l}_S^{\text{group}}},$$

6.

Calculate the similarity values between node v2 and node v5 of the following graph



$$\sigma_{\text{Jaccard}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{|\{v_1, v_3, v_4, v_6\}|} = 0.25,$$

$$\sigma_{\text{Cosine}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{\sqrt{|\{v_1, v_3, v_4\}||\{v_3, v_6\}|}} = 0.40.$$

7. What are word clouds? Demonstrate it with a code snippet.

A word cloud is a text visualization technique where words are displayed in varying sizes and colors based on their frequency or importance in a text corpus.

 Larger words appear more frequently or have higher importance.

 Smaller words appear less frequently.

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
# Sample text
text = """
Data visualization is an interdisciplinary field that deals
with the graphic representation of data.
It is a particularly efficient way of communicating when the
data is numerous as for example a time series."""
# Generate the word cloud
wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(text)
# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')  # Remove axes
plt.show()
```

8. Illustrate topic modeling using LDA

Topic modeling is a type of statistical modeling used to uncover the abstract topics that occur in a collection of documents.

- It is a form of unsupervised learning
- It relies on the co-occurrence patterns of words within the documents to discover latent topics.
- Unsupervised topic modeling method

LDA:

- Discovers hidden themes from text
- Represents documents as mixtures of topics
- Represents topics as mixtures of words

LDA operates on the following principles:

• Document-Topic Distribution: Each document is represented as a distribution over topics. This means that a document can belong to multiple topics, each with a certain probability.

• Topic-Word Distribution: Each topic is represented as a distribution over words. This implies that a topic is defined by a set of words, each associated with a probability of appearing in that topic.

• Generative Process: LDA uses a generative process to create documents. This involves:

• Choosing a distribution of topics for each document.

• For each word in the document, selecting a topic based on the topic distribution.

• Generating the word from the selected topic's word distribution.

Steps of LDA involves

• Preprocess text

• Extract hashtags

• Frequency & co-occurrence analysis

• LDA topic modeling

• Interpret and visualize results

EX:

```
!pip install pandas sklearn
import pandas as pd
import re
from collections import Counter
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation
from google.colab import files
uploaded = files.upload()

#Example: dataset.csv  → use your filename
```

```python
filename = list(uploaded.keys())[0]
df = pd.read_csv(filename)

print("Dataset loaded successfully!")
df.head()
def clean_raw_text(t):
    """Decode strings like b'...' into readable text."""
    if isinstance(t, str) and t.startswith("b'"):
        try:
            return eval(t).decode("utf-8", errors="ignore")
        except:
            return t
    return t

df['clean_text'] = df['tweet'].apply(clean_raw_text)

def extract_hashtags(text):
    if isinstance(text, str):
        return re.findall(r"#\w+", text.lower())
    return []

df['hashtags'] = df['clean_text'].apply(extract_hashtags)
all_tags = [tag for tags in df['hashtags'] for tag in tags]
tag_counts = Counter(all_tags)

print("\nTop 20 Hashtags:")
for tag, count in tag_counts.most_common(20):
    print(tag, ":", count)

print("\nRunning LDA Topic Modeling...")

vectorizer = CountVectorizer(stop_words='english',
                             max_features=5000)
X = vectorizer.fit_transform(df['clean_text'].astype(str))

lda = LatentDirichletAllocation(
    n_components=5,  # You can change no. of topics
    random_state=42
)
lda.fit(X)

feature_names = vectorizer.get_feature_names_out()

def print_topics(model, feature_names, n_top_words=15):
    for idx, topic in enumerate(model.components_):
```

```
        print(f"\nTopic {idx+1} Keywords:")
        print(", ".join([feature_names[i] for i in topic.argsort()[-
n_top_words:]]))

print_topics(lda, feature_names)
```

9. Explain the ethical considerations in Social Media Analytics.

Social media analytics involves collecting and analyzing vast amounts of user-generated data to derive insights. While this provides significant benefits, it also raises serious privacy and ethical challenges. Addressing these challenges responsibly is essential to protect user rights and maintain trust.

1. Unauthorized Data Collection

Data is often collected from social media platforms without users' explicit knowledge or permission, especially through scraping techniques.

Organizations should collect data only through authorized APIs and follow platform policies.

Data collection should be lawful, justified, and limited to stated purposes.

2. Lack of Informed Consent

Users are frequently unaware of how their data is being analyzed or used, as consent mechanisms are unclear or hidden in lengthy terms and conditions.

Adopt informed and explicit consent, clearly explaining what data is collected, why it is used, and how long it will be retained.

3. Exposure of Personally Identifiable Information (PII)

Social media data may contain sensitive information such as names, usernames, locations, images, or contact details.

Apply data anonymization and pseudonymization techniques to remove or mask personal identifiers before analysis.

4. Data Re-identification Risk

Even anonymized datasets can sometimes be re-identified when combined with other datasets.

Use strong anonymization methods, limit data sharing, and avoid combining datasets in ways that increase re-identification risks.

5. Surveillance and User Profiling

Continuous monitoring of user activity can lead to intrusive surveillance and unfair profiling.Follow data minimization principles by collecting only necessary data and avoiding excessive tracking of individuals.

6. Data Breaches and Security Risks

Large-scale social media datasets are vulnerable to cyberattacks and data leaks.Ensure secure data storage and access control using encryption, authentication, and role-based access mechanisms.

7. Bias and Ethical Misuse

Biased analytics or AI models can lead to discrimination, manipulation, or unfair targeting of specific groups.

Implement ethical AI and bias mitigation by auditing models, using diverse datasets, and ensuring fairness in analysis.

8. Transparency and Accountability

Lack of clarity about how analytics systems operate reduces trust and accountability.

Maintain transparency and accountability by documenting data practices, model decisions, and enabling audits.

9. Compliance with Privacy Regulations

Failure to comply with legal frameworks can result in legal and ethical consequences. Ensure compliance with privacy regulations such as GDPR, CCPA, and national data protection laws, respecting user rights like data access and deletion.

10. Explain the role of social media in AD performance and customer engagement

Social media has become a powerful platform for digital marketing, enabling organizations to promote their products and services effectively while engaging customers in meaningful ways. By leveraging data analytics, interactivity, and targeted communication, social media significantly enhances both advertisement performance and customer engagement.

1. Precise Audience Targeting

Social media platforms allow advertisers to target audiences based on demographics, interests, location, behavior, and online activity. This ensures that advertisements reach the most relevant users, increasing click-through rates and conversion rates. Precise targeting reduces wastage of advertising budgets and improves campaign effectiveness.

2. Cost-Effective Advertising

Compared to traditional advertising media such as television and print, social media advertising is more cost-effective. Businesses can run campaigns with flexible budgets using pay-per-click or pay-per-impression models. Even small businesses can reach large audiences at lower costs, achieving better return on investment (ROI).

3. Real-Time Performance Tracking

Social media platforms provide real-time analytics on advertisement performance, including impressions, reach, clicks, and engagement levels. Advertisers can continuously monitor campaign performance and make immediate adjustments to improve results. This real-time feedback enables quick decision-making and better campaign control.

4. Data-Driven Optimization

Social media analytics tools help analyze user interactions and campaign outcomes. Based on data insights, advertisers can optimize ad creatives, messaging, audience segments, and posting schedules. Techniques such as A/B testing further enhance the effectiveness of advertisements through data-driven improvements.

5. Personalized Content Delivery

Social media enables personalized advertising by delivering content tailored to individual user preferences and behaviors. Personalized ads and recommendations increase relevance, leading to higher engagement and better customer experience. AI-driven personalization strengthens the connection between brands and users.

6. Influencer and Community Marketing

Influencer marketing plays a significant role in enhancing advertisement reach and credibility. Influencers help brands connect with their followers in an authentic way. Additionally, brand communities on social media encourage interaction, discussions, and peer influence, further boosting engagement.

7. Improved Customer Feedback and Support

Social media platforms allow customers to provide instant feedback, reviews, and queries.

Brands can respond quickly to customer concerns, improving satisfaction and trust. This two way communication strengthens customer relationships and enhances brand image.

8. Brand Awareness and Loyalty

Consistent advertising and engagement on social media increase brand visibility and recall. Interactive campaigns, personalized communication, and prompt responses foster trust and loyalty among customers. Over time, this leads to long-term customer relationships and brand advocacy.