STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI - 600 086
(For candidates admitted from the academic year 2015-16 & thereafter)

SUBJECT CODE : 15MT/PC/RT24

M. Sc. DEGREE EXAMINATION, APRIL 2018
BRANCH I – MATHEMATICS
SECOND SEMESTER

COURSE : CORE
PAPER : RESEARCH METHODS AND TOOLS
TIME : 3 HOURS **MAX. MARKS : 100**

THEORY

Answer any two questions: **(2 X 10 = 20)**

1. What is a research problem? Briefly describe how to formulate the research problem.
2. Discuss about data collection.
3. Explain format of a report.

PRACTICAL

SUBJECT CODE : 15MT/PC/RT24

Answer any four without omitting any section **(4 X 20 = 80)**

SECTION – A

1. Create .tex file for the following document.

Summarizing and Understanding Large Graphs

Abstract

How can we succinctly describe a million-node graph with a few simple sentences? Given a large graph, how can we find its most 'important' structures, so that we can summarize it and easily visualize it? How can we measure the 'importance' of a set of discovered subgraphs in a large graph? Starting with the observation that real graphs often consist of stars, bipartite cores, cliques and chains, our main idea is to find the most succinct description of a graph in these 'vocabulary' terms. To this end, we first mine candidate subgraphs using one or more graph partitioning algorithms. Next, we identify the optimal summarization using the Minimum Description Length (MDL) principle, picking only those subgraphs from the candidates that together yield the best lossless compression of the graph—or, equivalently, that most succinctly describe its adjacency matrix.

1 Introduction

Given a large graph, such as the Facebook social network, what can we say about its structure? As most real graphs, the edge distribution will likely follow a power law [21], but apart from that, is it random? If not, how can we efficiently and in simple terms summarize which parts of the graph stand out, and how? The focus of this paper is exactly finding short summaries for large graphs, in order to gain a better understanding of their characteristics.

Why not apply one of the many community detection, clustering or graph-cut algorithms that abound in the literature [11, 15, 29, 38, 48], and summarize the graph in terms of its communities? The answer is that these algorithms do not quite serve our goal. Typically they detect numerous communities without explicit ordering, so a principled selection procedure of the most "important" subgraphs is still needed. In addition to that, these methods merely return the discovered communities, without characterizing them (e.g., clique, star), and, thus, do not help the user gain further insights in the properties of the graph.

PROBLEM 1. (GRAPH SUMMARIZATION - INFORMAL)

- **Given**: *a graph*
- **Find**: *a set of possibly overlapping subgraphs*
- *to most succinctly describe the given graph, i.e., explain as many of its edges in as simple possible terms,*
- *in a* **scalable** *way, ideally linear on the number of edges.*

and our contributions can be summarized as:

1. **Problem Formulation:** We show how to formalize the intuitive concept of graph understanding using principled, information theoretic arguments.
2. **Effective and Scalable Algorithm:** We design VOG which is near-linear on the number of edges.
3. **Experiments on Real Graphs:** We empirically evaluate VOG on several real, public graphs spanning up to millions of edges. VOG spots interesting patterns like 'edit wars' in the Wikipedia graphs (Fig. 1).

In this section we describe the first contribution, the MDL formulation of graph summarization. To enhance readability, we list the most frequently used symbols in Table 1.

In general, the Minimum Description Length principle (MDL) [52], is a practical version of Kolmogorov Complexity [39], which embraces the slogan *Induction by Compression*. For MDL, this can be roughly described as follows. Given a set of models $\mathcal{M}$, the best model $M \in \mathcal{M}$ minimizes

$$L(M) + L(\mathcal{D} \mid M) \ ,$$

where

- $L(M)$ is the length in bits of the description of M, and
- $L(\mathcal{D} \mid M)$ is the length, in bits, of the description of the data when encoded using the information in M.

2. Create .tex file for the following document.

CHAPTER 2 59th ANNUAL REPORT 2012-2013

INDIA TEA IN THE INTERNATIONAL PERSPECTIVE

Broad over view of the Global and Indian Tea Scenarios:

Global Tea Scenario

More than 30 countries spread over all the continents except North America with wide range of agro-climatic conditions between 42'N (Georgia) and 35'S latitude (Argentina) grow tea. The estimated global production in 2012 was 4625 million kg. With the world consumptic around 4440 million kg, the global production an absorption remained finely balanced.

Major tea producing and exporting countries ar China, India, Kenya and Sri Lanka and they accour for 78% and 71% of world production and expor respectively. (Table-1)

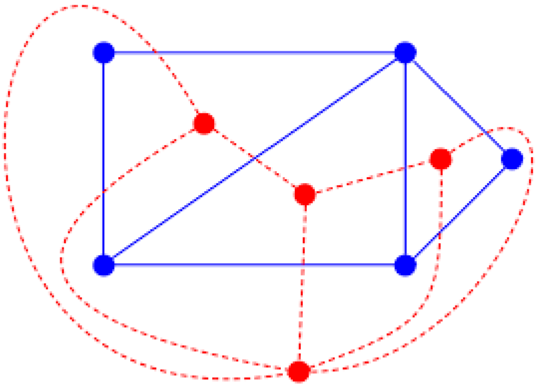
Table-1.

Production and Export share of major producing and exporting countries

Country	2012			
	Production		Export	
	Million Kg	%share	Million Kg	%share
China	1789.75	39	321.79	18
India	1126.33	24	208.26	12
Kenya	369.56	8	430.21	24
Sri Lanka	328.4	7	306.04	17
Others	1010.59	22	510.21	29
World Total	**4624.63**	100	**1776.51**	100

Section – B

3. Draw the following figure using layers in Flash.



4. Create a flash movie using tweening and interacting for a single piece of paper dropping through the air

Section – C

5. Using Mathcad solve the following.

a. Find AB and BA if $A=\begin{bmatrix}22 & 34 & 45\\ 12 & 0 & 8\\ 1 & 23 & 55\end{bmatrix}$ and $B=\begin{bmatrix}2 & 4 & 5\\ 1 & 0 & 3\\ 1 & 2 & 15\end{bmatrix}$

b. Solve the equation $\begin{vmatrix}x & 23 & 55\\ 23 & x & 55\\ 23 & 55 & x\end{vmatrix}=0$

c. Find the sum squares of integers between 1 & 300 which are divisible by 3 and 5.

d. Differentiate $x^3 e^x \sin x$.

e. Evaluate $\int_1^2 \log x\,dx$.

(4+4+4+4+4)

6. Using Mathcad do the following:

a. Draw the curve $y=3\sin(x) + e^x$

b. Trace the curve $x^2 + y^3 = 50$

c. plot the surface $f(x, y) = \sin(x) - \cos(y)$, $-5 <= x,y <= 5$

(6 + 6 +8)

⅄⅄⅄⅄⅄⅄⅄⅄⅄⅄⅄⅄