Reet Barik
PhD Qualifying Examination

The Examination Committee:
Jana Doppa (Chair), Larry Holder, Ananth Kalyanaraman
Written report due: 23:59 PST, Tuesday, August 11, 2020
Oral exam: 3-5pm PST, Friday, August 14, 2020 via Zoom

Summer 2020

Instructions

Your qualifying examination, assigned on **Friday, August 6, 2020**, consists of two parts: (i) a written report and (ii) an oral presentation. Please email your written report as a PDF document to Prof. Jana Doppa (jana.doppa@wsu.edu), Chair of the Examination Committee. The deadlines mentioned above are to be strictly adhere to. Be sure to follow all the guidelines for the two components of this examination as described below. Direct any questions you have concerning this examination to the chair of the examination committee.

Guidelines for the Written Report

Congratulations! You are about to commence work on the written part of your Qualifying Examination for the doctoral degree in Computer Science from Washington State University. As such, you need to be aware of all rules that pertain to this examination. The examination procedure is listed in the Graduate Program Policies and Procedures manual with which you should be familiar. The purpose of this document is to inform you of what is expected of you regarding the structure and content of the written component of the exam. The goal of the QE is to assess your ability to do self-directed research in your chosen area. *Research* includes the abilities to (a) evaluate existing work, (b) set it in the context of other work in your area, and (c) suggest novel, logical, and productive ways to extend it. A good written QE exhibits all of these traits. Furthermore, all written QE's should observe the following guidelines:

1. Your work must be your own. You may not communicate with any other person regarding your work on this examination unless authorized by the QE Committee.
2. The Committee is expecting a document of between 10 and 20 pages, double-spaced, in a 12-point font. However, you may exceed 20 pages, as needed, to ensure your document adequately addresses the examination.
3. You are expected to cite other work, both online and printed, but in doing so:
   (a) You should never copy the other work verbatim. This especially includes cutting and pasting online information. (Committee members know how to use search engines, too!)

(b) Direct quotes are often useful in the arts and humanities, but are used to a more limited extent in computer science. In those instances in which you do need to quote a source directly, you must place it in double quotes (“”). If it is longer than two lines, you must indent it in order set it off from the rest of your text. Note that you must cite both the source and *page number* of any direct quote.

(c) Most of your discussion of other work should paraphrase, not quote, the part(s) of it relevant to your topic. (Space and time constraints should limit you here in any case.)

(d) An exception to this may be made in the use of figures, i.e., diagrams and images, formulas, and tables, that cannot be easily paraphrased, but their source should always be cited in their captions and/or context. There should be no more than a few such uses in your write-up.

(e) You must always place a citation immediately before each instance of a description of other work, e.g., According to Knuth [1], ... You don't need to do this for every sentence: If you start a paragraph this way, readers will assume that the whole paragraph applies to the cited work. On the other hand, if you discuss the same paper on 3 or 4 different pages, each page should include the citation.

(f) The particular style of the citation is up to you. Examples include number-only ([1]), author last name and date in parentheses ((Knuth, 1968)), author-and-number (Knuth [1]), or abbreviated-author-year ([Knu68]).

As with any examination (including take-home) at WSU, you are subject to WSU Standards of Conduct for Students (a copy of which is in your Student Handbook), particularly the section on Academic Integrity. At the discretion of your QE Committee, violations of these standards may be grounds for your failing the Examination and may in addition be reported to the Graduate Studies Committee, the WSU Office of Student Conduct and lead to further action on the School, Graduate School, or University level.

Guidelines for the Oral Presentation

Prepare an oral presentation that summarizes your written report using either PowerPoint or PDF slides. Plan your presentation to last no longer than 45 minutes. Following your presentation, committee members will ask questions regarding your report and the areas it addresses; questions may also be asked during your presentation if there is a need for clarification. Please bring your own laptop to use for your presentation.

Exam Questions

Parallelizing graph operations remains a challenging task because of the difficulties associated with data dependencies and irregular data movement. Many techniques have been proposed in the past to efficiently parallelize graph operations.

In a recent paper [1], Balaji and Lucia present an approach to speedup graph operations on shared memory (multithreaded) machines using a combination of two techniques, namely, data duplication and vertex reordering.

For this QE, you will read the above paper and any other related paper that you deem appropriate to answer the following questions:

Q1) Present a thorough critical evaluation of the paper [1]. Your critique should include: a clear statement of the problem and objectives, the main approach ideas and how they address the key objectives of the problem, and a discussion of the merits of the proposed approach including strengths, weaknesses, and any tradeoffs. You are encouraged to put the contributions in context by comparing them against closely related work published either before this paper or since.

Q2) In the second part of the QE response, you will answer a related question but focused on vertex reordering. Since the goal of vertex reordering is mainly to improve locality during computation (and thereby performance), MinHashing (briefly reviewed in [2]), which is a type of locality sensitive hashing, becomes relevant. For this question, please go through some relevant literature on MinHashing to get some background, and subsequently propose ideas/ways in which one can possibly adapt the technique for vertex reordering in graphs. Focus on the main ideas for this part of the question.

The papers necessary to answer this QE are attached. If you are unable to locate any other paper that you deem necessary in any of the WSU libraries or online, please contact the Chair of the QE committee. You are advised to search for papers in Google Scholar prior to approaching the chair for assistance. Good luck!

References

[1] V. Balaji and B. Lucia. "Combining data duplication and graph reordering to accelerate parallel graph processing," *In Proc. 28th International Symposium on High-Performance Parallel and Distributed Computing*, pp. 133-144, 2019.
DOI: 10.1145/3307681.3326609.

[2] E. Cohen, "Min-hash sketches," Technical Report, 2016. http://www.cohenwang.com/edith/Surveys/minhash.pdf.