DESIGNING HEURISTIC ALGORITHMS FOR LARGE SEMANTIC GRAPHS: AN EXPLORATION IN APPROXIMATING SUBGRAPH MATCHES FOR IMPERFECT FRAGMENTS

By

Joshua Greenman

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: COMPUTER SCIENCE

Approved:

Dr. Mark Goldberg, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York

December 2010
(For Graduation December 2010)

ABSTRACT

Graphs have become an essential means for storing data regarding networks of all varieties. As a result, the demand to create increasingly efficient & portable solutions to network flow and graph theory problems is ever-growing. We introduce methods of providing approximate solutions to the well-known subgraph isomorphism problem, utilizing the concept of semantically typed data to define the structure of our graphs. Our two algorithms, Frequency-Vector and Type-Matrix, perform structural indexing and analysis to assist the process of locating hidden adversarial networks within a database graph from a given fragment. Since the fragment is not guaranteed to contain a precise match to any subgraph in the database, we must make approximations based on known relationships to locate valuable results.