| **REPORT DOCUMENTATION PAGE** | *Form Approved* *OMB No. 0704-0188* |
|---|---|



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 28-02-2011 | Final | Feb 2007 to Nov 2010 |

**4. TITLE AND SUBTITLE**

Social Networks Analysis: Classification, Evaluation, and Methodologies

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

FA9550-07-1-0134

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Eunice E. Santos

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Virginia Polytechnic Institute & State University, Blacksburg, VA 24061
University of Texas, El Paso, TX 79912

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Office of Scientific Research, Arlington, VA

**10. SPONSOR/MONITOR'S ACRONYM(S)**

AFOSR

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

AFRL-OSR-VA-TR-2012-0153

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

A distribution

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

We proposed the Social Network Analysis: Classification, Evaluation and Methodology (SNA-CEM) framework. SNA-CEM consists of Methodology, Evaluation and Classification components, each encapsulating critical aspects of the framework. The Methodology component deals with mathematically representing various SNA methodologies. Evaluation component consists of performance techniques and metrics to measure the utility of SNA methodologies. The Classification component uses the measures from the Evaluation component and representations from the Methodology component to group methods into categories, based on their similarity with respect to utility and performance. The utility measures used for evaluating methodologies include solution quality and time performance. We also focus on large-scale network size and dynamic changes in networks and research new capabilities in performing social networks analysis utilizing parallel and distributed processing.

**15. SUBJECT TERMS**

social networks analysis, graph theory, parallel and distributed processing, performance analysis

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Eunice E. Santos |
| | | | | | 19b. TELEPHONE NUMBER *(Include area code)* |
| | | | | | 915 747 5170 |

## INSTRUCTIONS FOR COMPLETING SF 298

**1. REPORT DATE.** Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

**2. REPORT TYPE.** State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

**3. DATES COVERED.** Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

**4. TITLE.** Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

**5a. CONTRACT NUMBER.** Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

**5b. GRANT NUMBER.** Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

**5c. PROGRAM ELEMENT NUMBER.** Enter all program element numbers as they appear in the report, e.g. 61101A.

**5d. PROJECT NUMBER.** Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

**5e. TASK NUMBER.** Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

**5f. WORK UNIT NUMBER.** Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

**6. AUTHOR(S).** Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES).** Self-explanatory.

**8. PERFORMING ORGANIZATION REPORT NUMBER.** Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES).** Enter the name and address of the organization(s) financially responsible for and monitoring the work.

**10. SPONSOR/MONITOR'S ACRONYM(S).** Enter, if available, e.g. BRL, ARDEC, NADC.

**11. SPONSOR/MONITOR'S REPORT NUMBER(S).** Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

**12. DISTRIBUTION/AVAILABILITY STATEMENT.** Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

**13. SUPPLEMENTARY NOTES.** Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

**14. ABSTRACT.** A brief (approximately 200 words) factual summary of the most significant information.

**15. SUBJECT TERMS.** Key words or phrases identifying major concepts in the report.

**16. SECURITY CLASSIFICATION.** Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

**17. LIMITATION OF ABSTRACT.** This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

**Social Networks Analysis: Classification, Evaluation and Methodologies**

**PI:** Eunice E. Santos
**AFOSR Grant No:** FA9550-07-1-0134
**AFOSR PM:** Terrence Lyons

# 1 Introduction

With large amounts of social data sets being readily available through web based social media and other sources, computational solutions are required for effective social modeling and data analyses. Social network is a powerful paradigm to represent, visualize, interpret and analyze information. Social Network Analysis (SNA) employs graph theoretic methodologies to mathematically define, analyze and quantify relevant metrics in social networks allowing for their interpretation and classification. Research in SNA has led to the emergence of several methodologies that have come to be widely used. Measures such as centrality, connectivity, degree and clique sizes have become standard in SNA.

However, large network sizes and dynamism continue to be important issues in SNA. As more people use online social networking apps and with the emergence of mobile computing apps, the social networks that have to be processed continue to grow. Additionally, real time social information is available leading to issues of dynamism.

Although there has been interesting research in SNA, it has been scattered and narrowly applied. There is a need for an overarching framework that provides a common representation for the different methodologies, making it easy to identify their similarities and differences. This will be helpful in designing new algorithms for SNA and understanding, a priori, their performance and utility. The framework should also take into account other critical aspects such as performance evaluation and methodology classification.

In pursuant of this goal, we proposed the Social Network Analysis: Classification, Evaluation and Methodology (SNA-CEM) framework. SNA-CEM consists of Methodology, Evaluation and Classification components, each encapsulating the critical aspects of the framework. Methodology component deals with mathematically representing various SNA methodologies. Evaluation component consists of performance techniques and metrics to measure the utility of SNA methodologies. The Classification component uses the measures from the Evaluation component and representations from the Methodology component to group methods into categories, based on their similarity with respect to utility and performance. In our project we have focused on methodologies that have a recursive structure and can be partitioned in a parallel/distributed environment. The utility measures used for evaluating methodologies are solution quality and time performance.

because that each processor only focuses on a small problem (the local subgraph) while the serial algorithm deals with the original large network, which introduces additional overheads such as reading discontinuous memories.

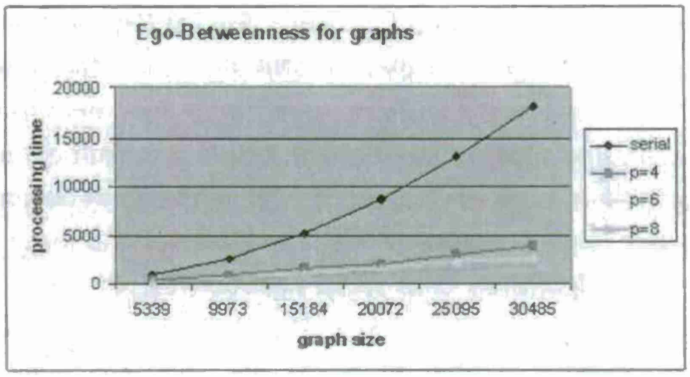


Figure 1: Ego-Betweenness centrality for graph with average degree=8 (Density II) [1]

In order to validate the anywhere aspects of the ego-betweenness algorithm, we looked at processing time for adopting 64 random changes in the graphs. Although the absolute value of time cost is an important factor to evaluate the system performance, graph size and density affect time cost. It may be more useful to understand algorithm performance with respect to graph size and density. Each relative cost is calculated as: $\mu = \max(r, c) / C$, where $\mu$ is the relative cost, *r* is the time cost for adopting random changes, *c* is the time cost for adopting max degree changes, C is the time cost for calculating each node's ego-betweenness centrality for static graphs by using the serial algorithm. From the experimental results, we saw that the relative cost decreases (even though the absolute time cost for adopting dynamic changes actually grows as graph size increases). This indicates that when the graph size becomes larger and larger, the portion of affected obtained results becomes smaller and smaller. The maximum relative cost for adopting one edge change is about 0.055%. Using theoretical and experimental analysis, we have demonstrated that our methodology for ego-betweenness centrality measurement can efficiently handle graph's dynamism.

### 3.1.2 Maximum Clique Enumeration

A clique of a graph is a sub-graph which has an edge between every pairs of nodes. A maximum clique cannot be contained in another clique. Identifying maximum cliques is an NP-hard problem. Our methodology generates the maximal cliques by coming with incrementally large cliques. This has a natural anytime property that can be exploited in the IA phase. We also developed anywhere approaches to deal with two types of dynamism: edge addition and edge removal.

To validate our methodology, we compare its performance with a serial algorithm based on Zhang's Algorithm. We study the performance for graphs from size 5000 to 30000 with various densities and no maximal cliques with size larger than 3. For brevity, we focus on the analysis for graphs of density II (average degree of 8), with the experimental results in Figure 2. Figure 2 shows that time for finding all maximal cliques increases with graph size. We also see that as the number of processors used in our system increases, the time cost for solving the problem decreases. Also, our parallel approach can solve the maximal clique enumeration problem faster than the typical serial algorithm.

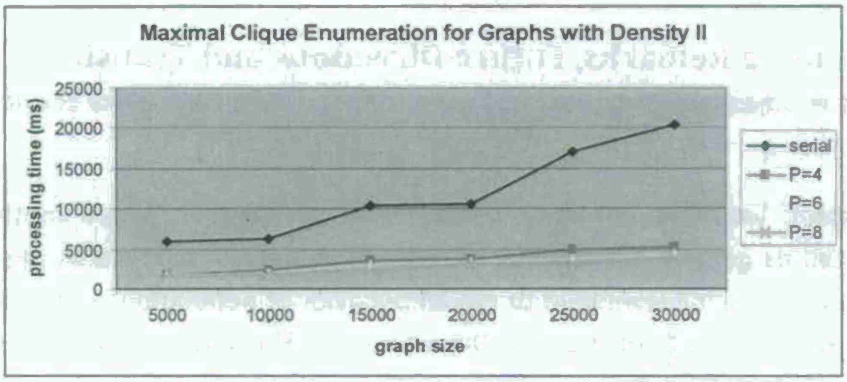


**Figure 2: Time costs for enumerating maximal cliques in graphs with average node degree=8[4]**

| S | *5000* | *10000* | *15000* | *20000* | *25000* | *30000* |
|---|---|---|---|---|---|---|
| $RC_4$ | 3.00% | 3.47% | 2.20% | 1.46% | 1.59% | 1.71% |
| $RC_8$ | 1.38% | 2.00% | 1.22% | 1.36% | 0.69% | 0.92% |
| $RC_{16}$ | 2.08% | 1.49% | 0.91% | 1.21% | 0.50% | 0.57% |

**Figure 3: Relative cost for adopting one dynamic edge change for maximal clique S-Size of the graph, $RC_x$ – Relative cost for graphs with average degree is x [4]**

We tested the anywhere approach for maximal clique enumeration on two sets of 8 dynamic edge changes: random edge changes and max degree edge changes. We measured the time cost of our system to handle each edge change. For each set, we take the average value as the time cost for adopting one dynamic change. Figure 3 shows that our anywhere approach for maximal clique enumeration can effectively handle dynamic graphs. The relative cost for adopting a random edge change is less than 3.5%.

## 3.2 Real World Scenarios

In accordance with our goal 3, we have applied the anytime-anywhere algorithms to real world scenarios. This was done as part of our transition to other projects dealing with socio-cultural behavioral models. The Culturally Infused Social Networks (CISN)[5] framework, which was developed as part of a project titled "A Framework for Adversarial Social Networks" funded by Defense Threat Reduction Agency (DTRA)[6], models and analyzes complex social processes by incorporating fine grained socio-cultural information onto individual nodes in social networks. We deployed anytime-anywhere algorithms for measuring closeness centrality in modeling gang violence in Haiti. Gangs in Haiti had become a problem as they operate with impunity in many areas. They gained the support of the people by providing basic services to the populace. Social networks representing the interactions between residents of a town in Haiti, and cultural fragments representing their ideology, were generated to model the scenario. The anytime-anywhere methodology provided the capability to extend CISN to large and dynamic graphs in the Haiti scenario.

## 3 Concluding Remarks, Future Directions and Transition

The goals and objectives of this project were met and discussed in the previous sections of this report.

In this project, we have not only validated the anytime-anywhere methodology but also demonstrated its general applicability to real world scenarios. The SNA design methods from SNA-CEM are also being leveraged to model behavior of populations in cross-border epidemics [7], as part of a project funded by the Department of Homeland Security (DHS). Understanding cross-border immigration during epidemics will help border security and border health agencies to be better prepared.

The anytime-anywhere framework for SNA is also leveraged to model and analyze social processes in network centric systems in the context of Network Centric Operations/Network Centric Warfare (NCO/NCW), as part of a project funded under the Army High Performance Computing Research Center (AHPCRC) initiative. The social relations between human actors in the network are key factors in understanding decision making and situational awareness in NCO/NCW. As part of the project, parallel and distributed SNA algorithms will be developed to analyze these processes in very large and dynamic social networks.

## Publications


[1] E. E. Santos, L. Pan, D. Arendt, and M. Pittkin, "An effective anytime anywhere parallel approach for centrality measurement in social network analysis," in *IEEE International Conference on Systems, Man and Cybernetics, SMC '06*, vol. 6, 2006, pp. 4693 –4698.

[2] E. E. Santos, L. Pan, D. Arendt, H. Xia, and M. Pittkin, "An anytime anywhere approach for computing all pairs shortest paths for social network analysis," *Integrated Design and Process Technology*, 2006.

[3] E. E. Santos, L. Pan, and D. Arendt, "An anytime-anywhere methodology for social networks analysis," in *Sunbelt XXVIII, International Sunbelt Social Network Conference*, 2008.

[4] L. Pan and E. E. Santos, "An anytime-anywhere approach for maximal clique enumeration in social network analysis," in *Proceedings of the IEEE International Conference on System, Man, and Cybernetics*, 2008.

[5] E. E. Santos, E. Santos Jr, L. Pan, and J. T. Wilkinson, "Culturally infused social network analysis," in *Proceedings of IC-AI Conference*, 2008, pp. 449–455.

[6] E. E. Santos, E. Santos Jr, J. T. Wilkinson, J. Korah, K. Kim, D. Li, and F. Yu, "Modeling complex social scenarios using culturally infused social networks," Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, 2011.

[7] E. E. Santos, E. Santos Jr., J. Korah, J. E. Thompson, K. Kim, R. George, Q. Gu, J. Jurmain, S. Subramanian, and J. T. Wilkinson, "Intent-Driven Behavioral Modeling During Cross-Border Epidemics," Proceedings of the Third IEEE International Conference on Social Computing, 2011.