REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
01-09-2022	Final Report	1-Jan-2019 - 31-Dec-2021

4. TITLE AND SUBTITLE
Final Report: Models and Algorithms for Higher Order Network Inference

5a. CONTRACT NUMBER
W911NF-19-1-0031

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Stanford University
3160 Porter Drive
Suite 100
Stanford, CA 94304 -8445

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

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

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
73348-MI-YIP.12

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Johan Ugander
UU	UU	UU	UU		19b. TELEPHONE NUMBER 605-725-0550

RPPR Final Report
as of 08-Sep-2022

Agency Code: 21XD

Proposal Number: 73348MIYIP **Agreement Number: W911NF-19-1-0031**
INVESTIGATOR(S):

Name: Ph.D. Johan Ugander
Email: jugander@stanford.edu
Phone Number: 6057250550
Principal: Y

Organization: **Stanford University**
Address: 3160 Porter Drive, Stanford, CA 943048445
Country: USA
DUNS Number: 009214214 EIN: 941156365

Report Date: 31-Mar-2022 Date Received: 01-Sep-2022
Final Report for Period Beginning 01-Jan-2019 and Ending 31-Dec-2021
Title: Models and Algorithms for Higher Order Network Inference
Begin Performance Period: 01-Jan-2019 **End Performance Period:** 31-Dec-2021
Report Term: 0-Other

Submitted By: Ph.D. Johan Ugander Email: jugander@stanford.edu
Phone: (605) 725-0550

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 10 **STEM Participants:** 12

Major Goals: The work under this YIP was focused on leveraging recent work by the PI that connects higher-order choice modeling -- approaches that go beyond “independence assumptions” between the alternatives on offer, sometimes called context effects -- and network science. The goal has been to develop new modeling tools that enrich both camps. The main applications are in modeling competition networks (match-ups) as well as social network formation, with further applications to complex decision-making domains where individuals report set-wise or ranked preferences (ranked choice voting, school choice).

Accomplishments: See uploaded PDF.

Training Opportunities: Nine (9) PhD students, two (2) MS students, and one (1) BS student participated in research under this award. Many of these students gave presentations on work output from the award, and were given direct mentorship on presentation skills from the PI. The PI organized weekly group meetings throughout the year for all students that dedicate approximately 50% of the agenda to career and professional development topics. During the pandemic, students were encouraged take advantage of the ease with which many workshops and conferences could be accessed at very low cost.

RPPR Final Report
as of 08-Sep-2022

Results Dissemination: The work products of the grant were the focus of departmental colloquia/seminars at Cornell, University of Michigan, Texas A&M, as well as an invite keynote address at the 2021 NeurIPS Workshop on Human and Machine Decisions.

Supported PhD students gave many talks on their work, both at conferences were results were presented in proceedings, but also broader dissemination as is suitable for multi-disciplinary work. Among such broader talks, Jan Overgoor gave talks at the IC2S2 Conference in June 2020 and a spotlight talk at the SIAM Network Science Workshop in July 2020. Many such talks were given during the “zoom era” when dissemination was made much cheaper by the lack of need to cover travel or lodging costs.

Beyond talks, the work under this ARO YIP has diligently included code releases with each paper that contains any relevant code. The code repository for the "choosing to grow" project, which encompasses the code for the WWW19 and KDD20 papers, has attracted a healthy amount of interest and attention (many emails, but also 23 stars on Github as of August 2022), in addition to the attention the main work has attracted (47 citations for the first paper, 8 citations for the second, as of August 2022).

PI Ugander also regularly gave Stanford campus talks to a general audience, aimed at undergraduates interested in data science and computational social science, as well as one summer school talk (Bay Area Summer Institute in Computational Social Science). These talks all featured the work support by the ARO YIP to some extent.

Honors and Awards: 1) Outstanding Problem-Solution Paper Award, AAAI Conference on Web and Social Media (ICWSM) 2021
2) Reviewer Award, Neural Information Processing Systems (NeurIPS) 2020
3) Faculty of the Year Award, Institute for Computational Mathematics and Engineering (ICME), Stanford University, 2020
4) Best Paper Award, AAAI Conference on Web and Social Media (ICWSM) 2020
5) Outstanding PC Award, ACM WSDM 2019 and 2020

Protocol Activity Status:

Technology Transfer: I remain interested in visiting ARL, but pandemic circumstances have made it impractical.

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)
Participant: Amel Awadelkarim
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Stephen Ragain
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Arjun Seshadri
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Jan Overgoor
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Samir Khan
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Jackson Eilers
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: George Supaniratisai
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Faculty
Participant: Austin Benson
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yatong Chen
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Alex Chin
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Rahul Makhijani

RPPR Final Report
as of 08-Sep-2022

Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kristen Altenburger
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Hao Yin
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Network Science
Publication Identifier Type: DOI Publication Identifier: 10.1017/nws.2020.20
Volume: Issue: First Page #: 1
Date Submitted: 8/28/20 12:00AM Date Published: 6/1/20 7:00AM
Publication Location:
Article Title: Measuring Directed Triadic Closure with Closure Coefficients
Authors: Hao Yin, Austin Benson, Johan Ugander
Keywords: triadic closure, configuration model
Abstract: Recent work studying clustering in undirected graphs has drawn attention to the distinction between measures of clustering that focus on the "center" node of a triangle vs. measures that focus on the "initiator," a distinction with considerable consequences. Existing measures in directed graphs, meanwhile, have all been center-focused. In this work, we propose a family of directed closure coefficients that measure the frequency of triadic closure in directed graphs from the perspective of the node initiating closure. We observe dramatic empirical variation in these coefficients on real-world networks, even in cases when the induced directed triangles are isomorphic. To understand this phenomenon, we examine the theoretical behavior of our closure coefficients under a directed configuration model.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

CONFERENCE PAPERS:

RPPR Final Report
as of 08-Sep-2022

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The World Wide Web Conference
Date Received: 30-Aug-2019 Conference Date: 13-May-2019 Date Published:
Conference Location: San Francisco, CA, USA
Paper Title: Choosing to Grow a Graph: Modeling Network Formation as Discrete Choice
Authors: Jan Overgoor, Austin Benson, Johan Ugander
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The World Wide Web Conference
Date Received: 30-Aug-2019 Conference Date: 13-May-2019 Date Published: 13-May-2019
Conference Location: San Francisco, CA, USA
Paper Title: Decoupled Smoothing on Graphs
Authors: Alex Chin, Yatong Chen, Kristen Altenburger, Johan Ugander
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Conference on Machine Learning
Date Received: 30-Aug-2019 Conference Date: 09-Jun-2019 Date Published: 09-Jun-2019
Conference Location: Long Beach, CA
Paper Title: Discovering Context Effects from Raw Choice Data
Authors: Arjun Seshadri, Alex Peysakhovich, Johan Ugander
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: Proceedings of the 2019 ACM Conference on Economics and Computation
Date Received: 28-Aug-2020 Conference Date: 24-Jun-2019 Date Published: 24-Jun-2019
Conference Location: Phoenix, AZ
Paper Title: Fundamental Limits of Testing the Independence of Irrelevant Alternatives in Discrete Choice
Authors: Arjun Seshadri, Johan Ugander
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The World Wide Web Conference
Date Received: 30-Aug-2019 Conference Date: 13-May-2019 Date Published: 13-May-2019
Conference Location: San Francisco, CA, USA
Paper Title: Parametric Models for Intransitivity in Pairwise Rankings
Authors: Rahul Makhijani, Johan Ugander
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining
Date Received: 28-Aug-2020 Conference Date: 24-Aug-2020 Date Published: 24-Aug-2020
Conference Location: Virtual
Paper Title: Scaling Choice Models of Relational Social Data
Authors: Jan Overgoor, George Pakapol Supaniratisai, Johan Ugander
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 08-Sep-2022

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: Neural Information Processing Systems 2020
Date Received: 28-Aug-2021 Conference Date: 09-Dec-2020 Date Published: 09-Dec-2020
Conference Location: Virtual
Paper Title: Learning Rich Rankings
Authors: Arjun Seshadri, Stephen Ragain, Johan Ugander
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining
Date Received: Conference Date: 14-Aug-2021 Date Published: 14-Aug-2021
Conference Location: Virtual Event Singapore
Paper Title: Choice Set Confounding in Discrete Choice
Authors: Kiran Tomlinson, Johan Ugander, Austin Benson
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: Proceedings of the International AAAI Conference on Web and Social Media
Date Received: 28-Aug-2021 Conference Date: 07-Jun-2021 Date Published: 07-Jun-2021
Conference Location: Virtual
Paper Title: Which Node Attribute Prediction Task Are We Solving? Within-Network, Across-Network, or Across-Layer Tasks
Authors: Kristen Altenburger, Johan Ugander
Acknowledged Federal Support: **Y**

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Johan Ugander
Signature Date: 9/1/22 3:57AM

ARO YIP: Models and Algorithms for Higher Order Network Inference
1/1/2019–12/31/2021

Johan Ugander*

September 1, 2022

1 Major Goals

The work under this YIP was focused on leveraging recent work by the PI that connects higher-order choice modeling – approaches that go beyond "independence assumptions" between the alternatives on offer, sometimes called context effects – and network science. The goal has been to develop new modeling tools that enrich both camps. The main applications are in modeling competition networks (match-ups) as well as social network formation, with further applications to complex decision-making domains where individuals report set-wise or ranked preferences (ranked choice voting, school choice).

2 Accomplishments

2.1 Thrust 1: Higher order models of competition

This project has grown into a very successful research program. A high-impact paper from the first year (Overgoor et al., "Choosing to grow a graph", WWW2019) was followed by a more detailed/technical follow-up ("Scaling Choice Models of Relational Social Data", KDD2020) in the second year. Well-received talks were given at SIAM Network Science and The International Conference on Computational Social Science. This work has helped cement choice-based methods as an important new approach to network science. As an indicator of the early import of that work, a 2019 commentary in Nature Communication, discussing the state of statistical methods for evaluating scale-free distributions in network data, described the paper, saying: "Inferring general network growth processes is a challenging and relatively unexplored topic [12,13]. Ref. [14] (Overgoor et al.) probably represents the state of the art." This work under the award has had a significant impact on the toolkit and momentum of network modeling efforts.

*Department of Management Science and Engineering, Stanford University. E–mail: jugander@stanford.edu.

Work on advanced choice modeling, the engine of the above network formation application, was the subject of a series of impactful papers published at ICML 2019, EC 2019, and NeurIPS 2020. The work on contextual choices examines the choices people make from sets as resulting from pairwise interactions with the other available items, effectively a "pairwise push and pull" model of choice. That work, appearing at ICML 2019, was then extended to models of how people (and non-people) produce rankings, in work appearing at NeurIPS 2020. When runners compete in a race, what mechanisms produce the rankings? Is it just the "intrinsic" qualities of the athletes? Under a push and pull model, it matters what other runners are in the field. And indeed, we find that a pairwise push and pull model of ranking is more accurate than traditional random utility models. Overall, this line of work has also been extremely successful, and was the subject of an invite keynote talk at the 2021 NeurIPS Workshop on Human and Machine Decisions. The annual reports (and more significantly, the papers) contain more detail on the merits of the technical contributions in this work.

The tail end of the effort under this award has explored applications of these advanced ranking models to new domains. Work has been under way to examine how these models of ranking data can be used to improve models of how parents rank schools when submitting preference lists in school choice systems, as well as high-fidelity models of empirical ranked choice voting behavior and other ranked data domains (notably "ranked social networks", where subjects rank their contacts). The former application has seen promising preliminary results, and we now carrying forward the work under this YIP award to work with a team at Stanford that is developing a new school choice system for the San Francisco Unified School District (SFUSD). I am actively seeking funding to continue that work on richer models of school choice preferences, and have secured an award from the NSF to continue the work on the structure of ranked social networks.

The way in which this thrust has produced high-impact work that has shaped the direction of the field makes me very proud. Work under this award has also connected me with the above exciting domains, unfolding as the best research projects do, leading to an explosion of downstream ideas. I am grateful for how this YIP award has helped launch of research program. As a small note, it funded a significant portion of the research that went into my tenure file, and I was granted tenure at Stanford in June 2022.

2.2 Thrust 2: Higher order recursive definitions on networks

This thrust was part of the approved YIP proposal, which itself was pared down from a larger proposal that had been written for a five-year project, but was revised down to a three-year YIP. At the time of that revision (reducing the five year plan of work to three) it felt reasonable to attempt to complete half the goals of this thrust, as described in the revised statement of work. The goal was to take two recent papers on recursive higher-order graph machine learning methods, SCRank and REV2, and work out a generalization of the two methods where SCRank and REV2 were special cases. The problem with the literature here is: these methods compute properties recursively in a manner that raises many questions about the stability of the computation, the convergence guarantees, and son on. The stated goal was to come up with a generalization and prove general (convergence, etc.) properties for the generalization, thus providing a general theory of recursive higher-order models.

The work on this effort fizzled, and no research papers have been written. I spent half a day visiting Google Research in 2019 and talking to the SCRank authors about their work, which had excited me at the time that I wrote the YIP. They said that they were unexcited about the method and hadn't found it that useful, so cautioned me away from going too far with it. At the same time, REV2 did not have sustained impact beyond the interest in it that I was seeing in 2018 when I wrote my proposal. While this work had been planned for the third year of the award, between the lukewarm two-years-later interest in the two main algorithms I was looking to study, plus the pandemic, it felt more fruitful to focus on research that connected to the (very, very successful) first thrust. If other recursive algorithms for computing higher-order network statistics catch on, I still have some ideas for how the analysis might work, and I may end up pursuing this work some future day.

2.3 Opportunity Thrusts

The grant also supported a number of "opportunity thrusts," research efforts that pursued emerging ideas that intersected the goals of the originally proposed work. Yin et al. published a well-received journal paper in Network Science that made a significant contribution to how studies of triadic closure can be mathematically operationalized in directed graph settings. The answer is a lot more subtle than I thought it would be when the work began.

As a second example, a paper on the theoretical foundations of hypothesis testing when choices are well-described by utility models vs. have higher-order structure. The idea in this paper was not part of the proposal, but the question arose on the heels of the 2019 ICML paper: how powerful are the natural hypothesis tests for if these higher order choice models are needed? The paper was executed briskly from start to finish over the first six weeks of 2019 by PhD student Arjun Seshadri, pointing out that testing whether an agent is behaving according to a rational (utility maximizing) model of choice is very statistically hard. Put simply, there is only one way to be rational and (exponentially) many ways to be irrational, so a hypothesis test for rationality is very hard to construct (many directions of deviation makes for a lowpowered test). This work was accepted at EC with glowing reviews (the technical machinery is very innovative). The ideas in this work raise serious questions about the many studies of context effects, given that it is statistically very hard to rigorously reject that the deviations form rationality aren't just due to some statistical fluke.