Opportunity Title: Scalable Pattern Discovery Within Graph-Structured Data
Opportunity Reference Code: ICPD-2021-57

Organization	Office of the Director of National Intelligence (ODNI)
Reference Code	ICPD-2021-57
How to Apply	**Create and release your Profile on Zintellect** – Postdoctoral applicants must create an account and complete a profile in the on-line application system. **Please note: your resume/CV may not exceed 2 pages.**

OAK RIDGE INSTITUTE
FOR SCIENCE AND EDUCATION

ORISE GO

The ORISE GO mobile app helps you stay engaged, connected and informed during your ORISE experience – from application, to offer, through your appointment and even as an ORISE alum!

Visit ORISE GO ↗

GET IT ON Google Play

Download on the App Store

Complete your application – Enter the rest of the information required for the IC Postdoc Program Research Opportunity. The application itself contains detailed instructions for each one of these components: availability, citizenship, transcripts, dissertation abstract, publication and presentation plan, and information about your Research Advisor co-applicant.

Additional information about the IC Postdoctoral Research Fellowship Program is available on the program website located at: https://orise.orau.gov/icpostdoc/index.html.

If you have questions, send an email to ICPostdoc@orau.org. Please include the reference code for this opportunity in your email.
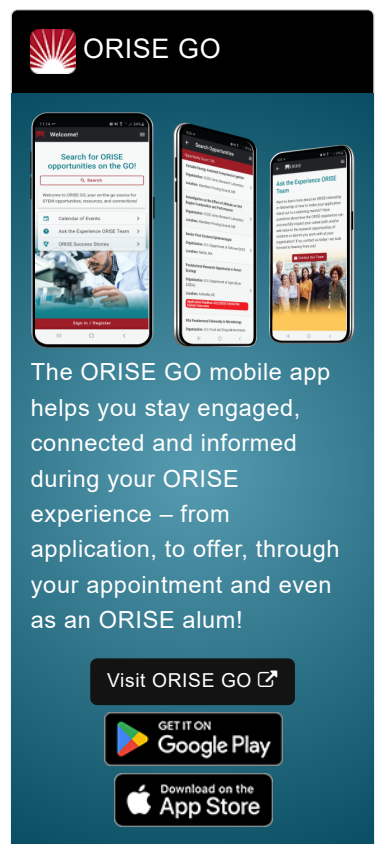
Application Deadline	2/26/2021 6:00:00 PM Eastern Time Zone
Description	**Research Topic Description, including Problem Statement:**

The Internet of Things, software vulnerability assessments, ontologies, and social networks, among other areas, can be modelled as graph-structured data and analyzed using complex graph theory and network analysis. Graph theory analysis can identify reoccurring events and community structures. The structure of a network is key to its function, and of particular interest is the discovery of repeating patterns across a single large complex network without prior information. A pattern in a network may represent a reoccurring event or a key functional block. The expectation is that the presence of patterns will not be limited to communities or functional blocks and could form in any part of the graph structure. Methods exist to detect the occurrences of a pattern within a network, such as subgraph isomorphism matching, however these require prior information of the pattern structure. The key to this problem statement is the unsupervised discovery of the patterns, using either machine learning (ML) or non-ML methods.

As the scale and complexity of these networks are increasing rapidly, there is a computational challenge to pattern discovery. Pattern discovery in small and static networks may be done manually, but it is becoming progressively less feasible and quality of analysis will be low. There are limited algorithms available to discover small patterns within moderate-sized heterogeneous networks, such as motif-detection and the graph-based data-mining system Subdue.[1] However, these do not scale to larger or more complex scenarios. Steps in the direction of unsupervised pattern discovery are apparent through research into Graph Neural Networks among other methods.

Complex network datasets are available; however, none have been identified to have existing labelled repeating patterns. These include but are not limited to:

Opportunity Title: Scalable Pattern Discovery Within Graph-Structured Data
Opportunity Reference Code: ICPD-2021-57

- https://www.complex-networks.net/datasets.html.
- https://snap.stanford.edu/data/.
- https://github.com/gephi/gephi/wiki/Datasets.
- http://konect.uni-koblenz.de/networks/.

Reference:

[1] Ketkar, N, Holder, L and Cook, D., "Subdue: compression-based frequent pattern discovery in graph data". 2005, ACM.

Example Approaches:

A scalable method for the discovery of large repeating patterns within a single complex network graph is required. Innovative solutions, using either ML or non-ML methods, are sought to develop the ability to identify repeating patterns across a single large and complex network. Our current focus is the discovery of large patterns (50-100 nodes) in complex (>100,000 nodes) static networks (i.e. not changing over time) with no prior knowledge of their structure, and their exact matches across the network. The key constraint is that there will be no prior knowledge of the patterns. Rather than searching for instances of a known pattern, the requirement is to discover the patterns that exist within the network data. Proposals may wish to consider how the method may apply or extend to:

- Identify reoccurring events across temporal datasets.
- Identify non-exact matches of pattern occurrences across the network.
- Create labelled datasets for method validation.

Proposals should include:

- Research assessment of algorithms and methods against the following criteria: ?
  - Scalability of method for networks with >100,000 nodes.
  - Ability to discover patterns in any part of the graph structure.
  - Ability to discover patterns and their exact matches.
- Implementation of most appropriate method(s).
- Methods for validation against open-source data; this may include identifying or creating a labelled dataset.
- Methods for improvement based on test results.

The potential to adapt the method for non-exact pattern matching and pattern discovery within a temporal dataset may also be included with the consideration for future work. However, the key focus should be on static networks.

Technical collaboration will be provided during the research process. Classified data will be used to test and assess methods, and feedback will be provided for further development.

Relevance to the Intelligence Community:

Pattern discovery using graph theory has several identified use cases across the Intelligence Community (IC) including:

Opportunity Title: Scalable Pattern Discovery Within Graph-Structured Data
Opportunity Reference Code: ICPD-2021-57

- The discovery of repeating functional blocks for analysis and understanding.
- Identification of structures that may inform risk areas, for example in software development and social networks.
- The analysis of large semantic/knowledge graphs to support deductions and predictions.
- Identification of interlinked events in communication networks.
- Assurance of deep learning models.

Successful pattern detection will provide insight into areas of understanding where human capability is a limitation. Further advancement of this work will enable improved analysis of temporal data. This may lead to identification of repeating or interconnected events across a network which cannot be identified by human analysis.

This work will contribute to a large number of capabilities across the IC.

Key Words: Large Complex Networks, Network Analysis, Scalable, Pattern Discovery, Graph Theory

Qualifications

Postdoc Eligibility

- U.S. citizens only
- Ph.D. in a relevant field must be completed before beginning the appointment and within five years of the application deadline
- Proposal must be associated with an accredited U.S. university, college, or U.S. government laboratory
- Eligible candidates may only receive one award from the IC Postdoctoral Research Fellowship Program

Research Advisor Eligibility

- Must be an employee of an accredited U.S. university, college or U.S. government laboratory
- Are not required to be U.S. citizens

Eligibility Requirements

- **Citizenship:** U.S. Citizen Only
- **Degree:** Doctoral Degree.
- **Discipline(s):**
  - **Chemistry and Materials Sciences** (12 👁)
  - **Communications and Graphics Design** (2 👁)
  - **Computer, Information, and Data Sciences** (17 👁)
  - **Earth and Geosciences** (21 👁)
  - **Engineering** (27 👁)
  - **Environmental and Marine Sciences** (14 👁)
  - **Life Health and Medical Sciences** (45 👁)
  - **Mathematics and Statistics** (10 👁)
  - **Other Non-Science & Engineering** (2 👁)
  - **Physics** (16 👁)
  - **Science & Engineering-related** (1 👁)
  - **Social and Behavioral Sciences** (27 👁)