Computational Network Science

NSF CSSI PI Meeting Feb 14 2020

Madhav Marathe

Email: marathe@virginia.edu

Phone: 540 808 6220

Network Simulation Science and Advanced Computing Division (NSSAC)

Biocomplexity Institute and Initiative

& Dept. of Computer Science

University of Virginia

Project Title: CINES (pronounced “science") *: A Scalable Cyberinfrastructure for Sustained Innovation in Network Engineering and Science*

Grant details: NSF Grant No.: OAC-1916805

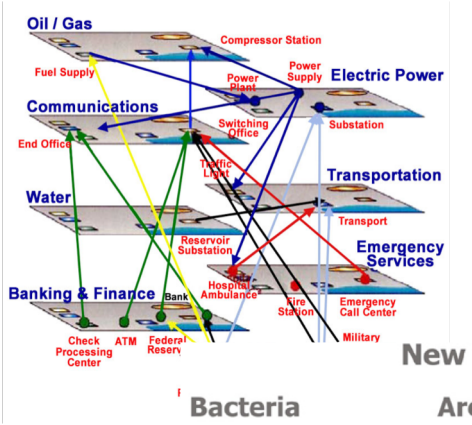
Time period*:* 5-year project started in October 2019

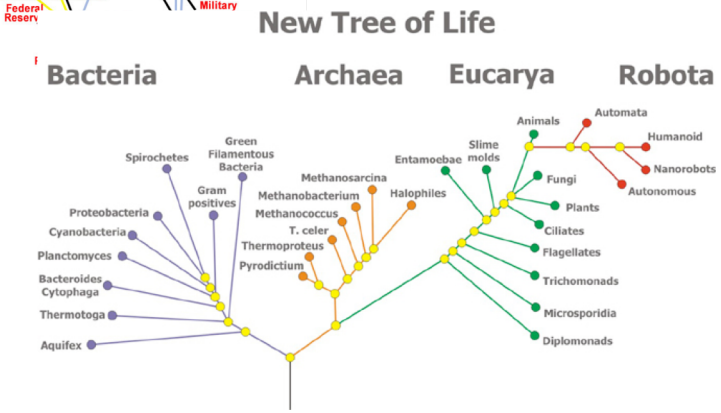
Context

Networks are pervasive

Networks are ubiquitous

- Networks are ubiquitous and are a part of our common vocabulary.
- Network science and engineering (NSE) has emerged as a formal field over the last 30 years & seen an explosive growth.
- Ideas from NSE have played a central role in the formation of companies, e.g. Akamai, Twitter, Google, Facebook, Uber, Lyft and LinkedIn.
- NSE concepts used to address fundamental problems in diverse fields, e.g., epidemiology and marketing and are now part of most university curricula.





A new multi-disciplinary field has emerged: differing viewpoints and needs

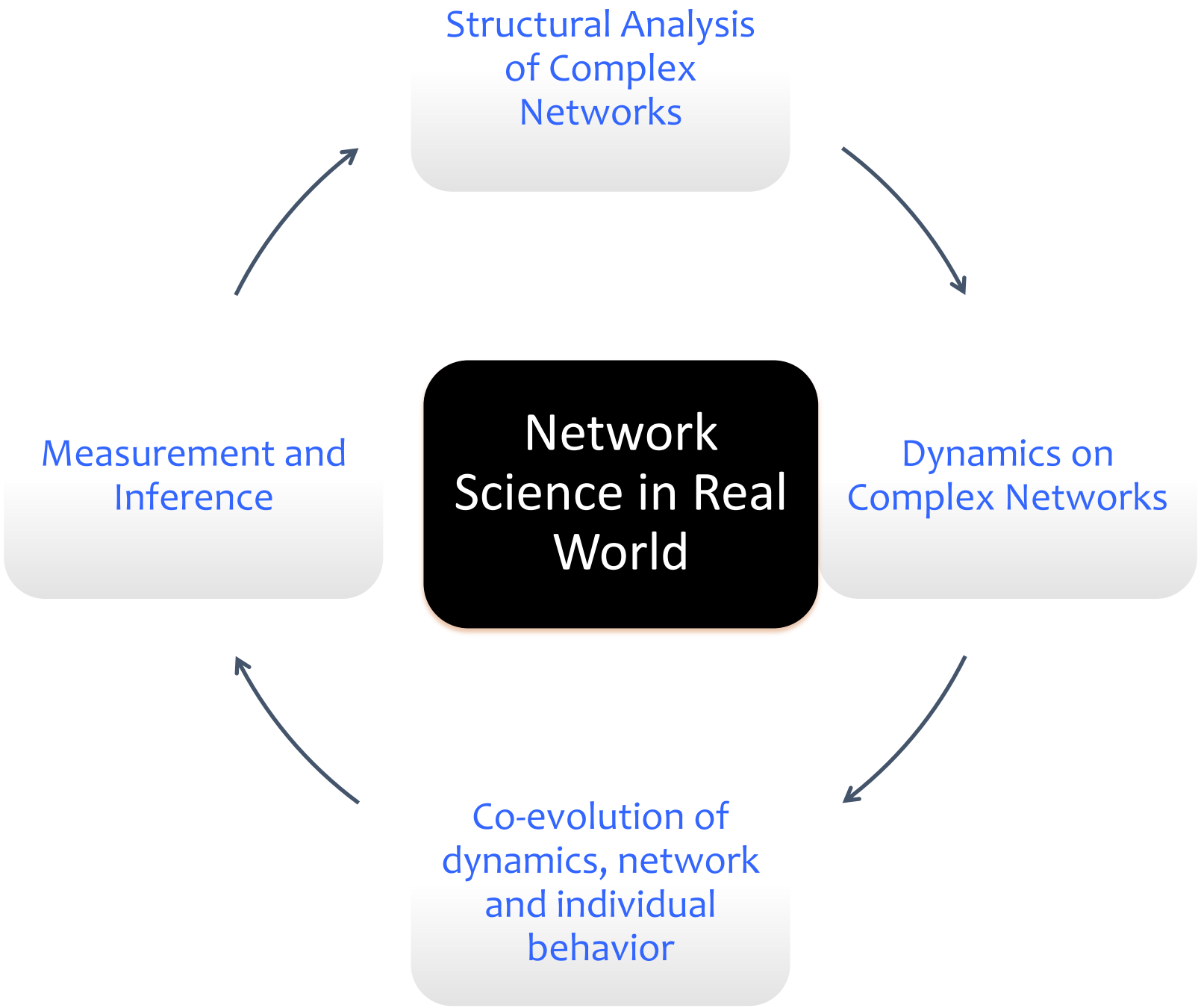
Opportunity, challenges

Engineers	Physicists and Mathematicians	Social Scientists and Economists	Computer Scientists
Understand how infrastructure networks work Design and control of these networks	Understand phase transitions and structural properties of networks	Network formation, Roles of nodes relationships	Understand phase transitions and structural properties of networks

- Resources are largely dispersed and stand-alone (in silos of isolated tools), of small scale, or home-grown for personal use.

Develop Pervasive computing technology to deliver Network Science technology to domain specialists and others who are not necessarily computing experts

Components of Network science and engineering

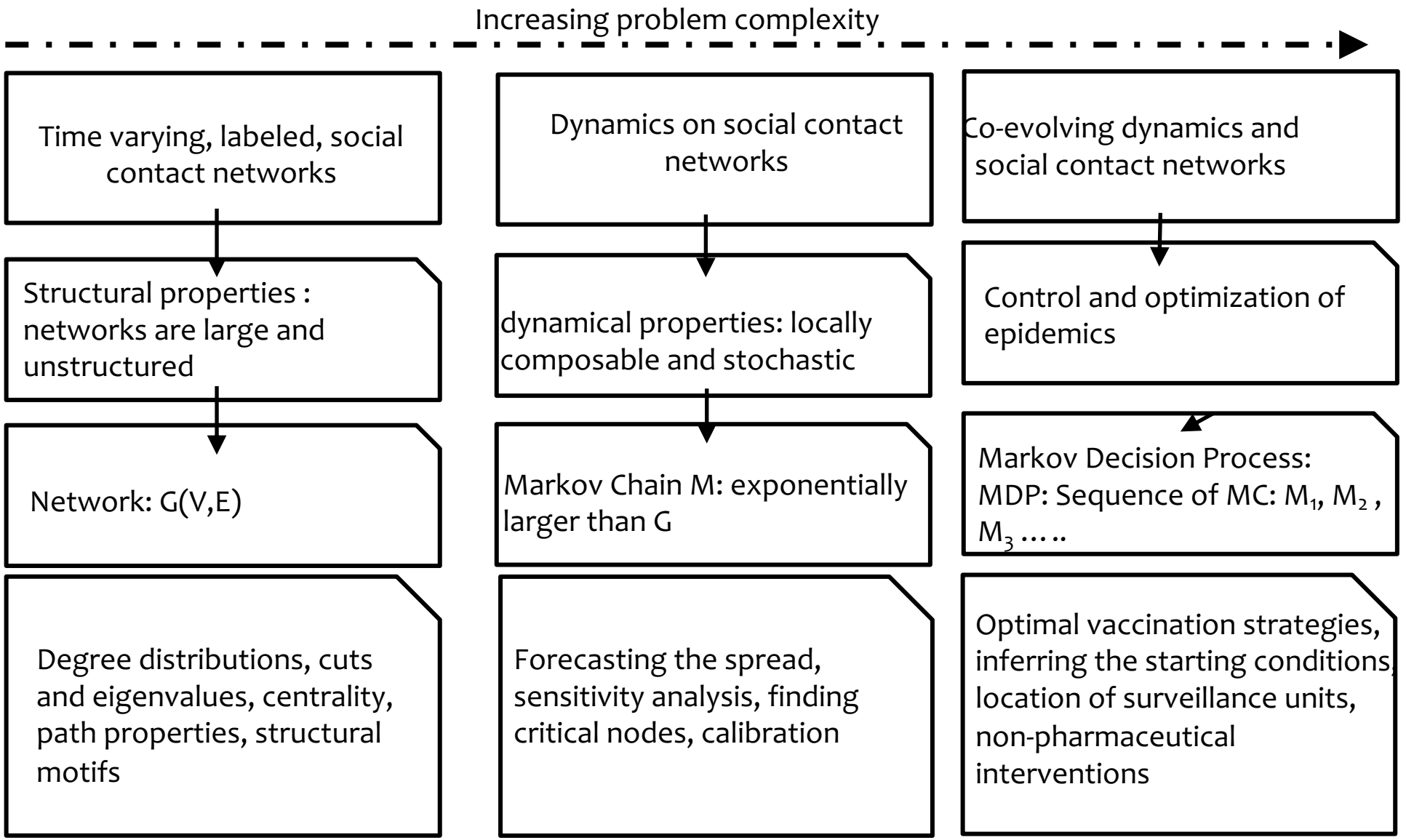


Research Challenges in NSE

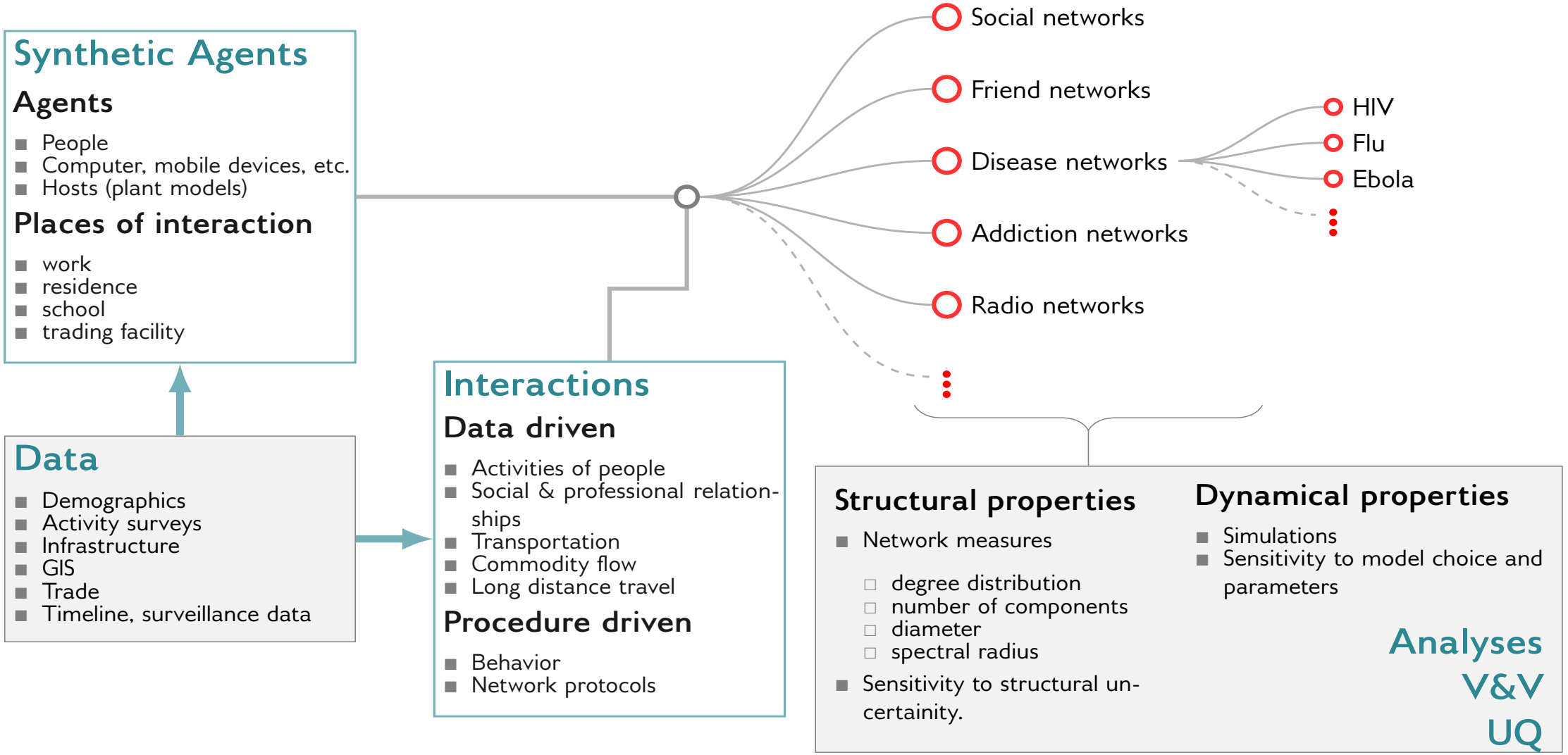
(National academy report)

1. **Dynamics**: Better understanding between structure and function
2. **Modeling** and analysis of large networks: Tools, abstractions, approximations
3. **Synthesis**: Design and synthesis of networks
4. **Rigor**: Increasing level of rigor and mathematical structure
5. **Abstracting** common concepts across fields
6. **Experiments**: Better experiments and measurements of network structure
7. **Robustness and Security**

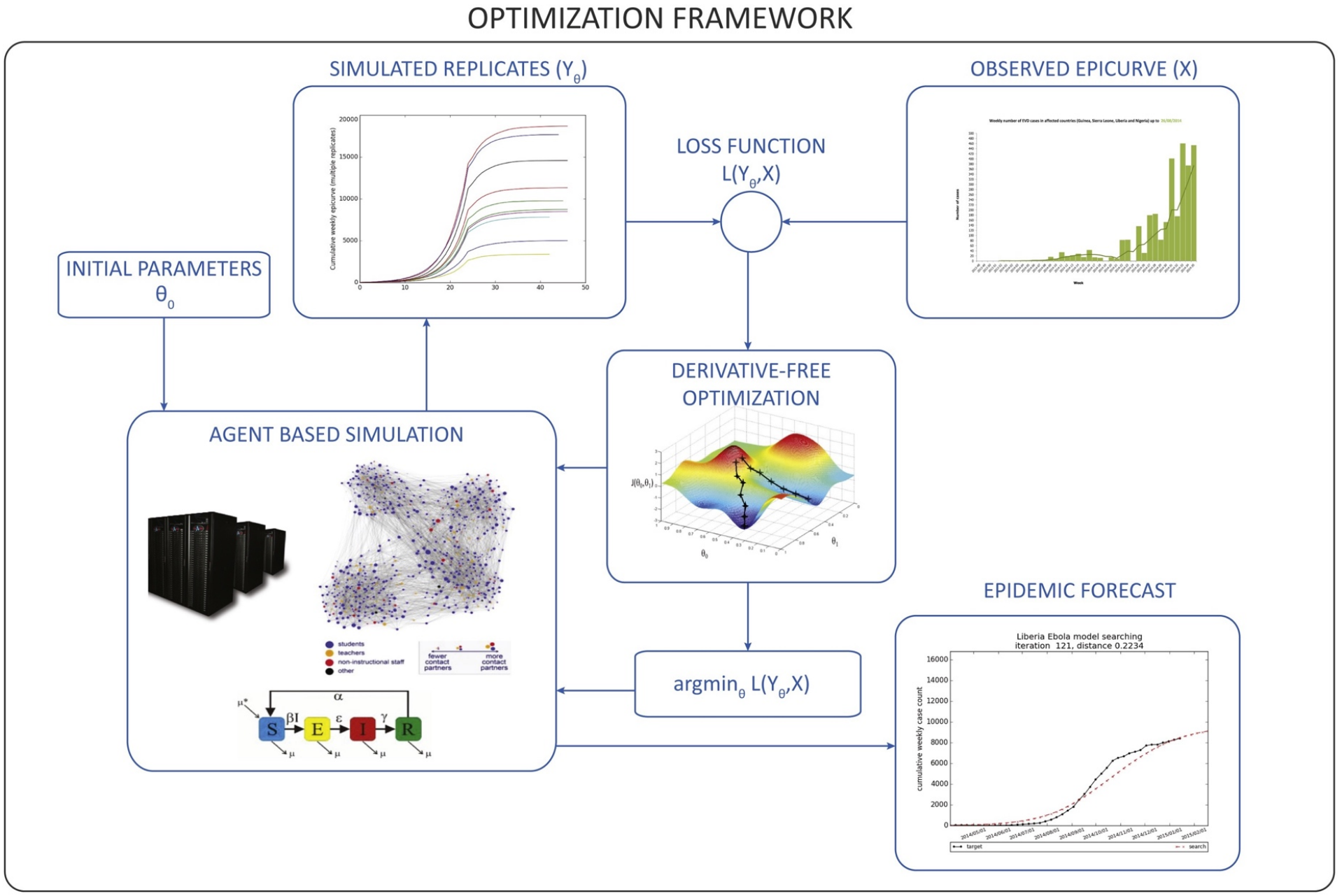
From structural analysis to co-evolution



Example of complex workflows: synthesizing socio-technical networks



Example of complex workflows: Forecasting disease dynamics

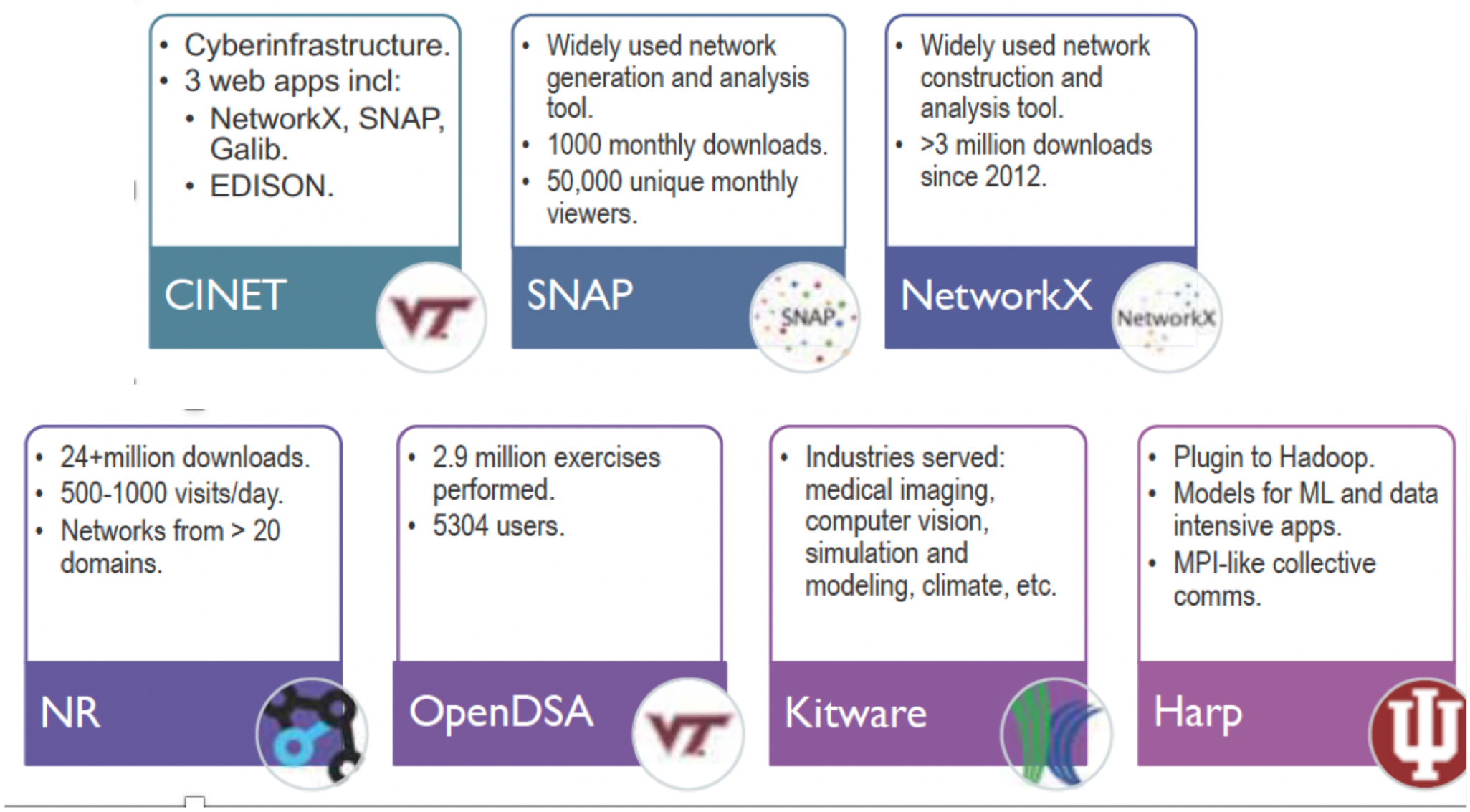


CINES Vision and team

How did CINES come about

- Team members worked on various components via funded programs from NSF (OCI, DIBBS, NETSE, SBE), DoD, DHS, NIH, etc. over the last 20+ years
- Conversations with Prof. David Easley (Cornell) when he was teaching the Network Science course at Cornell.
- Supporting decision makers at NIH/CDC (Public health), DoD (Disaster preparedness)

Team Capabilities



Team and advisory board

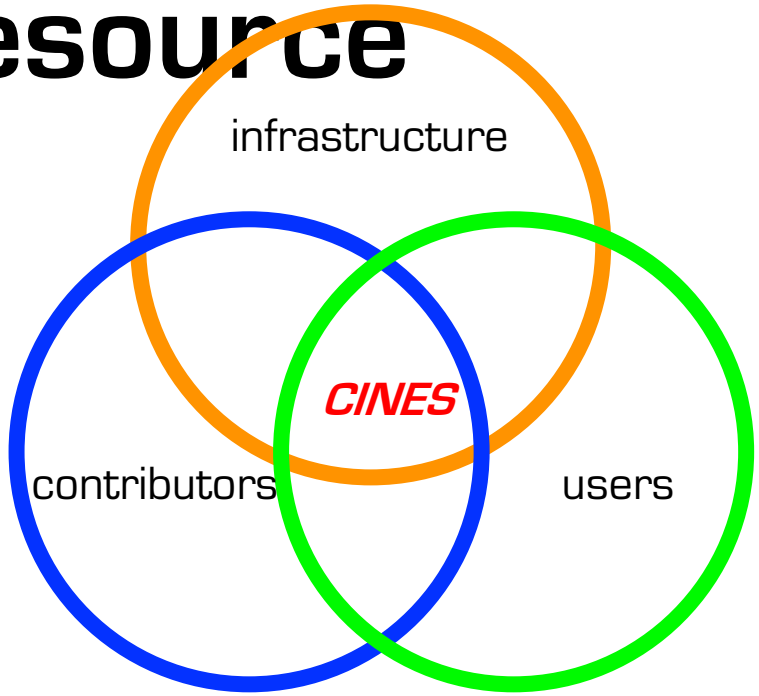
- **Indiana**: Geoffrey Fox, Judy Qiu, Gregor von Laszewski
- **Jackson State:** M. Natarajan Meghanathan
- **North Carolina A&T State**, NCAT: Albert Esterline
- **Stanford**: Jure Leskovec, Rok Sosic
- **University of Virginia**: Madhav V. Marathe Christopher Kuhlman, Dustin Machi and S. S. Ravi
- **Virginia Tech**: Catherine Amelink. Kristy Collins, Edward Fox and Naren Ramakrishnan, Yasuo Miyazaki
- **Los Alamos National Laboratory**: Aric Hagberg
- **Kitware**, Inc: Aashish Chaudhary
- **Network Repository**: Ryan Rossi, Nesreen Ahmed
- **New City Media**: David Poteet
- **Persistent Systems Inc**

- Richard Alo (Florida Agricultural and Mechanical)
- Noshir Contractor (Northwestern)
- Matthew Jackson (Stanford)
- Pamela Murray-Tuite (Clemson)
- Y. Narahari (Indian Institute of Science)
- Arun Phadke (Virginia Tech)
- Cliff Shafer (Virginia Tech)
- Zoltan Toroczkai (Notre Dame)
- Stanley Wasserman (Indiana)

Virginia Tech, Indiana U., Stanford U., Los Alamos National Lab, Jackson State U., North Carolina A&T, Kitware, Network Repository, New City Media , Persistent Systems and **25**+ US and international collaborators

CINES vision: a community resource

- A gateway for network science.
  - General users & Domain-specific users
  - Domain experts as well as non-CS analysts
- Community resource; users:
  - contribute data, codes.
  - support (teach, show) others how to use.
  - report bugs.
  - suggest new functionality, features.



CINES: a resource built by the community & for the community

CINES vision: contributor centric to make it sustainable

- Make simple interfaces for individuals to contribute (codes, data, machines, courses, ...)
- Give prominent and clear credit to contributors.
- Provide contributors with accessible metrics so that they can value their own contributions.

Expected CINES resources

What resources will CINES have: I

A digital library for 1-5 million networks:

- Social networks; Infrastructure networks; Biological networks
  - E.g. We are working with the NIH funded PATRIC project that has close to a million bacterial and viral genome data. A number of omic networks can be synthesized and network analytics tools can be used to gain insights
- Directly measured as well as synthesized from data
- Workflows to generate large number of networks
- Networks: **not just** undirected **but also** be directed, labeled, weighted and time-varying
- Digital library services for browsing, searching, organizing networks
- Users can add their networks easily and tag them

What resources will CINES have: II

A large code repository as a way to seed the code base

- Over 200 code fragments already available
- Structural analysis, e.g. finding shortest paths, spanning trees …
- Dynamics over networks: e.g. simulating the mis-information spread,
- Optimization and control: e.g. finding optimal nodes to delete to control the spread of mis-information
- Inference: e.g. Finding the most influential node, index case…
- Machine learning for networks: e.g. representational learning
  - E.g. Stanford team has pioneered representational learning for networks; new methods are being developed and will be made available to the community

Not just individual codes but workflows for users to seamlessly glueing different codes

What resources will CINES have: III

Educational material

- Modules to teach network science
- Repository for available courses, conferences, journals etc. on the web
- Web-apps so that the data and software can easily be used by social, behavioral and biological scientists
- Simple workflows for non-experts
- Visualization tools
- Easy access for schools and researchers that are resource constrained

What resources will CINES have: IV

Common framework to support reproducibility and standards

- Support journals and conferences to ensure the claims about network codes hold
  - Current practice is based on honor system: this will soon be unsustainable
  - Will reduce duplication of effort
- Lead to standards for important network theoretic problems,
  - e.g. yet another shortest path algorithm: researcher should try and easily compare performance with known methods.
  - Will support benchmarking of implementations

Timeline for the next two years

- **Summer 2020**: CINES version 1
  - Basic codes from NetworkX, Galib, SNAP and NR will be available to build larger workflows by glueing
  - A website that will provide information on courses, journals, conferences, and other related resources
  - Networks that have already been collected by team members.
- **Spring 2021**: CINES version 2
  - Network synthesis workflows based on social media & biological data
  - Simulations for dynamics over networks
  - Prototype digital library for network browsing, searching etc.

Summary

- We welcome participation from you all as we build CINES
  - A number of CI projects discussed already seem to have overlap (e.g. Multinet, Water CIs, Urban sustainability projects, health)
  - Infrastructure projects (e.g. workflows, monitoring, …)
- But we invite others as well, e.g. physicists to consider use of networks to study cosmology.

Extras