Software and IT-CAST

Why Does Software Cost So Much? Toward a Causal Model

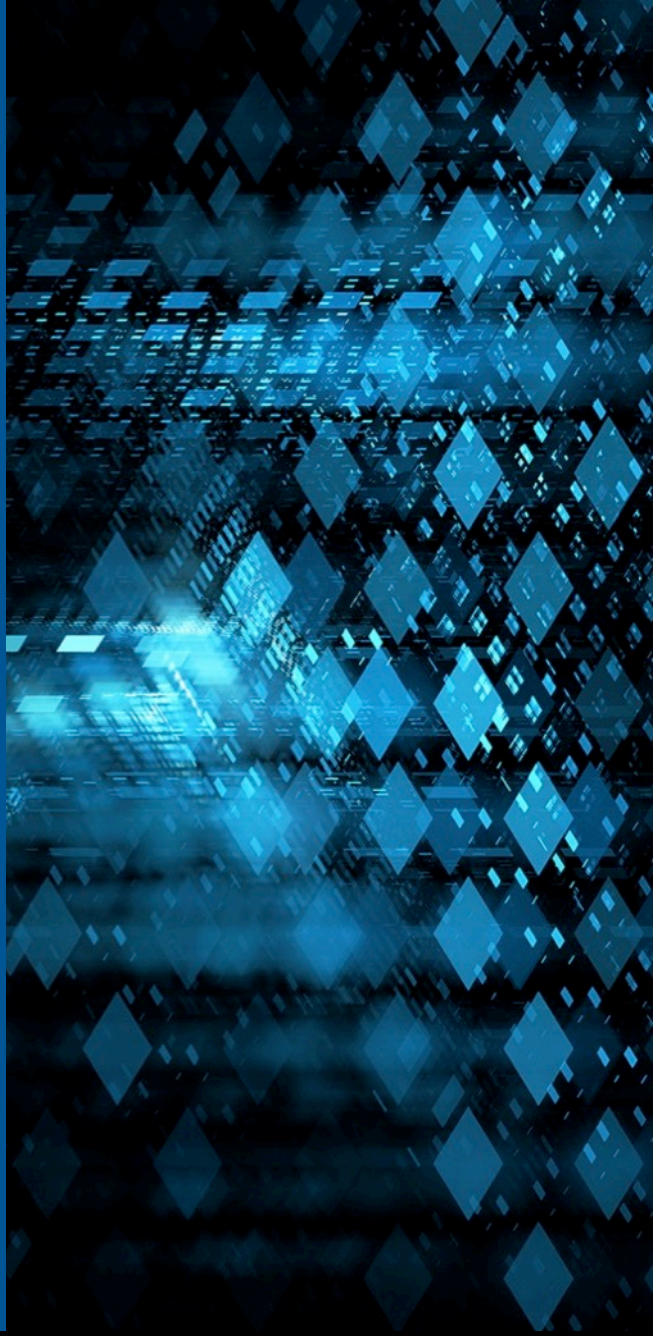
23 August 2017

Mike Konrad

Robert Stoddard

David Zubrow

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213



© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Copyright 2017 Carnegie Mellon University. All Rights Reserved.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

PSP℠ is a service mark of Carnegie Mellon University.

DM17-0537

Outline

- What is causal learning and modeling, and why do we care about It?
- Our technical approach
- Initial Results
- Conclusions

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Our Project: Bottom Line Up Front

Goal

- Demonstrate the benefit of causal modeling to the software cost domain
- Identify and quantify a causal network of factors that drive software effort and schedule

Actionable intelligence

- Enhance program control of software cost throughout the development and sustainment lifecycles
- Inform “could/should cost” analysis and price negotiations
- Improve contract incentives for software intensive programs
- Increase competition using effective criteria related to software cost

If you are interested in this approach, let’s work together.

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

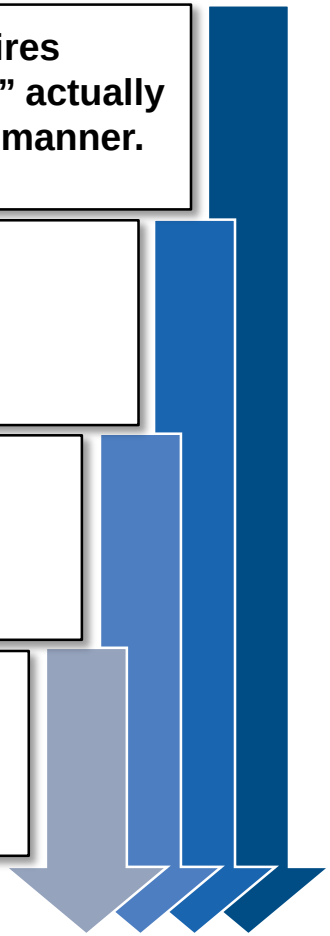
Why do we care about causal modeling?

Proactively controlling software costs requires knowing which of our "independent factors" actually *cause* outcomes to change in a predictable manner.

Just as correlation may be *fooled* by spurious association, so can regression

We must move beyond correlation to *causation, if we want to make use of cause and effect relationships*

Today, we can garner evidence of causation without the *expense* and challenge of conducting a controlled experiment

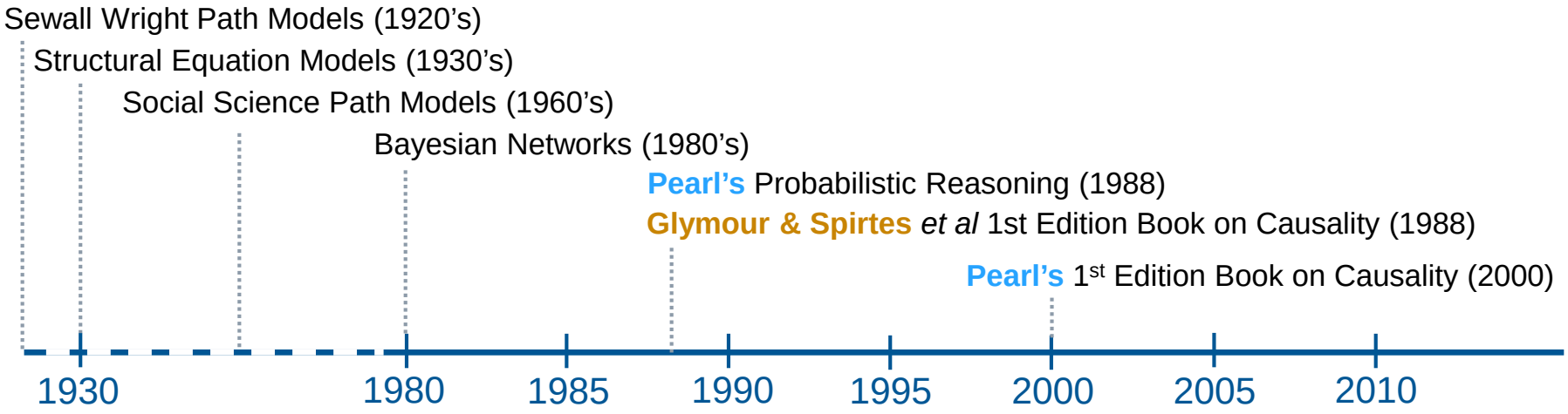


Establishing causation with observational data remains a vital need and a key technical challenge, but is becoming more feasible and practical.

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Significant Progress Toward Practicality



TETRAD – An Open Source Tool for Causal Learning

Carnegie Mellon University
http://www.phil.cmu.edu/tetrad/

University of Pittsburgh
http://www.ccd.pitt.edu/

For video tutorials from 2016 summer short course:
http://www.ccd.pitt.edu/training/presentation-videos/

Glymour & Spirtes *et al* 2nd Edition Book on Causality (2001)

Morgan Counterfactuals & Causality (2007)

Pearl's 2nd Edition Book on Causality (2009)

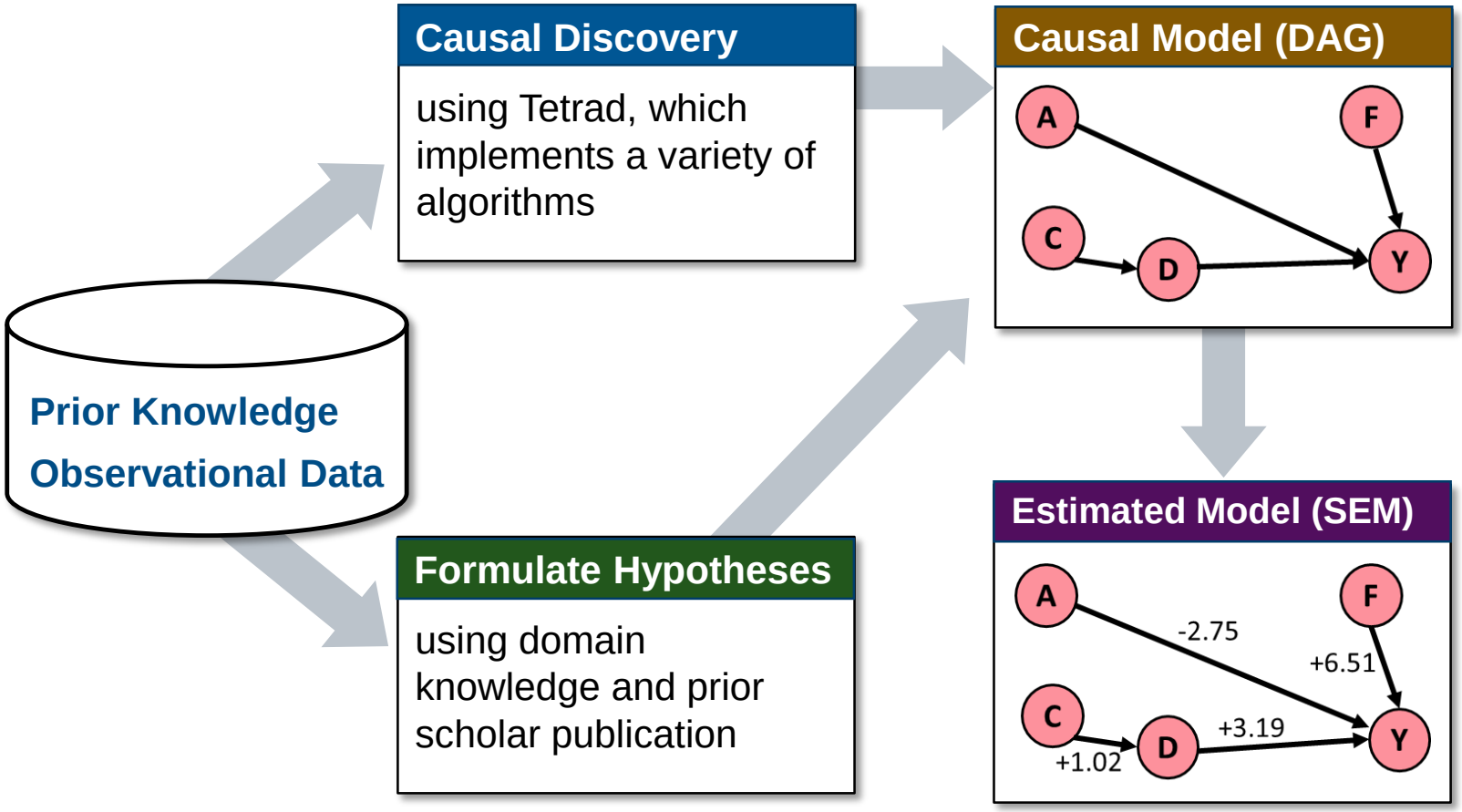
Morgan Counterfactuals & Causality (2014)

Morgan Handbook Social Science Causal Inference (2014)

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

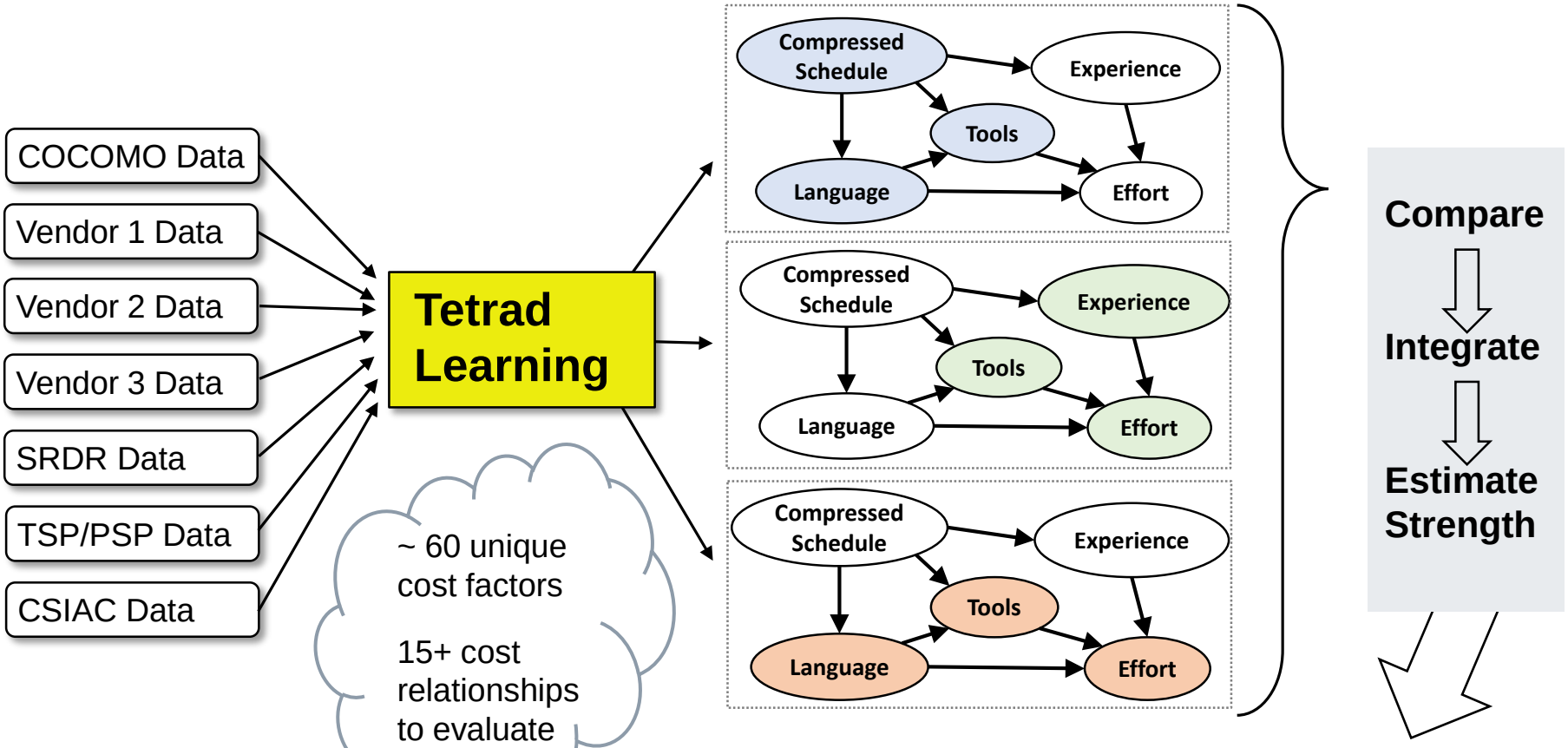
Basic Technical Approach



© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Integrating Models



Actionable Causal Models

Module Effort = f(factor1, factor2, factor3)
Module Post-Development Quality = g(factor1, factor4, factor5)
High-Reliability Module Cost = h(factor4, factor6, factor7)

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Example: PSM Performance Analysis Model

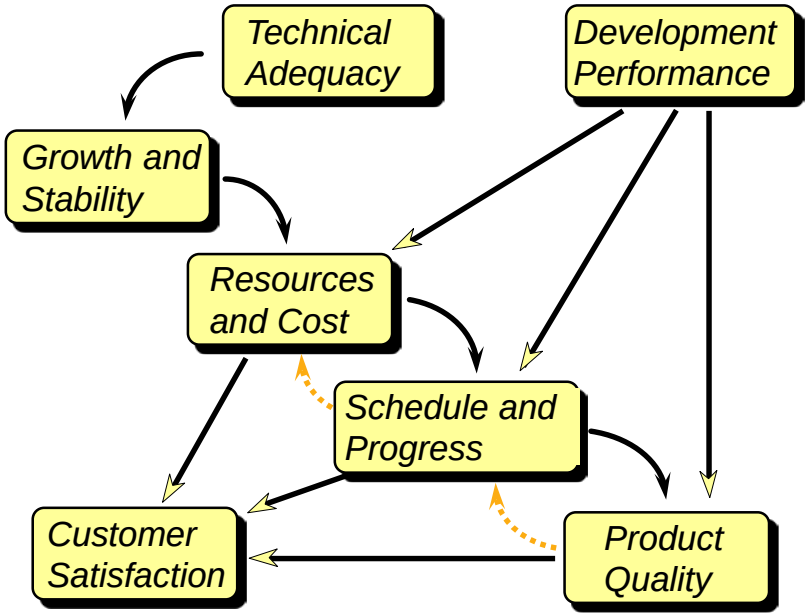
A familiar example of a causal model

Hard to find data sources to actually estimate the entire model

Consequently harder to empirically establish the causal relationships

Causal modeling methods allow for the integration of partial models

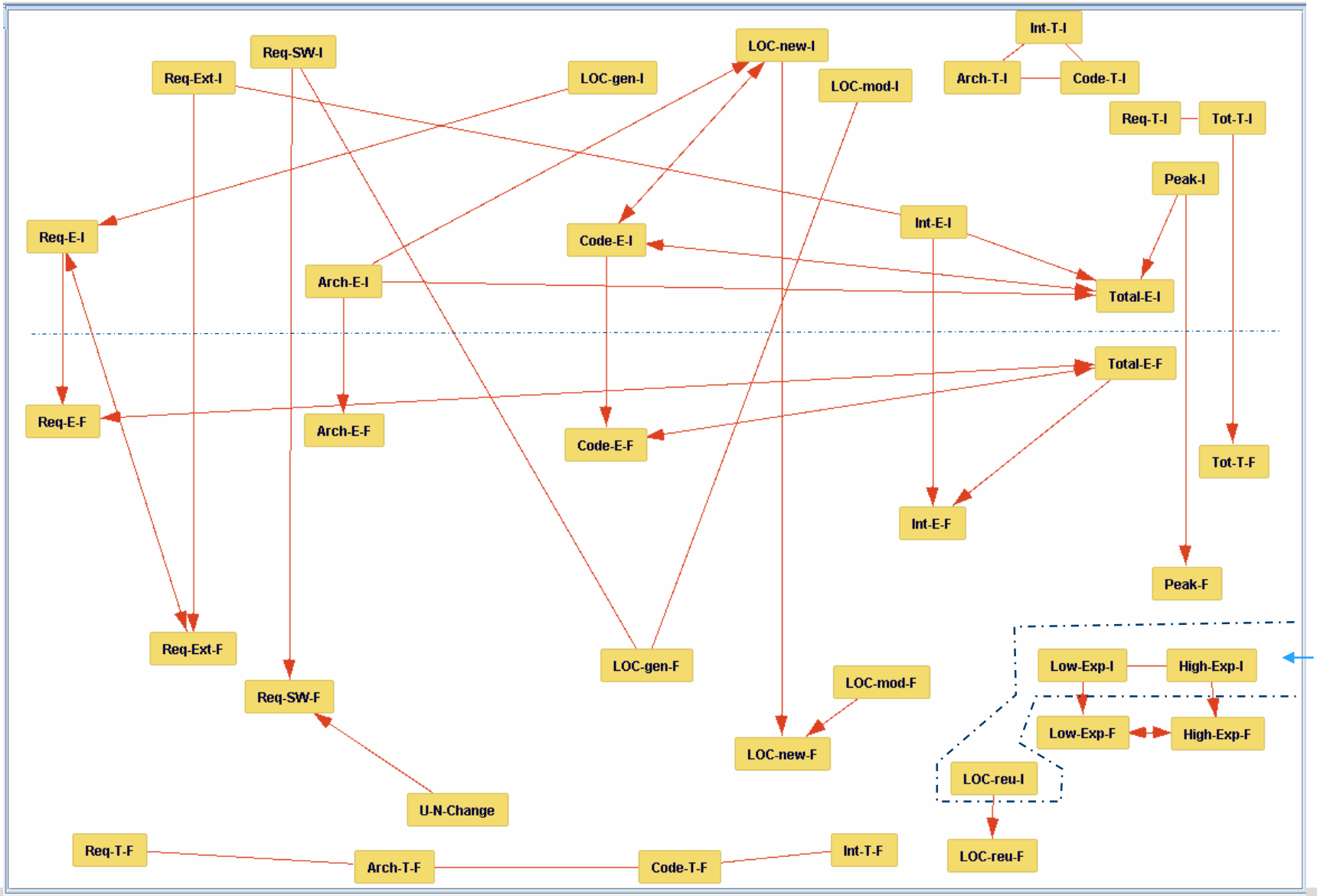
Opportunity for empirical support and refinement



© 2017 Carnegie Mellon University
[Distribution Statement A] Approved for public release and unlimited distribution.

Explaining Final Effort and Duration (Initial Results)[1]

181 pairs of matched initial-final SRDR reports reduced to 134 (complete Req…INT data).

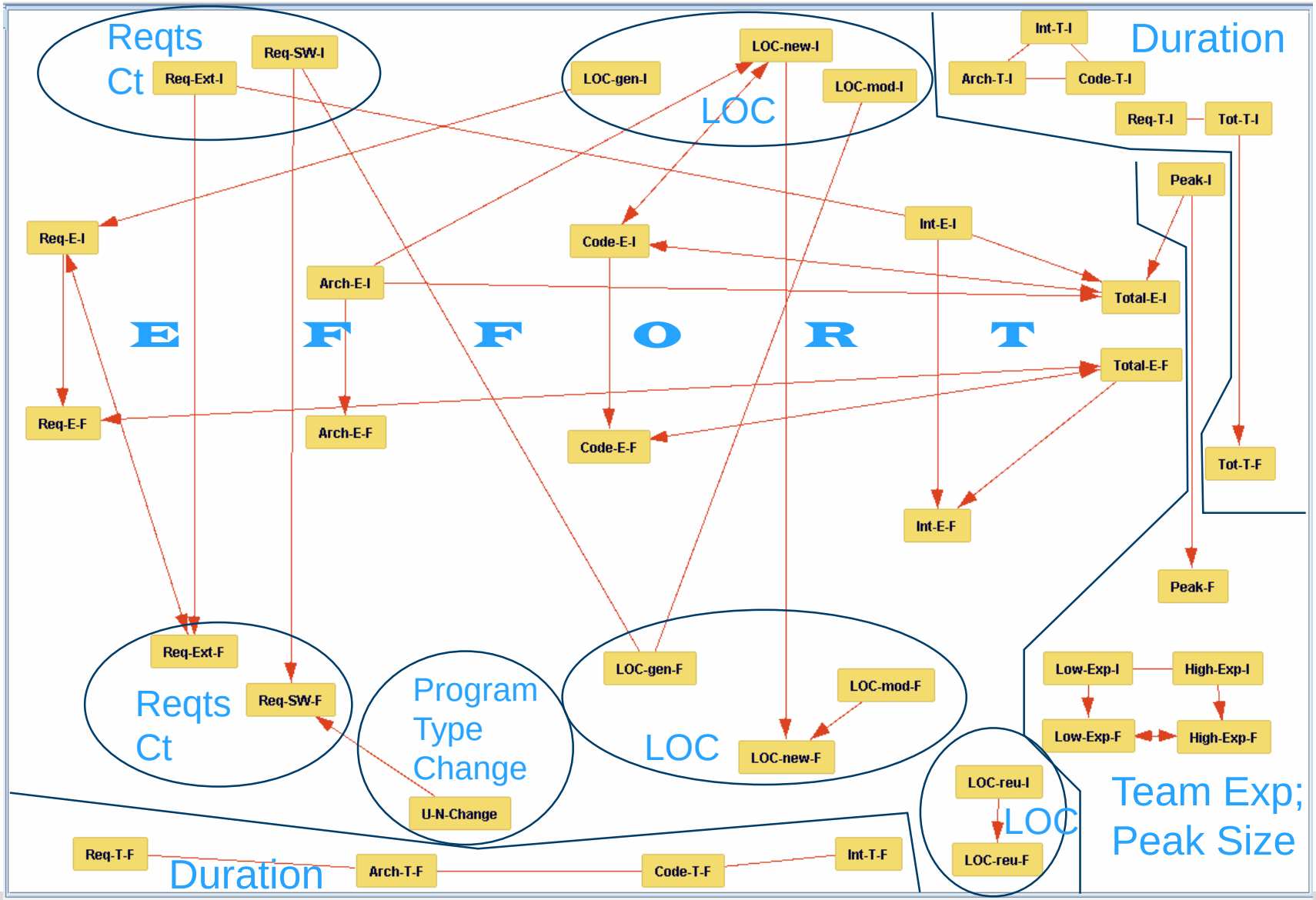


© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Explaining Final Effort and Duration[2]

Both this chart and previous analyzed with PC algorithm with Alpha set to .001.



© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

What Do These Initial Results Suggest?

Effort estimates for **Req, Arch, Code, INT** directly influence effort actuals.

- Not so for **Duration**

There are other cases where estimates of an attribute do not directly influence actuals for that attribute, suggesting challenges to estimation.

Total effort actual

- may be directly influenced by **Req effort** and **Code effort** actuals
- not directly influenced by **Arch effort** actual
- directly influences **INT effort** actual (after accounting for influence of initial INT effort estimate). Evidence of effort compression?

Cautions

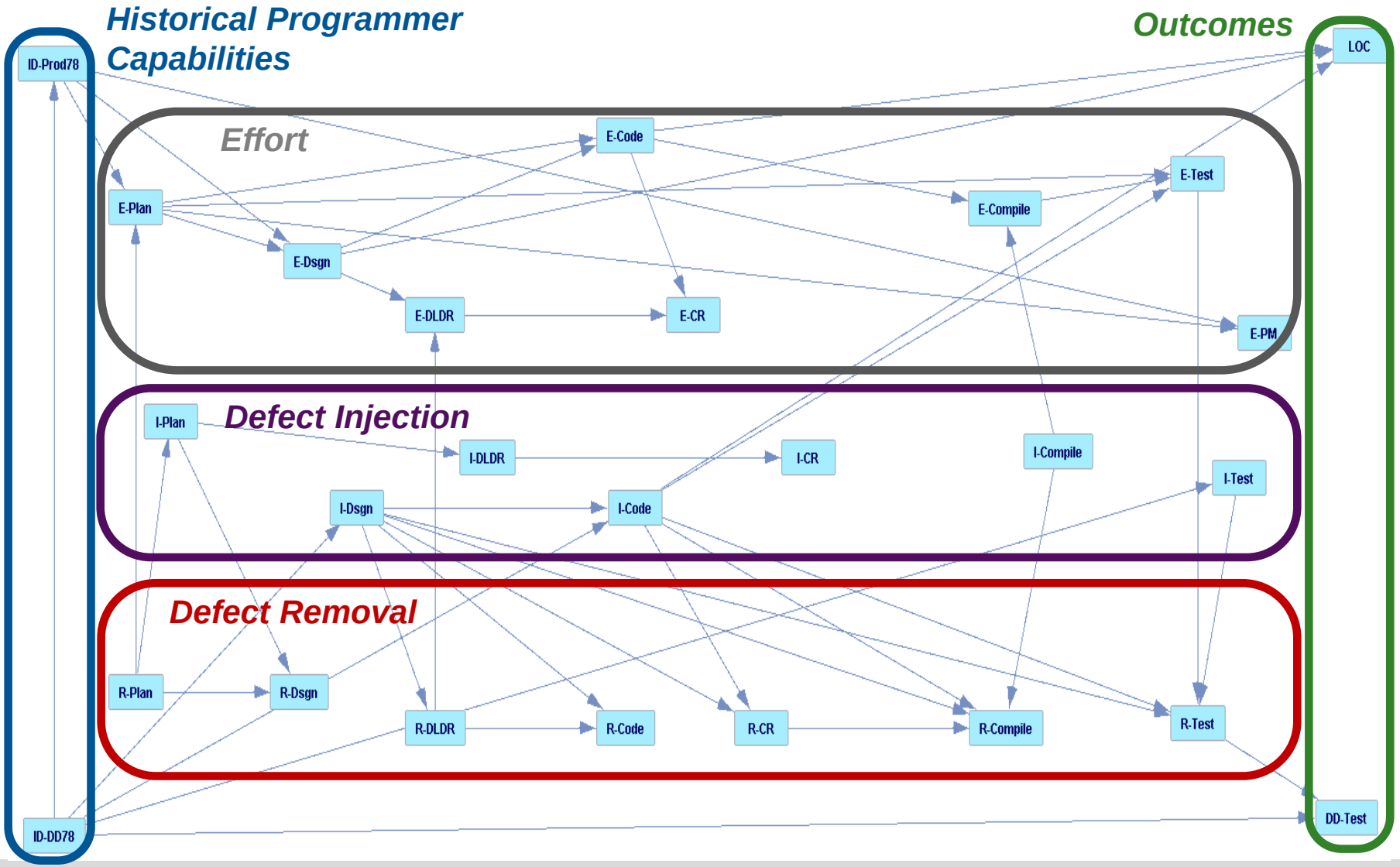
- Double-headed edges suggest unmeasured confounders (factors that are a common cause of factors connected by the edge).
- Undirected edges suggest insufficient data.

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Explaining Size and Defect Density – Need to Drill Deeper?

Data from 975 programmers during PSP training



© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Conclusions

Causal learning:

- has come of age from both a theoretical and practical tooling standpoint
- may be performed on data whether it be derived from experimentation or passive observation

Causal models:

- help separate true causes from spuriously-correlated factors
- help identify when unknown causes may likely exist
- lend themselves to actionable intelligence better than models based on correlation

We welcome collaborators interested in using these methods and tools.

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

QUESTIONS?

© 2017 Carnegie Mellon University

[Distribution Statement A] Approved for public release and unlimited distribution.

Contact Information

Points of Contact

Robert Stoddard
rws@sei.cmu.edu

Mike Konrad
mdk@sei.cmu.edu

Dave Zubrow
dz@sei.cmu.edu

William Nichols
wrn@sei.cmu.edu

David Danks
ddanks@cmu.edu

Kun Zhang
kunz1@cmu.edu

U.S. Mail

Software Engineering Institute
Customer Relations
4500 Fifth Avenue
Pittsburgh, PA 15213-2612, USA

Web

www.sei.cmu.edu
www.sei.cmu.edu/contact.cfm

Customer Relations

Email: info@sei.cmu.edu
Telephone: +1 412-268-5800
SEI Phone: +1 412-268-5800
SEI Fax: +1 412-268-6257

© 2017 Carnegie Mellon University
[Distribution Statement A] Approved for public release and unlimited distribution.