IMPROVING DESTINATION CHOICE WITH AI

Caliper

Transportation & Mapping Solutions

Maptitude • TransCAD • TransModeler

CONTEXT

Caliper

CONTEXT OF FHWA TMIP PROJECT

- Acknowledgement and thanks for FHWA sponsorship of this important work

- Part of larger project to improve travel forecasting through the use of **big data** and **AI**
  - ***Review of literature and practice***
  - Testing new methods
  - Implementation pilot projects with case studies
  - “Playbook” for incorporating AI in travel models
  - TMIP webinars to promote Playbook methods

CALIPER TEAM

Vince Bernardin, PhD
Project Manager

Andrew Rohne

Rama
Balakrishna, PhD

Howard Slavin, PhD
Senior Advisor

Wuping Xin, PhD
Deputy Project Manager

Kyle Ward

Srini Sundaram

Andres
Rabinowicz, DSc

EXPERT PANEL

Francisco Pereira, PhD
Panel Lead

Kara Kockelman, PhD

Mark Bradley

Joshua Auld, PhD

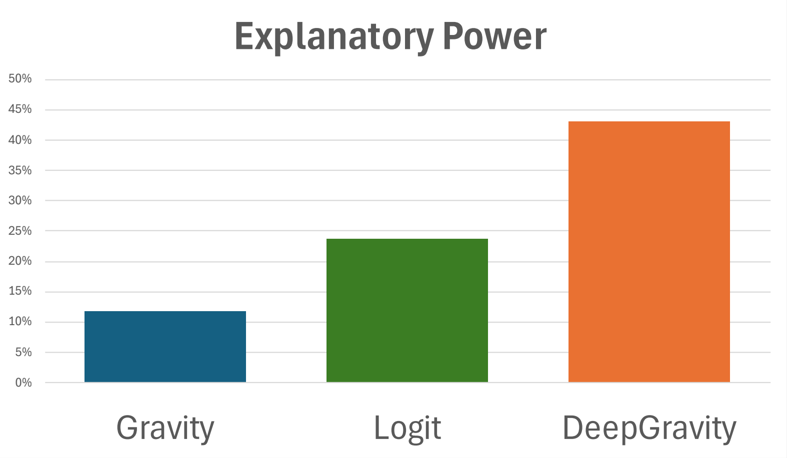
Brian Gregor, PE

Sabya Mishra, PhD

Dan Work, PhD

PROJECT FOCUS

- Focus on AI
  - References to TMIP resources on big data
- Focus on Practical Improvements for the Near- to Mid-Term
  - Methods to improve/replace individual model components
  - AI-DCMs
  - Primary focus on Destination Choice
    - Largest source of error in existing models – largest opportunity for improvement



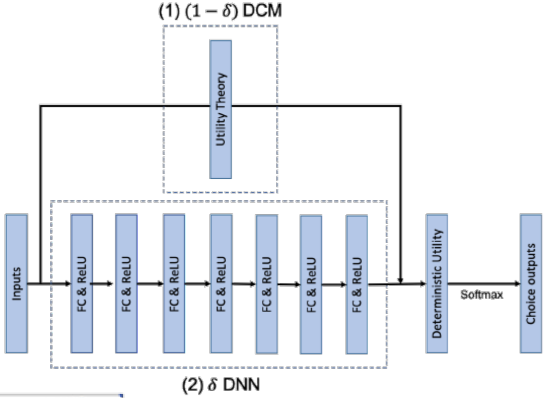
AI-DCM MODELS

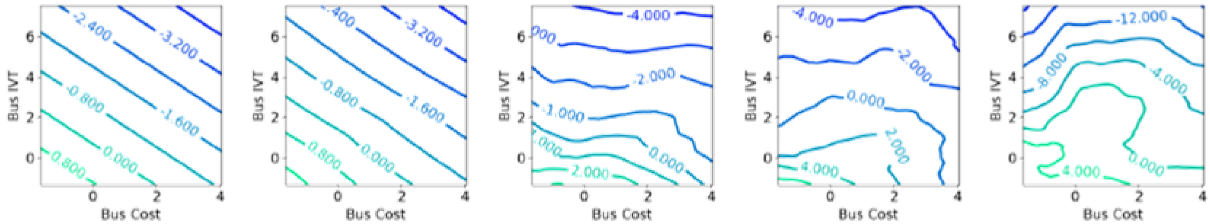
- Artificial Intelligence – Discrete Choice Models
- Combine neural networks and logit models
- Attempt to combine the best of both traditional and newer methods
  - Theoretical basis and interpretability of traditional models
  - Explanatory power and accuracy of AI
- Six types proposed so far
  - L-MNL
  - ResLogit
  - TB-ResNet
  - TasteNet
  - RUMnets
  - e-Logit

TB-RESNETS

- Ensemble of Logit and Deep NN
- Interpretable as a logit or DNN
- Utilities weighted average of logit and DNN
- Weight estimable from data

Fig. 1. Architecture of TB-ResNet. Both DCM and DNN are flexible





(a) MNL (50.6%) (b) MNL-ResNet ($\delta = 10^{-5}$; 53.1%) (c) MNL-ResNet ($\delta = 0.008$; 57.0%) (d) MNL-ResNet ($\delta = 0.05$; 56.1%) (e) DNN (55.8%)

Fig. 2. Utility functions of MNL-ResNets, MNL, and DNNs. Upper row: visualization of 2D utility functions, and percentages in the parentheses represent the prediction accuracy. Lower row:

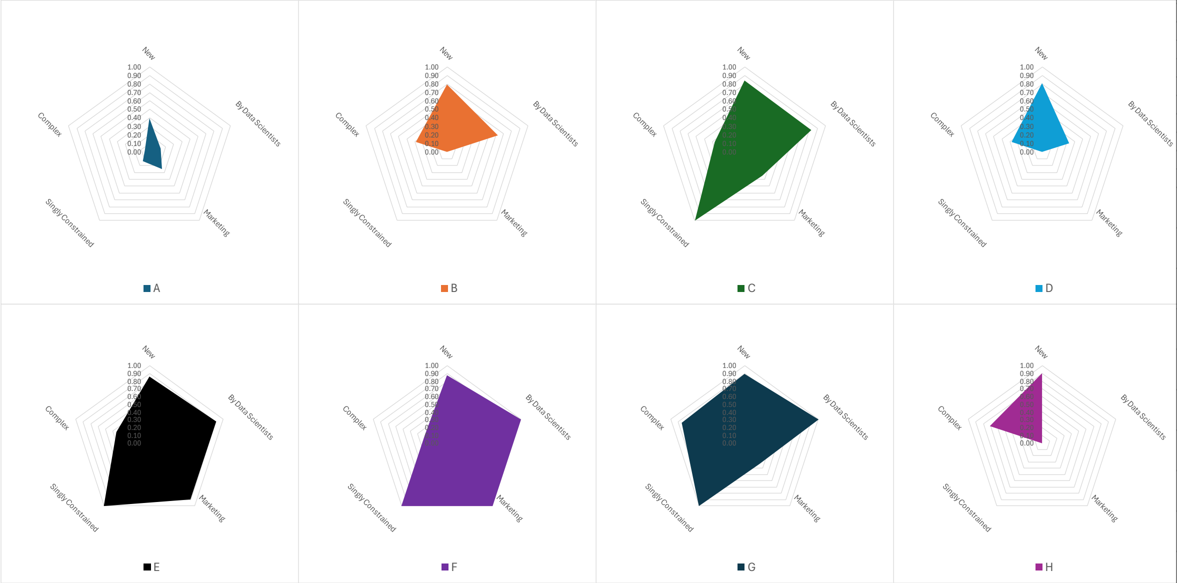
LITERATURE REVIEW

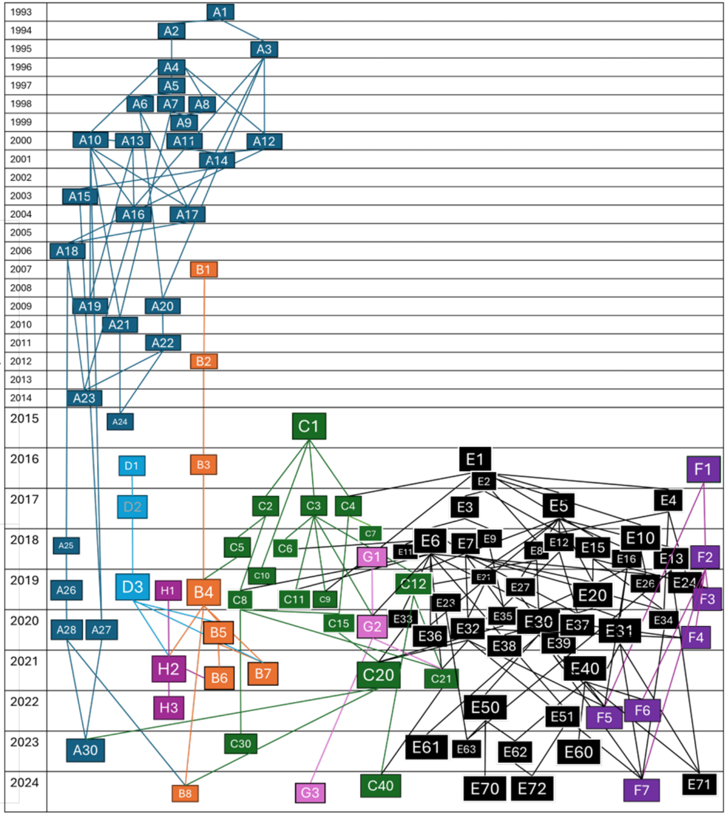
LITERATURE REVIEW

- Identified 354 papers from 1993 to present
- Explosion of papers from 2016, peaking in 2020, stabilized around 2018-19 levels
- Needed to prioritize, mostly based on citation rates
- Cursory review of 123 papers and 18 surveys/reviews
- Report summarizes 34 papers
  - Plus, a brief overview of 15 early papers
  - And appendix with 13 paper summaries
- Identified 8 branches of the literature

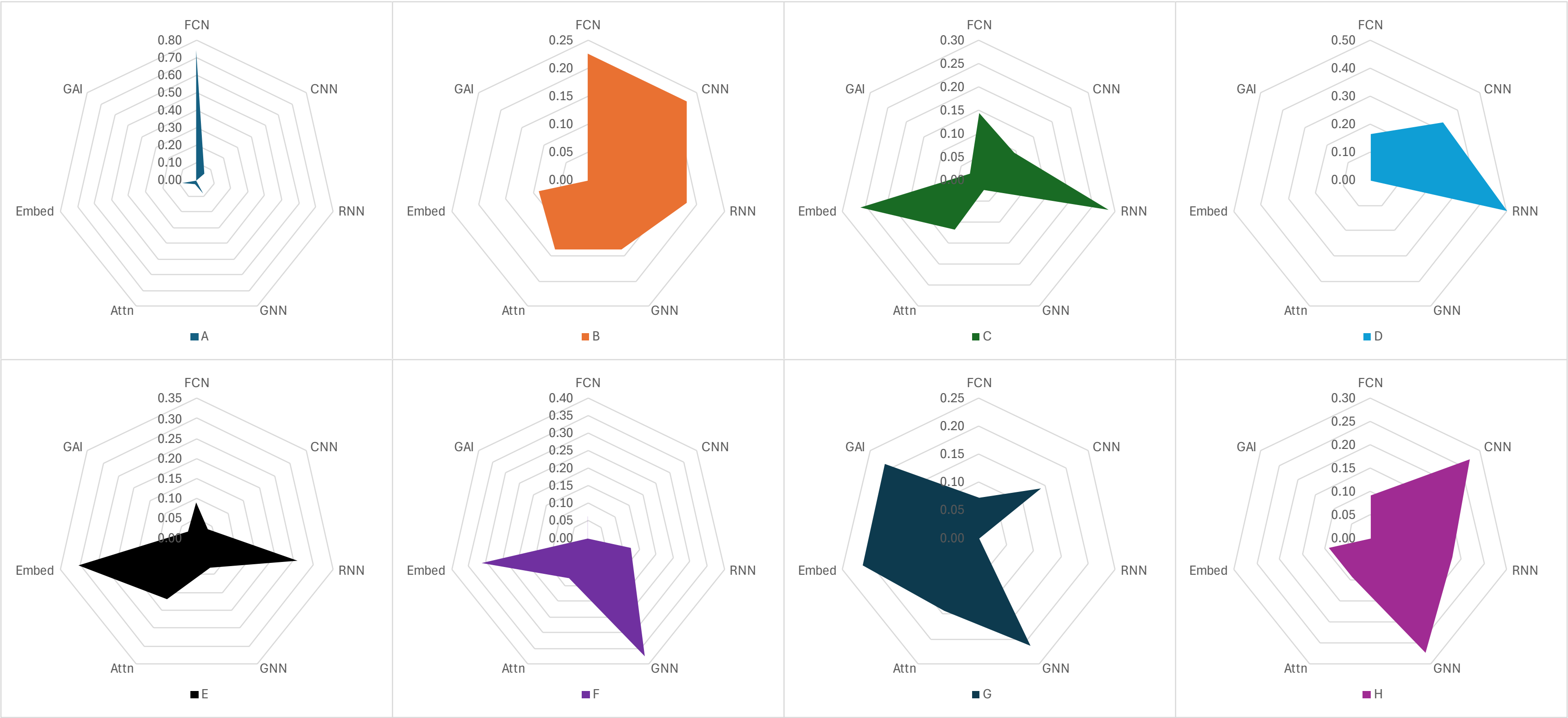
BRANCHES OF THE LITERATURE

- Eight branches of the literature
  - Based on citations, but vary across many dimensions





BRANCHES METHODOLOGICAL FOCUS



MODEL-BASED META-ANALYSIS

HOW TO COMPARE MODELS?

- 22 different metrics reported
  - 14 goodness-of-fit metrics
  - 8 error metrics
- Assumption:
  - **Relative** improvement in fit or decrease in error are comparable, though not identical, regardless of fit / error metric used
- Approach:
  - Model a latent generic fitness measure which minimizes squared error between modeled and published relative comparisons

Metric	Type	Normalized	% Papers Reporting
RMSE	Error	No	26.6%
k-Recall / HR	Fit	Yes	21.1%
k-Accuracy	Fit	Yes	20.2%
MAE	Error	No	13.8%
R2	Fit	Yes	12.8%
k-MAP	Fit	Yes	11.0%
k-Precision	Fit	Yes	10.1%
k-NDCG	Fit	Yes	8.3%
F1 / DSC	Error	Yes	9.2%
MAPE	Error	Yes	7.3%
MSE	Error	No	7.3%
MRR	Fit	Yes	7.3%
AUC	Fit	Yes	6.4%
ARV	Error	No	6.4%
Distance	Fit	Yes	5.5%
JSD	Fit	Yes	3.7%
sMAPE	Fit	Yes	3.7%
SRMSE	Fit	Yes	2.8%
LL	Error	Yes	1.8%
k-Top	Fit	Yes	0.9%
WMAPE	Error	Yes	0.9%
k-DCG	Fit	No	0.9%

LATENT FITNESS MODEL

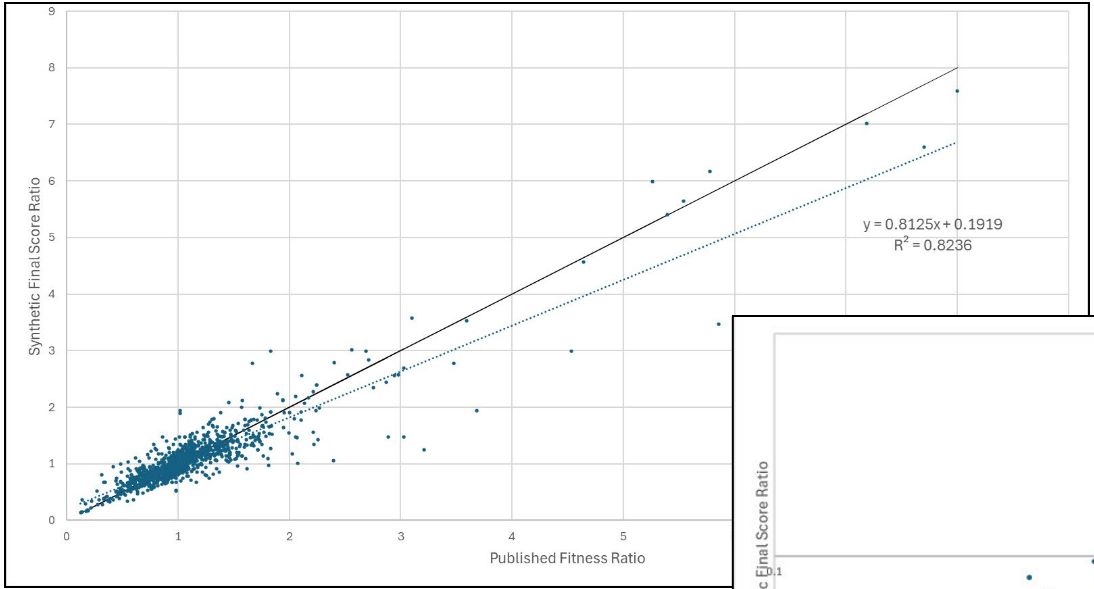
- Latent fitness score defined on unit interval [0,1]
- Binary logit model
  - Model specific constant
  - 10 methodological dummy variables
    - FCN
    - RNN
    - CNN
    - GNN
    - GCN
    - Attention
    - Embeddings
    - SSL
    - LLM
    - GAN
- LSE with regularization term
  - (squared difference from initial score calculated as normalized average of ratio of model's goodness-of-fit to other models)

DATA CONSTRUCT

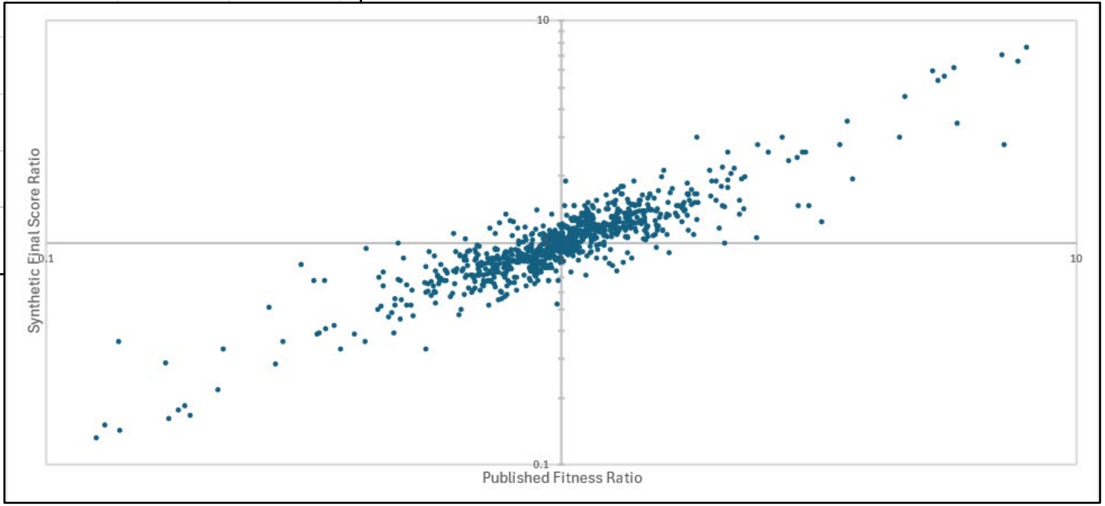
- 12 metrics used in meta-analysis
- Preference for normalized
  - 78% normalized used in meta-analysis
  - Highest preference for metrics normalized on the unit interval
- Observed Data:
  - 629 relative comparisons
  - Published in 81 papers
  - Which used 176 datasets

Metric	Type	Normalized	% Papers Reporting	% Comparisons in Meta-Analysis
RMSE	Error	No	26.6%	9.4%
k-Recall / HR	Fit	Yes	21.1%	12.7%
k-Accuracy	Fit	Yes	20.2%	21.8%
MAE	Error	No	13.8%	0.0%
R2	Fit	Yes	12.8%	2.4%
k-MAP	Fit	Yes	11.0%	0.0%
k-Precision	Fit	Yes	10.1%	2.4%
k-NDCG	Fit	Yes	8.3%	0.0%
F1 / DSC	Error	Yes	9.2%	16.5%
MAPE	Error	Yes	7.3%	8.0%
MSE	Error	No	7.3%	0.0%
MRR	Fit	Yes	7.3%	0.0%
AUC	Fit	Yes	6.4%	4.1%
ARV	Error	No	6.4%	0.0%
Distance	Fit	Yes	5.5%	2.2%
JSD	Fit	Yes	3.7%	3.7%
sMAPE	Fit	Yes	3.7%	3.3%
SRMSE	Fit	Yes	2.8%	0.8%
LL	Error	Yes	1.8%	0.0%
k-Top	Fit	Yes	0.9%	0.0%
WMAPE	Error	Yes	0.9%	0.0%
k-DCG	Fit	No	0.9%	0.0%

MODELED SCORE RATIOS VS. PUBLISHED



$r^2 = 0.824$



META-ANALYSIS RESULTS

- Best methods
  - GAI
    - GAN
    - LLM
  - SSL
  - GCN
- Small Sample Size for best
  - GAI (8)
  - SSL (6)
  - LLM (3)

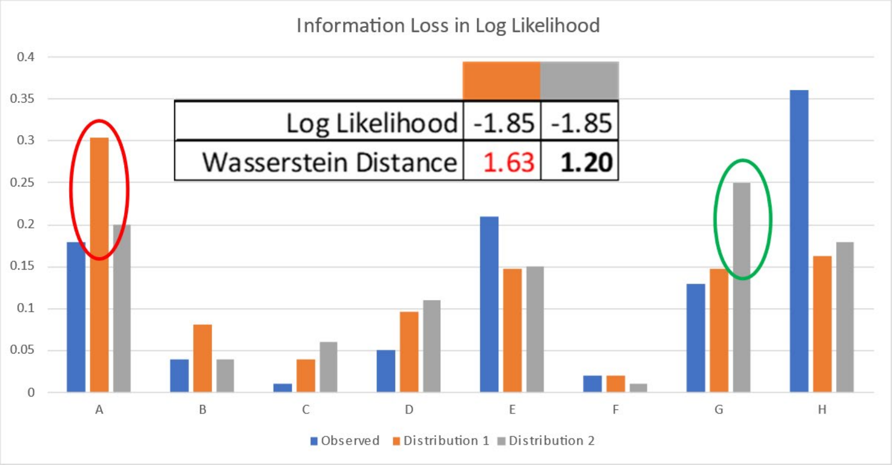
	Utility Coefficient	Factor	Avg. Score
FCN	-0.111	0.89	0.37
RNN	-0.250	0.78	0.35
CNN	0.014	1.01	0.41
GNN	0.046	1.05	0.39
GCN	0.066	1.07	0.44
Attention	-0.155	0.86	0.45
Embeddings	-0.162	0.85	0.41
SSL	0.110	1.12	0.43
GAN	1.790	5.99	0.79
LLM	0.518	1.68	0.66

RECOMMENDATIONS FOR NEXT PHASE

Caliper

PERFORMANCE MEASUREMENT

- Importance of Out-of-Sample (Holdout Sample) Validation
  - Standard practice of good data science
  - Extremely rare in travel forecasting practice
  - Key opportunity to improve the practice
- Choice of Metric
  - Huge variety of error / goodness-of-fit metrics
  - Minimum Wasserstein distance
    - Powerful in computer vision, with CNNs
    - Gives credit for getting close



NOW TESTING

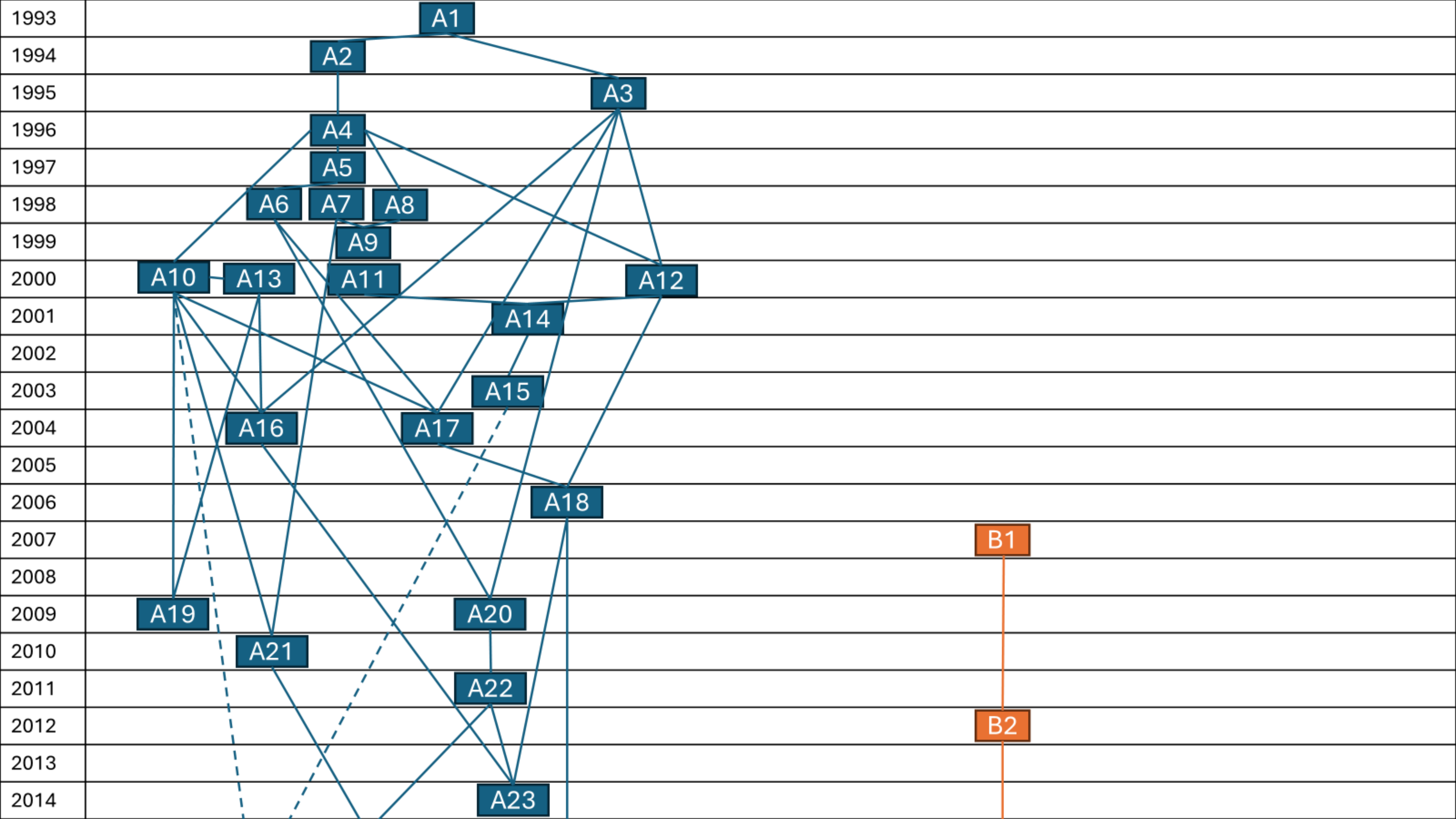
- Recommended models for testing in AI-DCMs
  - GAN: MoveSim/TrajGAN, highest scores
  - SSL GCN: STHGCN, #7 highest score, highest non-GAI, high confidence
  - MLP/FCN: DeepGravity, reference, average performance with minimal complexity

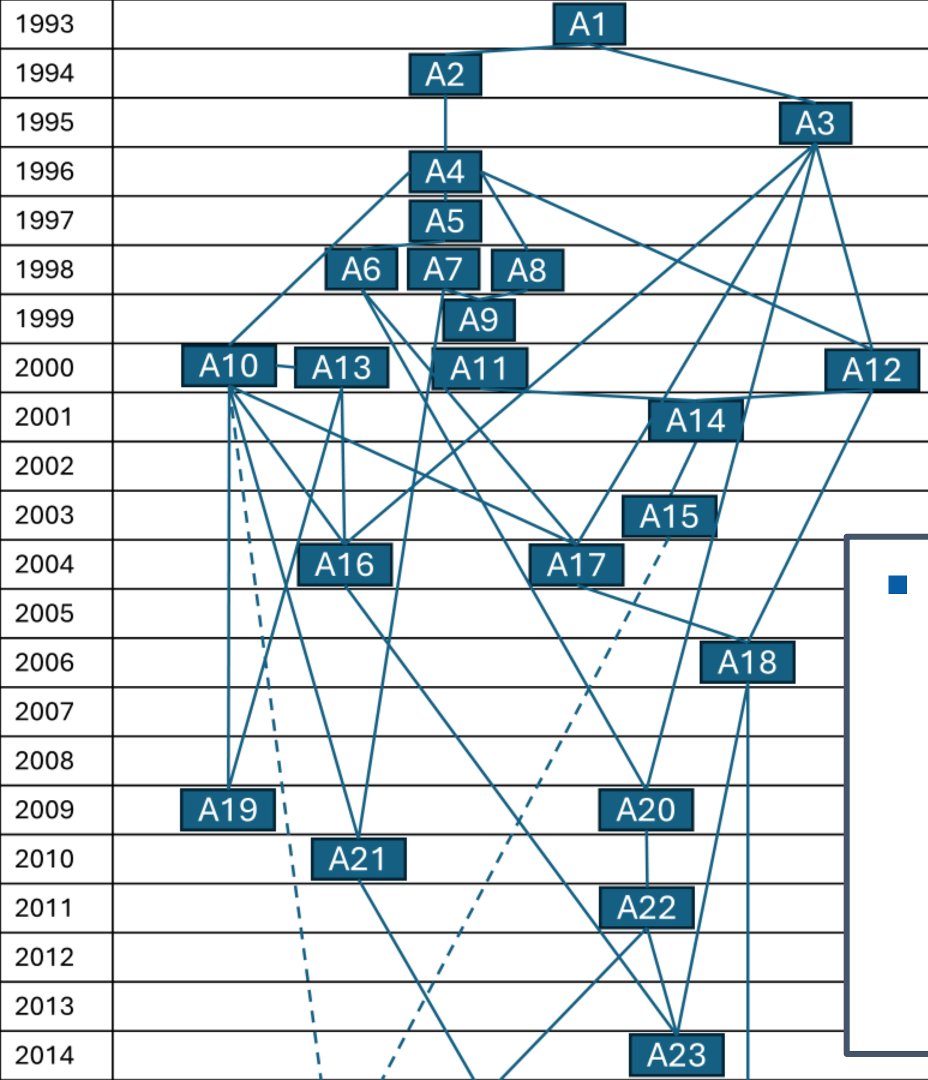
Rank	Model	Paper	Final Score	FCN	RNN	CNN	GNN	Attention	Embeddings	SSL	GAN	LLM
1	MoveSim	Feng et al. (2020a)	0.983	0	0	1	0	1	1	0	1	0
2	TrajGAN	Ouyang et al. (2018)	0.979	0	0	1	1	0	1	0	1	0
3	COLA	Wang et al. (2024)	0.950	1	0	0	0	1	1	0	1	0
4	LLM4POI	Li et al. (2024)	0.851	0	0	0	0	0	1	0	0	1
5	Geo-ALM	Liu et al. (2019b)	0.788	0	0	0	0	0	0	0	1	0
6	LLMove	Feng et al. (2024)	0.697	0	0	0	0	0	1	0	0	1
7	STHGCN	Yan et al. (2023)	0.675	1	0	1	1	0	1	1	0	0
8	CatDM	Yu et al. (2020)	0.669	0	1	0	0	0	1	0	0	0
9	EEDN	Wang et al. (2023b)	0.587	0	0	1	1	1	1	1	0	0
10	DRAN	Wang et al. (2022b)	0.551	0	0	0	1	1	1	0	0	0
43	DeepGravity	Simini et al. (2021)	0.412	1	0	0	0	0	0	0	0	0

CONTACTS

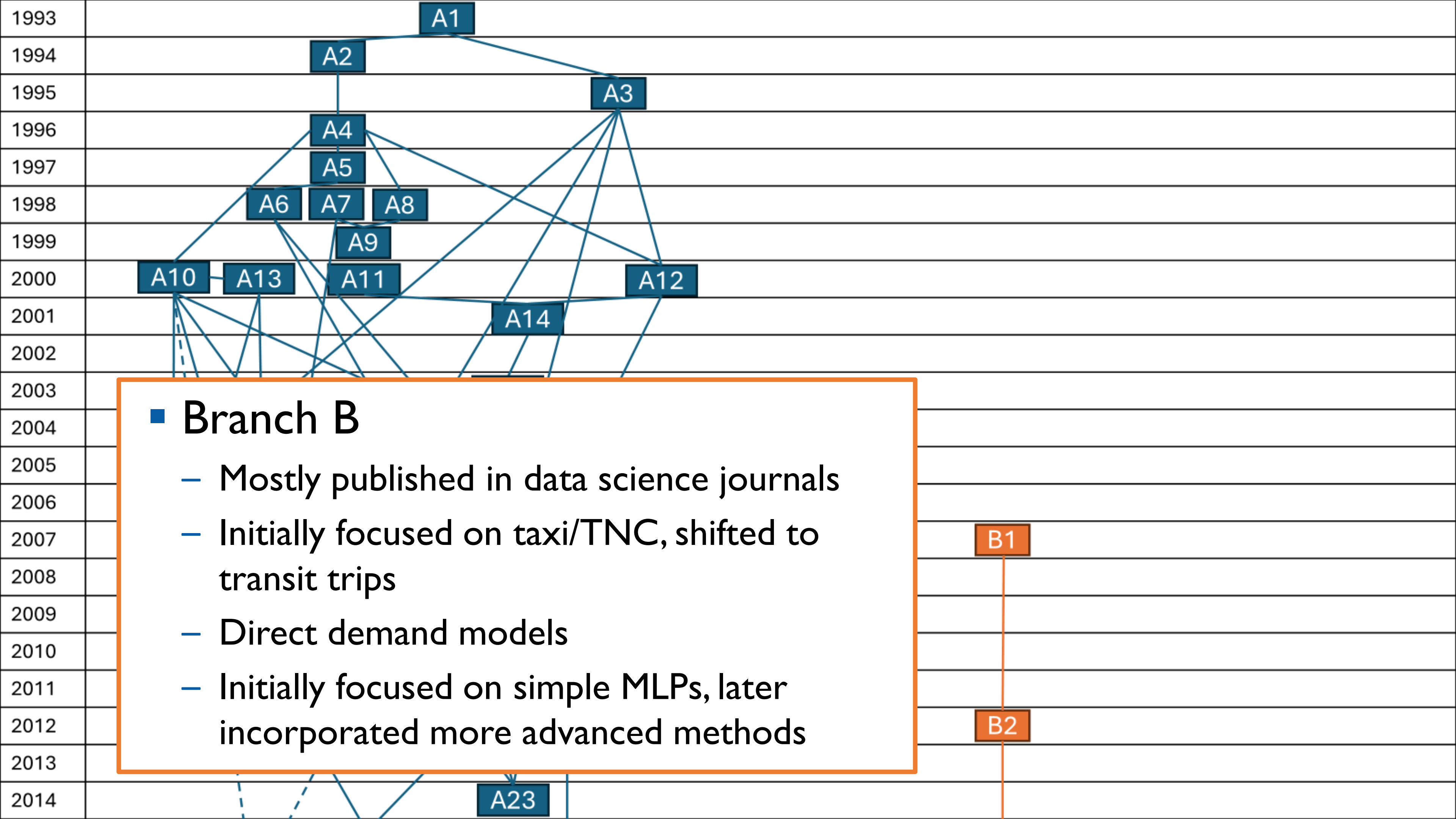
Vince Bernardin, PhD | Vice-President

vince@caliper.com | +1 812-459-3500

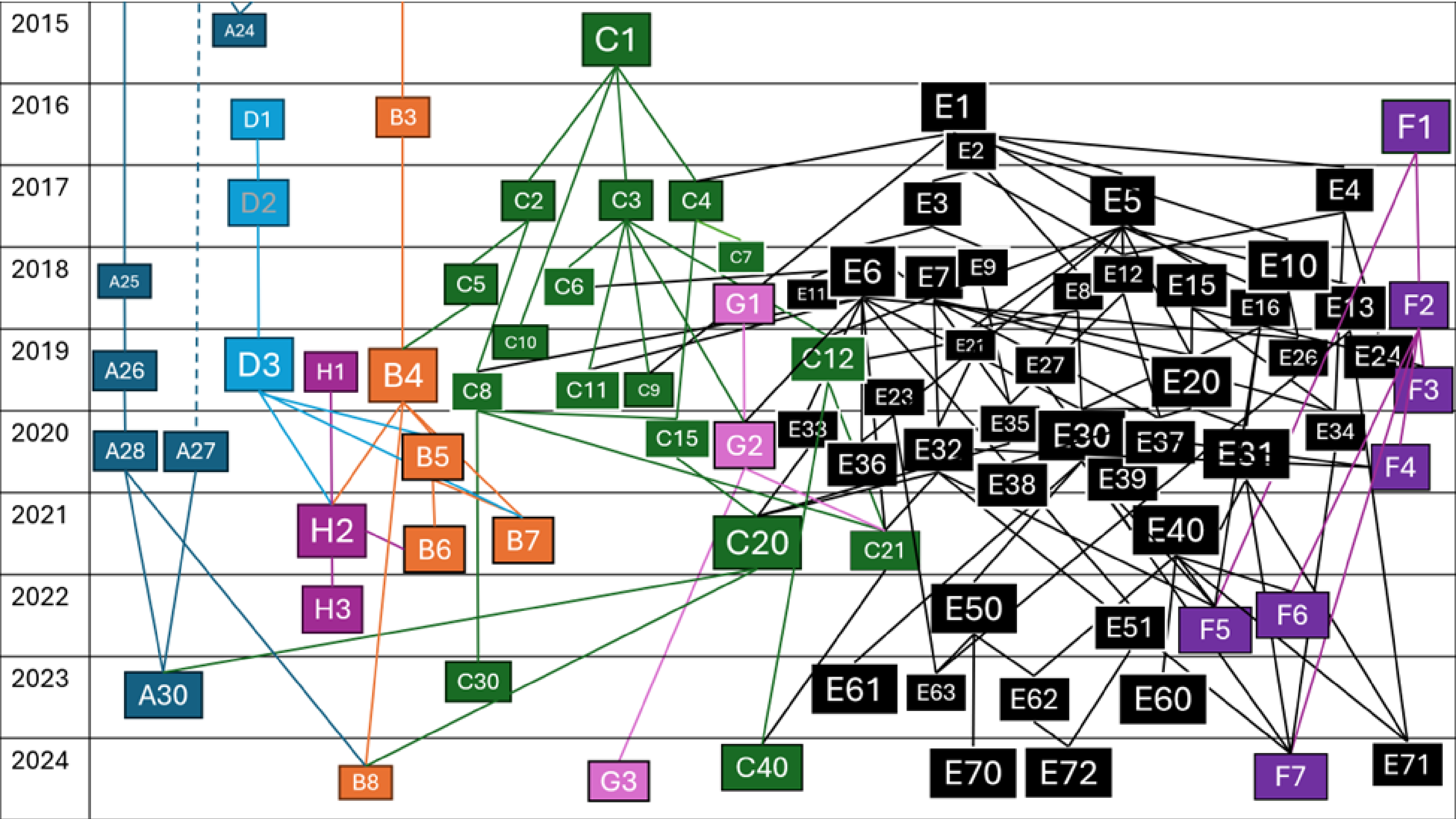

1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A13
A11
A12
A14
A15
A16
A17
A18
B1
A19
A20
A21
A22
B2
A23

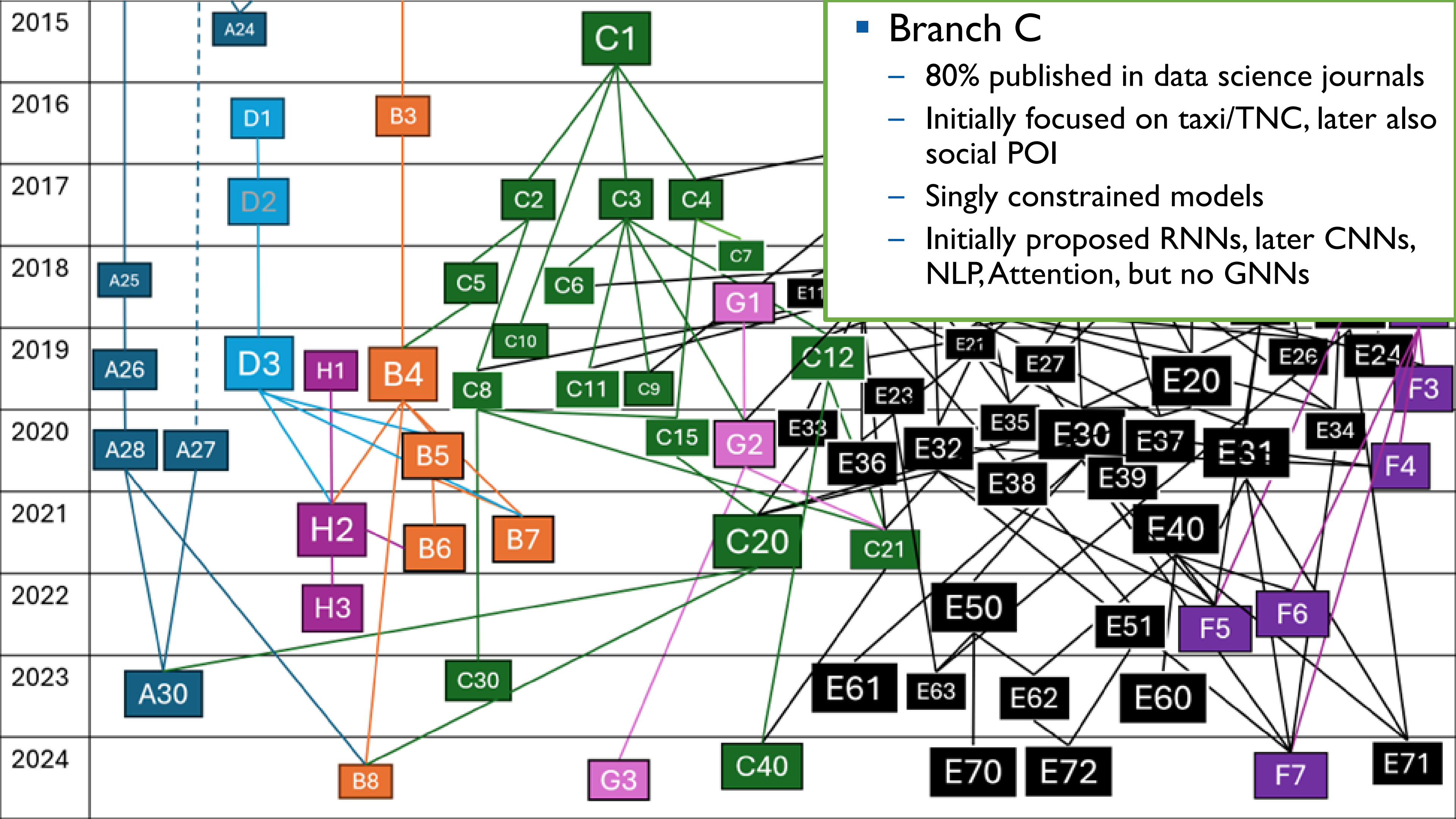


- Branch A
  - Mostly published in geography/GIS and transportation journals
  - Initially focused on commuting, later various applications
  - Direct demand models
  - Mostly focused on simple MLPs



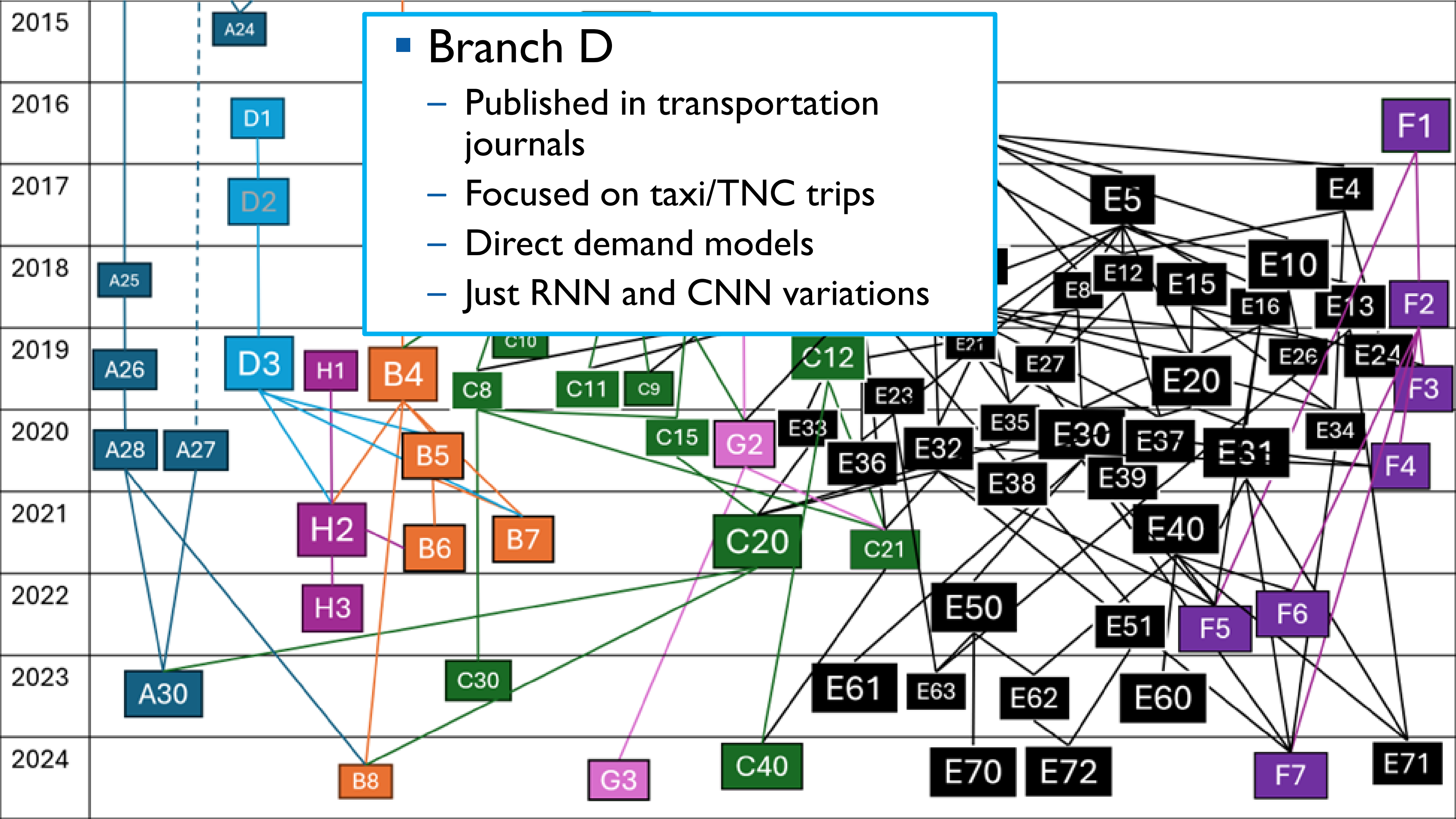
- Branch B
  - Mostly published in data science journals
  - Initially focused on taxi/TNC, shifted to transit trips
  - Direct demand models
  - Initially focused on simple MLPs, later incorporated more advanced methods


2015
A24
C1
2016
D1
B3
E1
F1
E2
2017
D2
C2
C3
C4
E3
E5
E4
2018
A25
C5
C6
C7
G1
E11
E6
E7
E9
E8
E12
E15
E10
E16
E13
F2
2019
A26
D3
H1
B4
C10
C8
C11
C9
C12
E21
E27
E26
E24
E20
F3
E23
2020
A28
A27
B5
C15
G2
E33
E36
E32
E35
E30
E37
E31
E34
E38
E39
F4
2021
H2
B6
B7
C20
C21
E40
2022
H3
E50
E51
F5
F6
2023
A30
C30
E61
E63
E62
E60
2024
B8
G3
C40
E70
E72
F7
E71

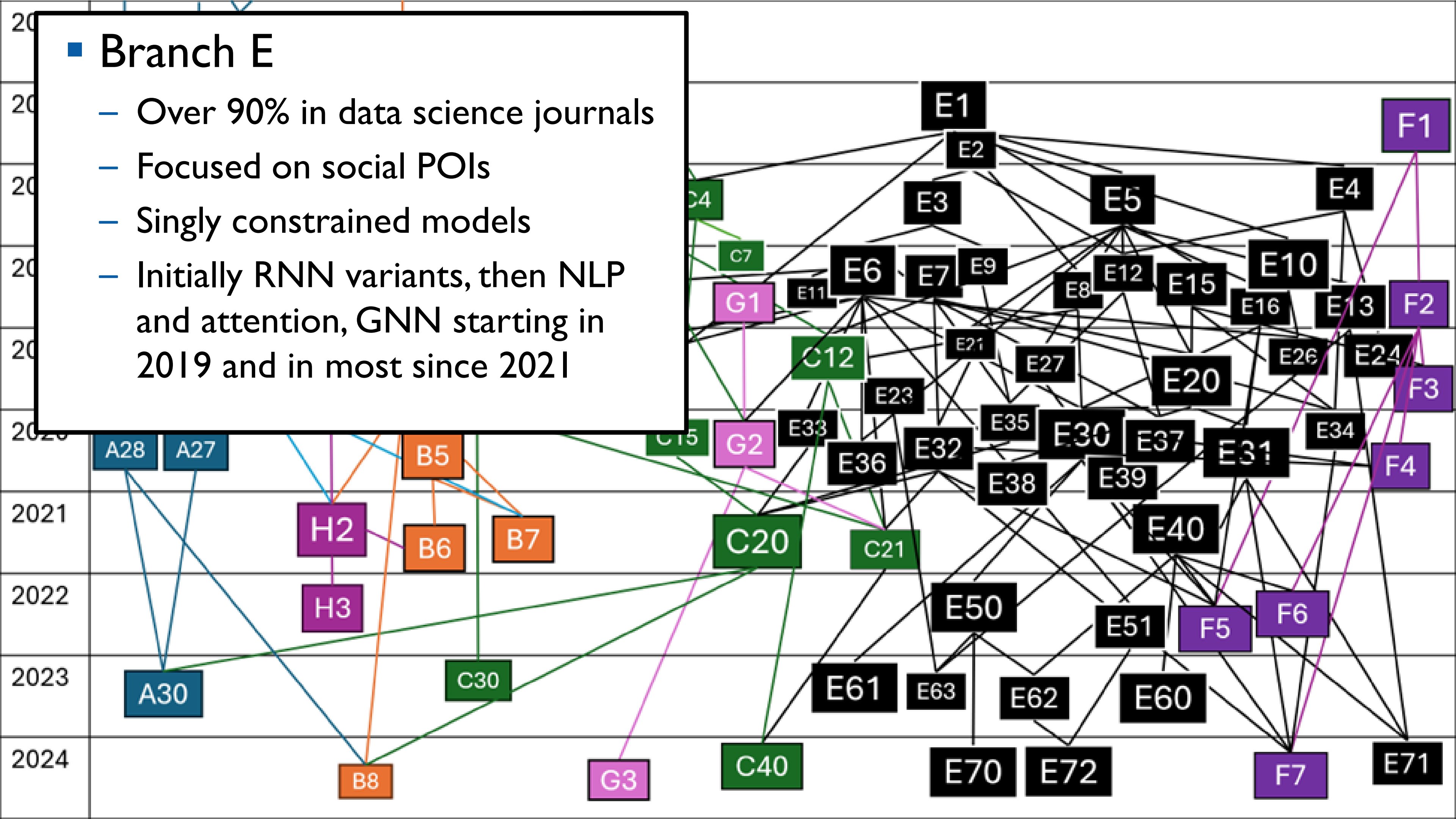


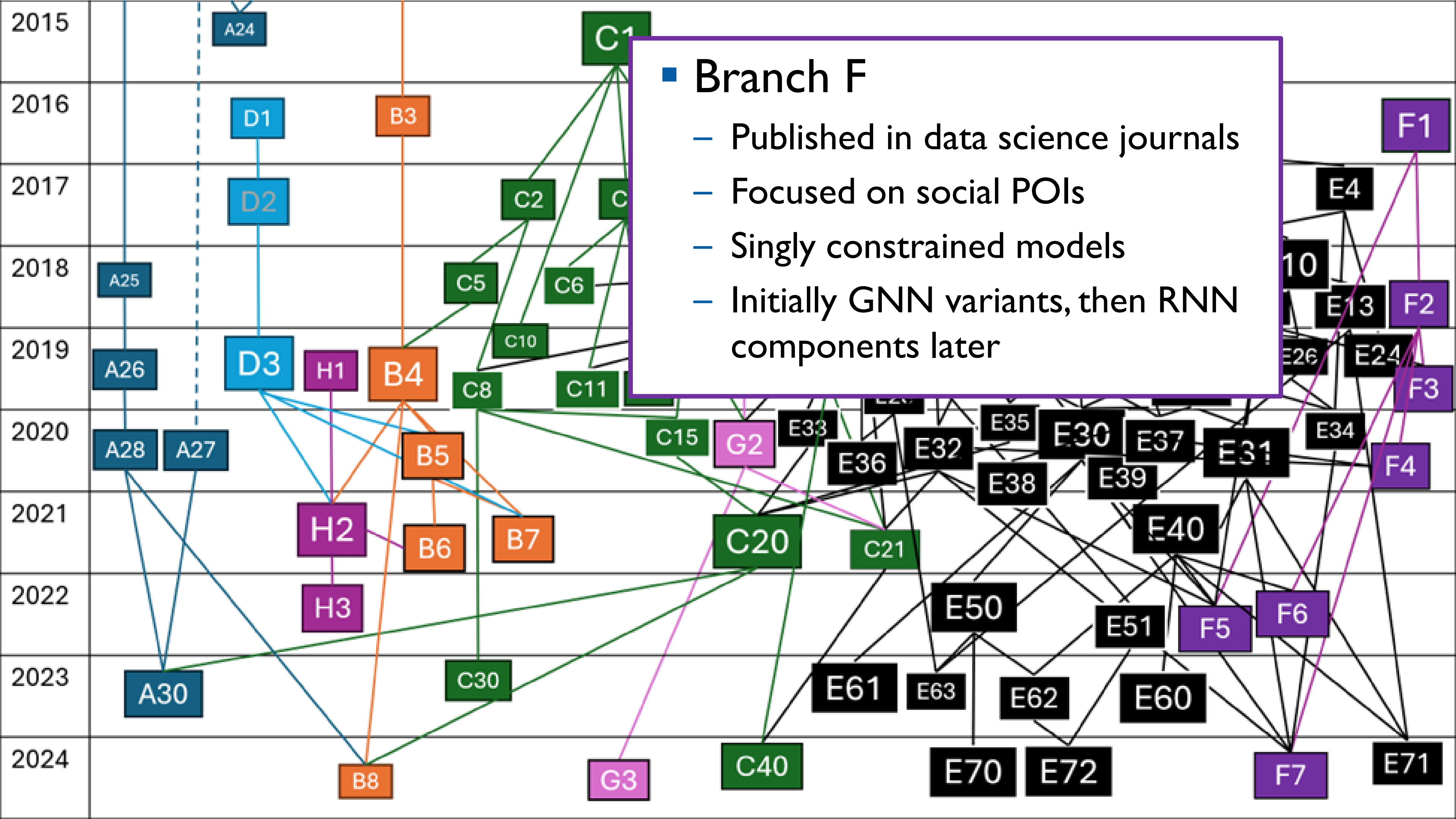
- Branch C
  - 80% published in data science journals
  - Initially focused on taxi/TNC, later also social POI
  - Singly constrained models
  - Initially proposed RNNs, later CNNs, NLP, Attention, but no GNNs

- Branch D
  - Published in transportation journals
  - Focused on taxi/TNC trips
  - Direct demand models
  - Just RNN and CNN variations

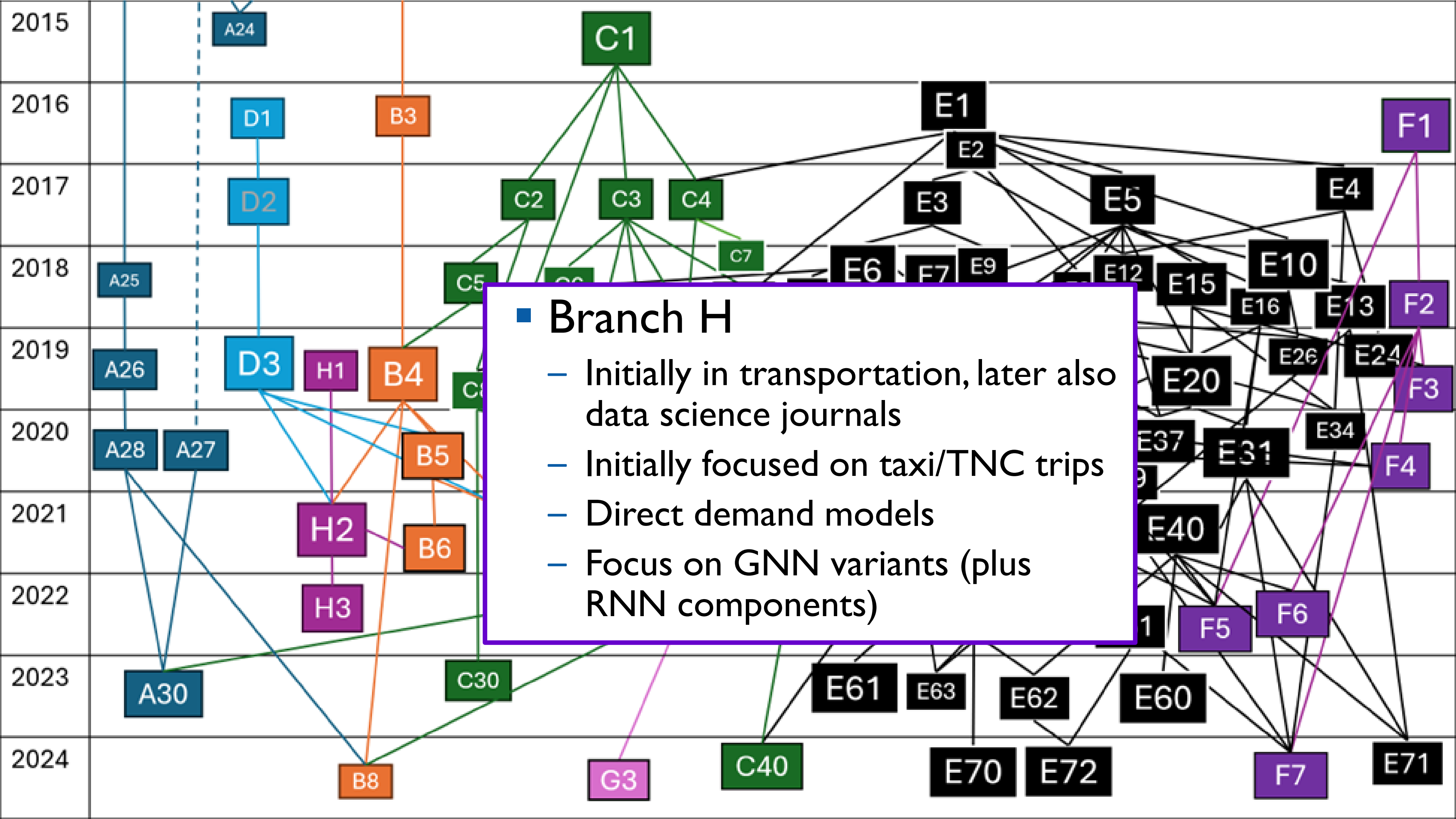


- Branch E
  - Over 90% in data science journals
  - Focused on social POIs
  - Singly constrained models
  - Initially RNN variants, then NLP and attention, GNN starting in 2019 and in most since 2021





- Branch F
  - Published in data science journals
  - Focused on social POIs
  - Singly constrained models
  - Initially GNN variants, then RNN components later



Branch H

- Initially in transportation, later also data science journals
- Initially focused on taxi/TNC trips
- Direct demand models
- Focus on GNN variants (plus RNN components)